The Laryngoscope

Intersegmenter Variability in High-Speed Laryngoscopy-Based Glottal Area Waveform Measures

Youri Maryn, PhD; Monique Verguts, MD; Hannelore Demarsin, MD; Joost van Dinther, MD; Pablo Gomez, PhD; Patrick Schlegel, MSc; Michael Döllinger, PhD

Objectives/Hypothesis: High-speed videoendoscopy (HSV) has potential to objectively quantify vibratory vocal fold characteristics during phonation. Glottal Analysis Tools (GAT) version 2018, developed in Erlangen, Germany, is software for determining various glottal area waveform (GAW) quantities. Before having GAT analyze HSV videos, segmenters have to define glottis manually across videos in a semiautomatic segmentation protocol. Such interventions are hypothesized to induce variability of subsequent GAW measure computation across segmenters and may attenuate GAT measures' reliability to a certain point. This study explored intersegmenter variability in GAT's GAW measures based on semiautomatic image processing.

Study Design: Cohort study of rater reliability.

Methods: In total, 20 HSV videos from normophonic and dysphonic subjects with various laryngeal disorders were selected for this study and segmented by three trained segmenters. They separately segmented glottis areas in the same frame sets of the videos. Upon analysis of GAW, GAT offers 46 measures related to topologic GAW dynamic characteristics, GAW periodicity and perturbation characteristics, and GAW harmonic components. To address GAT's reliability, intersegmenter-based variability in these measures was examined with intraclass correlation coefficient (ICC).

Results: In general, ICC behavior of the 46 GAW measures across three raters was highly acceptable. ICC of one parameter was moderate (0.5 < ICC < 0.75), good for seven parameters (0.75 < ICC < 0.9), and excellent for 38 parameters (0.9 < ICC).

Conclusions: Overall, high ICC values confirm clinical applicability of GAT for objective and quantitative assessment of HSV. Small intersegmenter differences with actual small parameter differences suggest that manual or semiautomatic segmentation in GAT does not noticeably influence clinical assessment outcome. To guarantee the software's performance, we suggest segmentation training before clinical application.

Key Words: High-speed videoendoscopy, glottis segmentation, reliability, quantitative parameters.

Level of Evidence: 2b

***Laryngoscope*, 00:1–8, 2019**

INTRODUCTION

There are various endoscopic imaging techniques to assess the integrity of phonatory vocal fold vibration: stroboscopy, kymography, and high-speed videoendoscopy (HSV). For background information on all these techniques and their advantages as well as disadvantages, the reader is referred to the respective chapters in Kendall and Leonard,[1] Ma and Yiu,[2] and Döllinger.[3] Regarding HSV, however, the following pro arguments can be raised. First, it enables assessment of instationary dynamics like voice onsets.[4] Second, it can be applied to all types of vocal fold vibrations, even the least periodic ones.[5] Third, it can be utilized in case of very short and unstable (for example due to voice breaks and spasms) phonations with varying period properties.[6] Fourth, it allows for a detailed spatiotemporal analysis of vocal fold vibrations.[7] Fifth, HSV can enable important refinements in the diagnosis and assessment of vocal fold disorders.[8,9] Sixth, when applying a high enough framerate, HSV even allows for analyzing the singing voice.[10] However, there are also arguments complicating the clinical implementation of HSV. It is expensive, lacks real-time visualization, requires a relatively long time for postrecording playback, necessitates a high volume of data storage, and still mainly involves visual perceptual interpretation by clinicians. From stroboscopy, for example, it is known to have mostly poor rater reliability.[11,12] Use of a visually supported rating form with clear definitions of the laryngoscopic items, such as the Voice-Vibratory Assessment with Laryngeal Imaging form, may enhance stroboscopy's as well as HSV's rater reliability.[13] Key

From the Department of Otorhinolaryngology–Head and Neck Surgery (Y.M., M.V., H.D., J.V.), European Institute for Otorhinolaryngology–Head and Neck Surgery, GasthuisZusters Antwerpen Sint-Augustinus, Wilrijk/Antwerp, Belgium; Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology–Head and Neck Surgery (P.G., P.S., M.D.), University Hospital Erlangen, Friedrich–Alexander University Erlangen–Nürnberg, Erlangen, Germany; Department of Speech, Language, and Hearing Sciences (Y.M.); Faculty of Education, Health, and Social Work (Y.M.), University College of Ghent, Ghent, Belgium; Faculty of Psychology and Educational Sciences (Y.M.), School of Logopedics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; Faculty of Medicine and Health Sciences (Y.M.), University of Antwerp, Antwerp, Belgium; Phonanium (Y.M.), Lokeren, Belgium; and Department of Otorhinolaryngology and Voice Disorders (M.V.), Diest General Hospital, Diest, Belgium.

Additional supporting information may be found in the online version of this article.

This Manuscript was received on August 23, 2019, and accepted for publication on November 26, 2019.

P.S. was funded by the Deutsche Forschungsgemeinschaft under grant numbers DO1247/8-1 and BO4399/2-1.

The authors have no other funding, financial relationships, or conflicts of interest to disclose.

Send correspondence to Youri Maryn, PhD, Oosterveldlaan 24, Wilrijk 2610, Belgium. E-mail: youri.maryn@gza.be

DOI: 10.1002/lary.28475

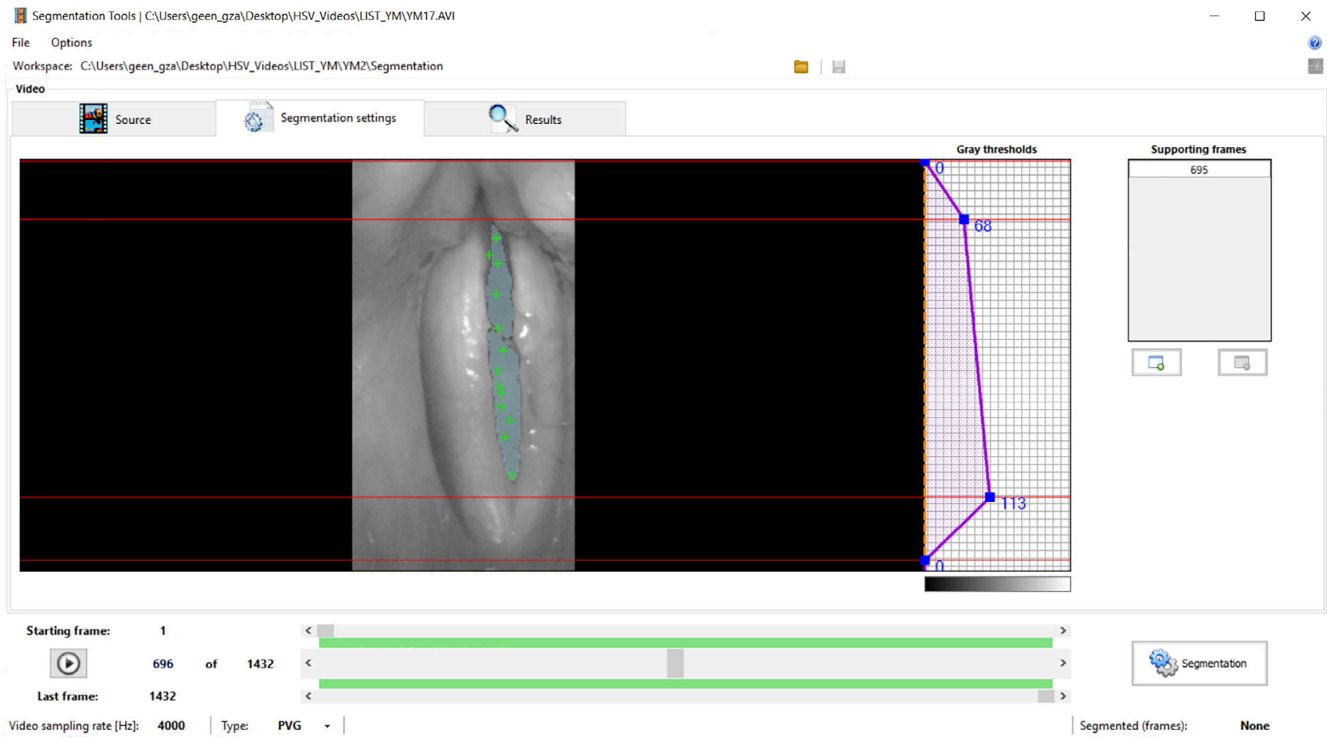

Fig. 1. Screenshot taken during the segmentation of one of the videos in this study to illustrate the user interface. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

TABLE I.
Gender, Age, and Laryngeal Diagnosis Together With the HSV Frame Rate and Segmented Interval Lengths of the 20 Subjects in the Study.

				High-Speed Video Characteristics	
Subject No.	Gender	Age, yr	Laryngeal Diagnosis	Frame Rate, fps	No. of Frames
1	F	47	Left vocal fold polyp with contralateral reactive mass	4,000	1,000
2	F	38	Normal	4,000	1,000
3	F	26	Firm nodules	4,000	690
4	F	34	Subtle nodules	4,000	1,000
5	M	66	Left vocal fold ectasia and right vocal fold hypotrophy	4,000	1,000
6	F	29	Right vocal fold fibrosis	4,000	1,000
7	F	29	Right vocal fold fibrosis	4,000	1,000
8	F	60	Bilateral Reinke's edema	4,000	1,000
9	F	40	Right vocal fold ectasia	8,000	1,000
10	F	19	Normal	4,000	651
11	F	19	Normal	4,000	1,000
12	M	42	Normal	8,000	1,000
13	F	37	Left vocal fold cyst	4,000	281
14	M	64	Bilateral sulcus vergeture	8,000	1,000
15	F	54	Subtle nodules	4,000	1,000
16	M	37	Subtle right vocal fold polypoid swelling	4,000	1,000
17	F	17	Subtle nodules	4,000	761
18	F	34	Left vocal fold cyst	4,000	1,000
19	F	51	Bilateral Reinke's edema	8,000	736
20	M	17	Normal	4,000	1,000

F = female; M = male.

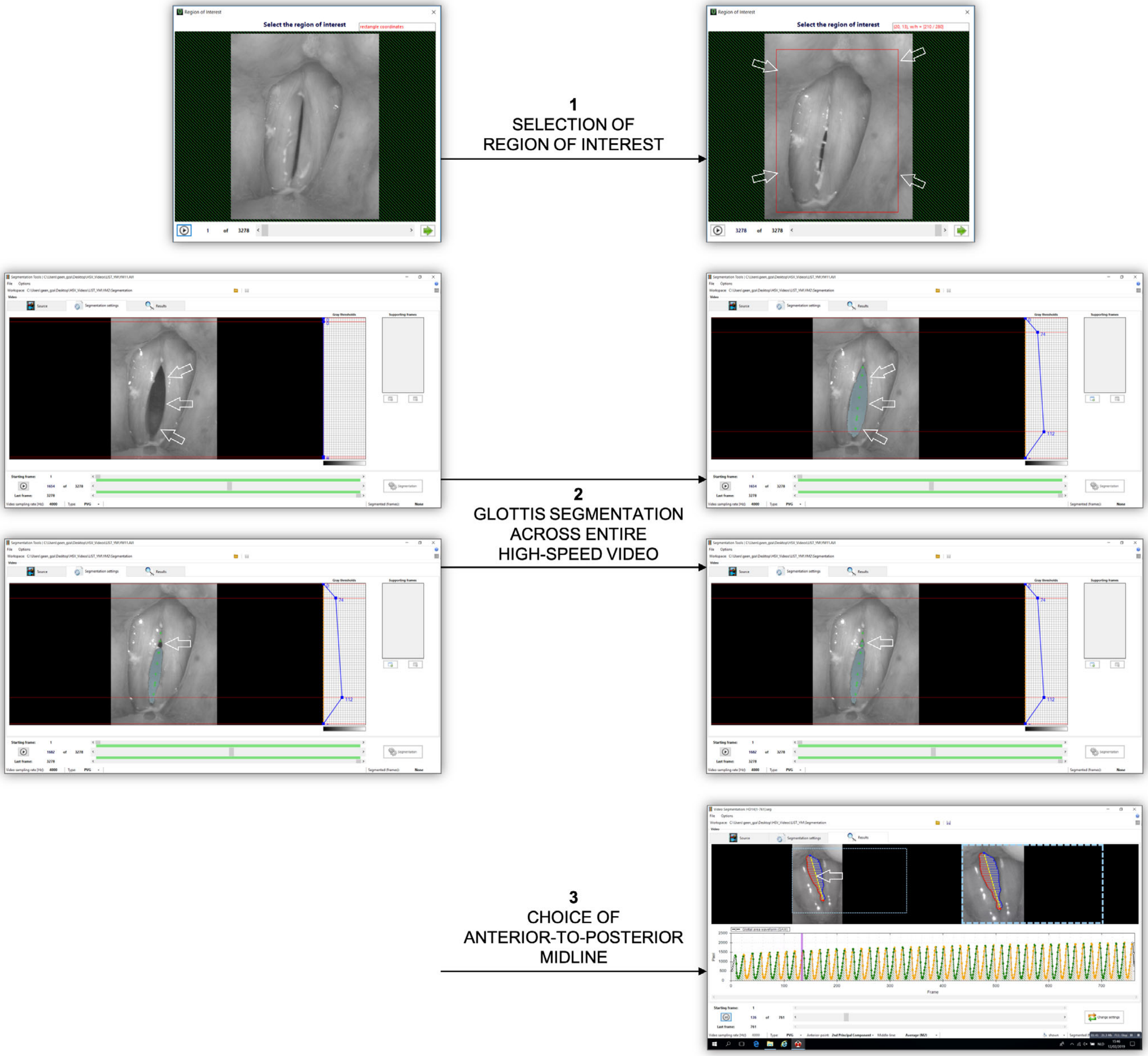

Fig. 2. Illustration of the three major consecutive subjective user interventions, as indicated by the white-lined arrows, during the segmentation process before retrieving the glottal area waveform measures in the program Glottal Analysis Tools. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

voice-vibratory items, such as vibratory amplitude, mucosal wave, and symmetry, are rated more reliable in HSV videos than in stroboscopic videos.[13,14] However, given the limited rater reliability in general, it is believed that endoscopic assessment of vocal fold vibration may benefit from more objective and quantitative analysis of HSV images.[15]

Many studies have been published on quantitative and computerized analysis of voice-vibratory characteristics in HSV, for example Döllinger et al.,[7] Patel et al.,[16] and Yamauchi et al.[17] One example of software for establishing glottal area waveform (GAW) from HSV videos and subsequently computing various GAW measures, is the Glottal Analysis Tools (GAT; University Hospital Erlangen, Erlangen, Germany). Depending on the length of the video, GAT's 2018 version calculates up to 63 GAW characteristics in terms of topologic GAW dynamics, GAW periodicity and perturbation, and GAW harmonic components.[18–20] However, before receiving GAW and its measures, the following manual user interaction is required: determination of laryngeal region of interest, segmentation of glottis (grey value threshold determination), and determination of the glottal midline. Although GAT offers an extensive user guide and a user-friendly interface (see Fig. 1 for a screenshot taken during a segmentation with GAT software), the requirement of these human interventions is hypothesized to induce a certain variability in these GAW measures across clinicians/researchers (i.e., segmenters) and thereby may reduce GAT's reliability as a clinical voice physiology assessment tool. As far as we know, this type of intersegmenter-based variability has never been investigated for HSV analysis before. Therefore, this study addressed the following clinically relevant research question: How reliable are HSV-based GAW measures?

MATERIALS AND METHODS

Subjects

The study group consisted of 20 subjects. There were 15 females ranging in age from 17 to 60 years (mean age = 35.6 years), and five males with ages ranging between 17 and 66 years (mean age = 45.2 years). Fifteen subjects were diagnosed with various vocal fold abnormalities, whereas five subjects had normal vocal folds. Gender, age, and laryngeal diagnosis are listed per subject in Table I.

High-Speed Videoendoscopy Equipment

A Machida (Chiba, Japan) LY-C30 rigid endoscope with 10-mm diameter and 70° view, a CUDA (Jacksonville, FL) Surgical E300 xenon lightsource with 300 W, a Photron (Tokyo, Japan) Fastcam MC2.1 unit with MC2 camera head, and a 15-to-36 lens adaptor were used. Laryngograph (London, United Kingdom) VoiceSuite10 software was applied to record and organize the high-speed videos. Sixteen of the videos used in this study were recorded with a frame rate of 4,000 fps. The four remaining videos were recorded with a frame rate of 8,000 fps. Whenever available, a 1,000-frame interval was extracted from the HSV videos prior to segmentation. This was possible in 15 videos. In the five remaining videos, there were less than 1,000 frames available (i.e., 281–761 frames) (Table I).

HSV Video Segmentation

Prior to receiving GAW measures, the HSV videos need to be segmented (i.e., glottis has to be differentiated from vocal fold tissue). Within such a segmentation procedure, three major human interventions are required. First, the laryngeal region of interest has to be selected. This means that the entire glottis has to be visible across the entire video interval of interest. Second, the glottis has to be defined across all images in the HSV video. Therefore, the following is required: rough determination of anterior and posterior vocal fold positions; determination of the glottis by seed points in one or more so called supporting frames, from there the glottis will be determined by grey-value thresholding over the video[21]; and adjustment of grey-value thresholds over one or more video frames, which is necessary when lighting conditions change during the recording. After these steps the glottal area is calculated. Third, for computing left–right symmetry characteristics, the glottal midline has to be determined. These steps, as well as all other considerations, are explained and illustrated in detail in Schlegel et al.[22] Figure 2 provides a graphical summary of the user interventions.

Three segmenters with different professional backgrounds were asked to segment the 20 selected HSV video intervals. Segmenter 1 (female, age = 36 years) was a laryngology consultant with more than 7 years of experience in specialized medical voice care and surgery. Segmenter 2 (male, age = 42 years) was a speech–language pathologist with more than 15 years of experience in clinical voice assessment, therapy, and research. These two segmenters visited the University Hospital in Erlangen, Germany to receive extensive training in the segmentation of HSV's with the GAT software over the course of 2 days. Segmenter 3 (female, age = 30 years) was in her third year as a trainee in otorhinolaryngology. She learned about HSV segmentation during a peer-learning session together with segmenters 2 and 3. All three clinicians segmented the 20 HSV video intervals individually during a single session of approximately 3 hours. This resulted in a set of 60 segmented HSV videos.

Glottal Area Waveform Measures

Once the glottis was delineated across the videos according to every segmenter's opinion, 62 out of 63 GAW measures were retrieved automatically from the GAT software: four related to fundamental frequency, seven related to amplitude perturbation, 11 related to period perturbation, four related to energy perturbation, 10 related to noise, four related to so-called mechanical features, two related to the first derivative of the GAW, nine related to symmetry, and finally 11 measures related to glottal area waveform topology itself. The 63rd measure, called GNE, was not calculated because it required a window length that exceeded the length of the signals examined in this study. Supporting Information, Appendix 1, lists all measures together with a short explanation of each. More profound elaboration and formula on how most of these measures are calculated are provided in Schlegel et al.[22] Additional information on the concept behind a lot of these GAW measures can also be found in Buder,[23] as several of these measures' algorithms have also been applied to acoustic waveforms instead of glottal area waveforms.

From the 63 GAW measures that are provided by the GAT software, 17 measures (i.e., Shim%, APQ3, APQ5, APQ11, PPQ3, PPQ5, PPQ11, RAPv1, RAPv2, EPQ3, EPQ5, EPQ11, HarmIntens, SNRv1, SNRv2, CPPv2, and GNE) are described in the program's user guide as not recommended for use due to various reasons (e.g., redundancy or being not suitable). These 17 measures were therefore excluded from further analysis, leading to a final study set of 46 GAW measures and resulting in a database with 2,760 data points (i.e., three raters * 20 HSL video portions * 46 GAW measures).

Statistics

To assess intersegmenter reliability of the semiobjectively gathered GAW measures (i.e., variability in GAW measures across segmenters), the intraclass correlation coefficient (ICC) was calculated for all parameters according to the guideline of Koo and Li.[24] The two-way random-effects ICC model was chosen because we imagined our reliability results to be generalized to any raters with the same characteristics as the selected raters in the present study. The single-rater ICC type was selected because we did not intend to use the mean value of three raters for assessment purposes. The consistency ICC definition indicates the degree to which one segmenter's GAW measure can be equated to another segmenter's GAW measure plus a systematic error. This definition was chosen because GAW measures are at ratio measurement level and therefore expected not to be equal between segmenters. IBM SPSS Statistics for Mac version 20.0 (IBM, Armonk, NY) was applied for the calculation of this ICC. Portney and Watkins[25] suggest that ICC values <0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values >0.90 indicate excellent reliability.

TABLE II.
Descriptive Data of the 46 GAW Measures for the Three Segmenters Ordered According to Reliability (in Terms of ICC).

	Segmenter 1					Segmenter 2					Segmenter 3					
GAW Measure*	Min	Max	Range	Mean	SD	Min	Max	Range	Mean	SD	Min	Max	Range	Mean	SD	ICC
ASI	0.84	1.25	0.41	0.96	0.11	0.78	1.20	0.42	0.95	0.09	0.74	1.19	0.44	0.96	0.11	0.653
DRS	0.58	1.42	0.85	0.98	0.19	0.74	1.47	0.73	0.97	0.18	0.73	1.35	0.62	0.97	0.19	0.762
CPPv1	10.9	26.2	15.3	20.1	4.0	8.0	26.9	18.9	19.5	5.2	2.2	26.1	23.9	19.9	5.7	0.780
HNR	7.20	19.00	11.80	12.79	3.53	8.00	21.10	13.10	13.19	3.32	7.20	18.00	10.80	13.04	3.36	0.820
HarmMax	328.0	1124.6	796.6	610.3	251.7	328.0	1240.0	912.0	644.6	288.1	304.0	1124.6	820.6	594.8	263.1	0.823
DRSI	0.58	0.96	0.38	0.84	0.11	0.68	0.96	0.28	0.84	0.09	0.66	0.96	0.30	0.83	0.09	0.860
RQ	0.74	2.97	2.23	1.53	0.59	0.75	3.11	2.36	1.44	0.53	0.73	2.80	2.07	1.46	0.55	0.891
SS	−0.21	0.07	0.28	−0.04	0.07	−0.23	0.10	0.33	−0.04	0.08	−0.26	0.10	0.36	−0.04	0.08	0.891
SubharmMin	61.6	412.0	350.4	226.3	91.9	61.6	608.0	546.4	261.3	118.3	61.6	608.0	546.4	254.8	118.7	0.913
SSI	0.02	0.21	0.19	0.08	0.05	0.02	0.23	0.22	0.08	0.05	0.02	0.26	0.24	0.08	0.06	0.913
OQ	0.56	1.00	0.44	0.92	0.15	0.57	1.00	0.43	0.94	0.13	0.56	1.00	0.44	0.93	0.15	0.915
ASI	0.69	0.96	0.27	0.88	0.07	0.73	0.96	0.23	0.88	0.07	0.72	0.97	0.25	0.88	0.08	0.918
PVI	0.20	22.90	22.70	4.00	6.63	0.20	14.70	14.50	3.07	4.56	0.30	23.50	23.20	4.03	6.39	0.922
AmplitQ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.925
JittMean	0.07	1.08	1.01	0.24	0.24	0.07	0.86	0.79	0.23	0.18	0.09	1.28	1.19	0.26	0.27	0.925
CQ	0.25	0.58	0.32	0.42	0.09	0.24	0.57	0.33	0.43	0.08	0.26	0.58	0.32	0.43	0.08	0.927
AVI	−0.98	2.69	3.68	0.55	0.82	−1.02	2.71	3.73	0.45	0.84	−1.02	2.67	3.68	0.42	0.84	0.933
Jitt%	1.70	20.70	19.00	5.93	5.38	1.60	16.50	14.90	5.54	4.06	2.00	24.20	22.20	6.36	5.98	0.933
PPF	1.70	21.60	19.90	6.02	5.61	1.60	17.00	15.40	5.62	4.21	2.00	25.00	23.00	6.47	6.23	0.934
TP	0.82	0.98	0.17	0.95	0.05	0.85	0.98	0.13	0.95	0.04	0.80	0.98	0.18	0.94	0.05	0.934
JittRat	16.9	206.7	189.8	59.3	53.8	15.8	164.7	148.9	55.0	40.8	20.0	241.8	221.8	63.5	59.8	0.935
JittFact	1.70	21.30	19.60	6.00	5.52	1.60	17.00	15.40	5.61	4.21	2.10	24.20	22.10	6.42	6.05	0.938
PA	−0.21	0.23	0.44	0.07	0.10	−0.21	0.23	0.43	0.07	0.09	−0.22	0.23	0.45	0.07	0.10	0.941
NNE	−26.3	−4.0	22.3	−19.9	5.8	−25.7	−4.2	21.5	−20.4	5.5	−25.6	−3.8	21.8	−20.7	5.7	0.944
SF	−30.8	−17.2	13.6	−23.2	3.9	−32.3	−17.6	14.7	−23.8	4.1	−32.0	−18.4	13.6	−24.0	4.0	0.945
GAI	0.38	1.81	1.43	0.93	0.42	0.38	1.76	1.38	0.88	0.36	0.36	1.78	1.42	0.90	0.41	0.957
MADR	0.22	2.87	2.65	1.50	0.80	0.21	2.82	2.61	1.45	0.76	0.21	2.93	2.72	1.48	0.82	0.970
EPF	0.80	31.90	31.10	4.97	7.04	0.80	30.40	29.60	4.49	6.71	1.00	36.30	35.30	4.66	7.87	0.971
ShimMean	0.08	1.65	1.57	0.40	0.47	0.07	1.92	1.85	0.37	0.49	0.07	1.44	1.36	0.35	0.42	0.972
SQ	0.74	1.77	1.03	1.24	0.29	0.75	1.82	1.07	1.25	0.33	0.73	1.74	1.02	1.23	0.30	0.972
APF	0.83	0.99	0.16	0.96	0.05	0.81	0.99	0.18	0.96	0.05	0.85	0.99	0.14	0.96	0.04	0.973
APF	0.90	18.70	17.80	4.65	5.59	0.80	21.70	20.90	4.32	5.85	0.80	16.50	15.70	4.05	5.01	0.973
WMCMax	0.93	1.00	0.07	0.99	0.02	0.95	1.00	0.05	0.99	0.01	0.93	1.00	0.07	0.99	0.02	0.976
PAI	0.01	0.24	0.23	0.12	0.07	0.00	0.27	0.27	0.12	0.08	0.00	0.25	0.25	0.12	0.07	0.976
AQ	0.42	0.63	0.21	0.54	0.06	0.43	0.64	0.22	0.54	0.06	0.42	0.63	0.21	0.54	0.06	0.977

(Continues)

TABLE II.
Continued

GAW Measure*	Segmenter 1 Min	Max	Range	Mean	SD	Segmenter 2 Min	Max	Range	Mean	SD	Segmenter 3 Min	Max	Range	Mean	SD	ICC
SI	−0.15	0.27	0.42	0.08	0.12	−0.15	0.28	0.43	0.08	0.13	−0.16	0.26	0.42	0.07	0.12	0.977
PQ	0.12	0.39	0.27	0.17	0.06	0.10	0.40	0.29	0.17	0.06	0.09	0.39	0.30	0.17	0.06	0.989
Stifness	631.7	2200.0	1568.2	1136.4	367.5	645.9	2274.5	1628.7	1147.1	383.9	617.2	2241.0	1623.8	1116.9	377.2	0.989
ALR	2.10	25.00	22.90	11.40	6.06	2.10	24.40	22.30	11.42	6.28	2.10	25.50	23.40	11.53	6.57	0.992
ClosVelocP	0.13	2.37	2.24	1.08	0.57	0.13	2.34	2.21	1.08	0.56	0.13	2.45	2.32	1.08	0.57	0.996
GGI	0.00	0.62	0.62	0.21	0.22	0.00	0.62	0.62	0.21	0.22	0.00	0.64	0.64	0.21	0.23	0.997
WMCMean	0.67	1.00	0.33	0.96	0.08	0.67	1.00	0.32	0.95	0.09	0.66	1.00	0.34	0.96	0.09	0.997
AccelP	165.8	6138.1	5972.3	1893.1	1565.3	161.6	6065.2	5903.6	1888.8	1539.3	162.0	6346.0	6183.9	1885.7	1527.2	0.997
WSI	0.48	1.00	0.52	0.87	0.15	0.48	1.00	0.52	0.86	0.15	0.48	1.00	0.51	0.86	0.15	0.997
CycleDur	1.60	7.60	6.00	4.38	1.53	1.70	7.70	6.00	4.40	1.53	1.60	7.50	5.90	4.39	1.51	1.000
f_O	132.0	619.3	487.3	263.2	113.3	129.8	619.2	489.4	262.6	113.1	133.8	619.5	485.7	262.8	112.7	1.000

*See Supporting Information, Appendix 1, in the online version of this article for full name of these measures.
GAW = glottal area waveform; ICC = intraclass correlation coefficient; Min = minimum; Max = maximum; SD = standard deviation.

Ethics

This retrospective study of these HSV videos was approved by the ethics committee of GasthuisZusters Antwerpen Sint-Augustinus (approval code 190203RETRO).

RESULTS

The descriptive statistics per segmenter of the 46 GAW measures are provided in Table II. The intersegmenter reliability data for these measures are also presented in Table II as well as in Figure 3. From both Figure 3 and Table II, it is clear that none of the 46 included GAW measures has poor (<0.5) intersegmenter reliability. Only one measure (i.e., amplitude symmetry with ICC = 0.653) showed moderate intersegmenter reliability. Seven GAW measures (15%) had good reliability: dynamic range symmetry, with ICC = 0.762; cepstral peak prominence, with ICC = 0.780; harmonics-to-noise ratio, with ICC = 0.820; maximum harmonic, with ICC = 0.823; dynamic range symmetry index = 0.860; and rate quotient and spatial symmetry, both with ICC = 0.891. All other 38 GAW measures (83%) had excellent intersegmenter reliability, with ICC values between 0.913 (for minimum subharmonic and spatial symmetry index) and 1.000 (for cycle duration and its reciprocal fundamental frequency).

DISCUSSION

HSV is an advanced imaging technique in the assessment of vocal fold physiology and pathology. To facilitate review and enable quantification of HSV videos, clinicians can utilize software for computer-aided glottis segmentation and more objective vocal fold oscillation analysis. GAT is an example of such software. First, it comes with a user-friendly and intuitive interface to define the glottal area across the entire HSV video. Second, upon this segmentation, GAT constructs a sequence of glottal area–time coordinates that can then automatically be queried for a large set of quantitative data, including data on periodicity, left–right symmetry, and glottic closure. However, because currently, multiple manual user interventions are required, GAT is still only semiautomatic and subjective to some extent, and differences in glottal area waveforms as well as their measures between clinicians and researchers (i.e., segmenters) might be anticipated. To explore the extent to which these GAW measures vary across segmenters, the glottal areas in 20 HSV video portions were specified by three trained segmenters, and intersegmenter reliability was calculated by means of ICCs. To our knowledge, this was the first time that such reliability has been examined in the domain of HSV. The results of this study clearly demonstrate that segmenter-based variability in all GAW measures is negligible. Except for amplitude symmetry, all GAW measures showed good reliability, and the majority of the measures showed even excellent reliability.

However, a couple of issues have to be noted. First, two of three segmenters in this study were trained in glottis segmentation with GAT by the developers of this

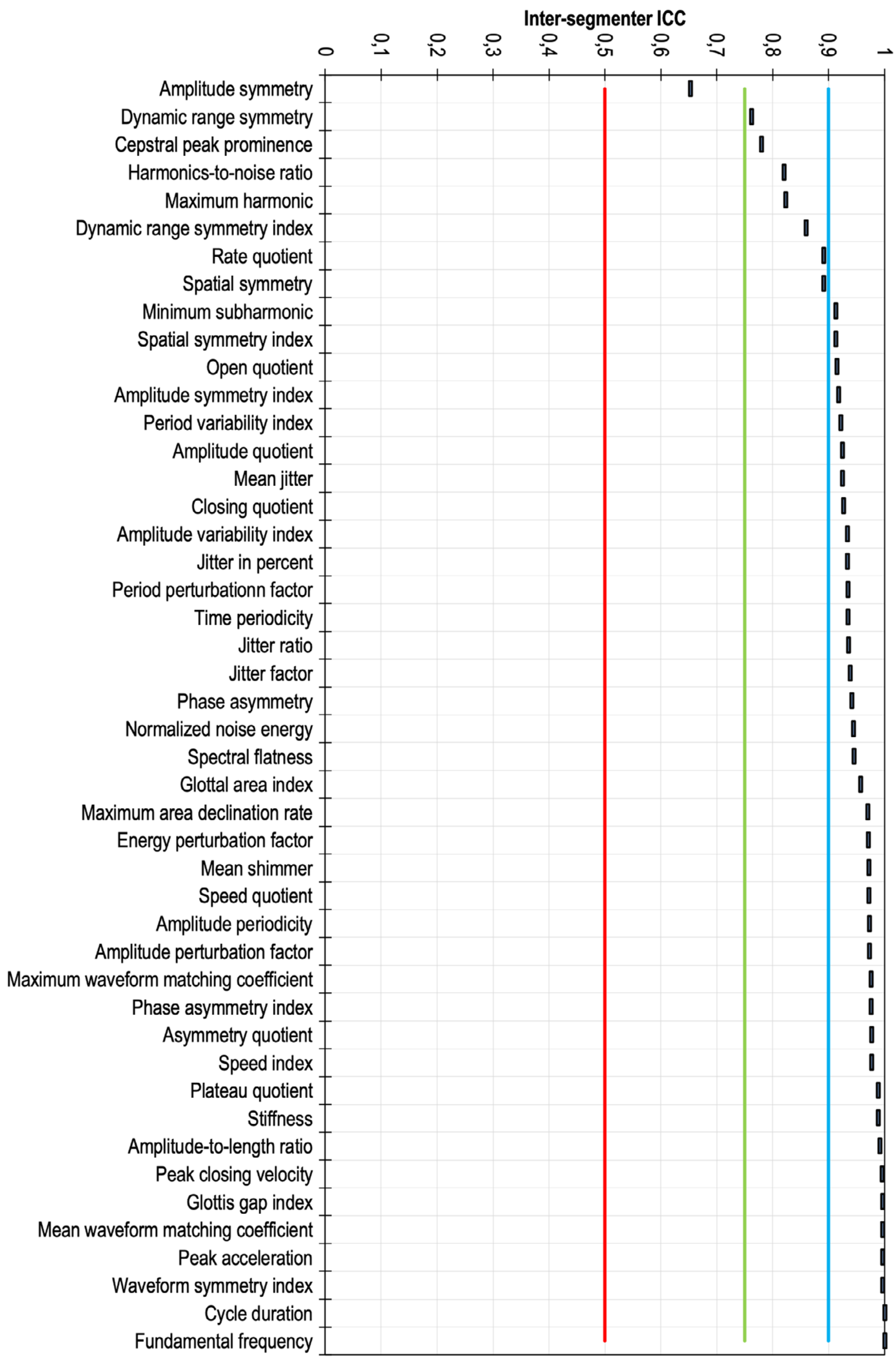

Fig. 3. The intersegmenter intraclass correlation coefficient (ICC) values for 46 glottal area waveform measures across three segmenters. Above red/green/blue line: respectively moderate/good/excellent reliability. Overall a very high reliability is given. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

program, and the third segmenter learned about this in a peer-learning session with the other two. As for example in auditory-perceptual rating of voice quality, it is acknowledged that training is important to align the raters' internal standards and reduce rating variability,[26–28] as well as for visual-perceptual ratings of laryngeal images/videos[13] or other issues to be rated visually, such as swallowing on videofluoroscopy[29] and oral health.[30] How reliable GAW measures would have been in segmenters with less instruction, and thus how generalizable the results are, will have to be investigated in a future study. However, given that glottis segmentation is a predominantly dichotomous task (i.e., choosing between glottis vs. nonglottis), it can be assumed that most laryngoscopically trained professionals yield similar reliability. Second, the quality of HSV video may influence the difficulty of manual glottis segmentation. HSV videos with sharp focus on the vocal folds and with clear contrast between the glottis and vocal folds/posterior commissure are easier to segment than blurry videos with less distinctive (i.e., more fluent) transition from glottis to nonglottis. Consequently, for postrecording manual segmentation of glottis and GAW measures' reliability, it is important to have HSV videos with sufficient quality in terms of endoscope-to-vocal folds distance, focus, and brightness.

CONCLUSION

With the large amount of data provided by HSV, fully automatic segmentation of the glottal area is hypothesized to dramatically reduce work time and workload of clinicians and provide them with more objective information than currently available in the clinical routine. Gloger et al.[31] have proposed such a method. However, they have only applied it in a limited dataset, so transfer of their findings to other datasets was impeded. With deep learning algorithms growing and even becoming the methodology of choice for automated medical image analysis,[32] robust and fast HSV image processing methods can be anticipated. First attempts were already suggested to apply such deep learning methods toward image quality enhancements in HSV[33] and also estimating physiological laryngeal parameters (e.g. subglottal pressure) from HSV recordings in combination with numerical lumped-mass models.[34]

BIBLIOGRAPHY

1. Kendall KA, Leonard RJ. *Laryngeal Evaluation: Indirect Laryngoscopy to High-Speed Digital Imaging*. New York, NY: Thieme Medical Publishers; 2011.
2. Ma EPM, Yiu EML. *Handbook of Voice Assessments*. San Diego, CA: Plural Publishing; 2011.
3. Döllinger M. The next step in voice assessment: high-speed digital endoscopy and objective evaluation. *Curr Bioinform* 2009;4:101–111.
4. Petermann S, Kniesburges S, Ziethe A, Schützenberger A, Döllinger M. Evaluation of analytical modeling functions for the phonation onset process. *Comput Math Methods Med* 2016;2016:8469139.
5. Thornton FN, Berry DA, Kniesburges S, Döllinger M, Alexiou C, Schützenberger A. Impact of subharmonic and aperiodic laryngeal dynamics on the phonatory process analyzed in ex vivo rabbit models. *Appl Sci* 2019;9:1963.
6. Patel RR, Liu L, Galatsanos N, Bless DM. Differential vibratory characteristics of adductor spasmodic dysphonia and muscle tension dysphonia on high-speed digital imaging. *Ann Otol Rhinol Laryngol* 2011;120:21–32.
7. Döllinger M, Dubrovskiy D, Patel R. Spatiotemporal analysis of vocal fold vibrations between children and adults. *Laryngoscope* 2012;122: 2511–2518.
8. Bohr C, Kräck A, Eysholdt U, Ziethe A, Döllinger M. Quantitative analysis of organic vocal fold pathologies in females by high-speed endoscopy. *Laryngoscope* 2013;123:1686–1693.
9. Zacharias SRC, Deliyski DD, Gerlach TT. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *J Voice* 2018;32:216–220.
10. Echternach M, Burk F, Burdumy M, et al. The influence of vocal mass lesions on the passaggio region of professional singers. *Laryngoscope* 2017;127:1392–1401.
11. Bonilha HS, Desjardins M, Garand KL, Martin-Harris B. Parameters and scales used to assess and report findings from stroboscopy: a systematic review. *J Voice* 2018;32:734–755.
12. Brunings JW, Vanbelle S, Akkermans A, et al. Observer agreement for measurements in videolaryngostroboscopy. *J Voice* 2018;32:756–762.
13. 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.
14. Olthoff A, Woywod C, Kruse E. Stroboscopy versus high-speed glottography: a comparative study. *Laryngoscope* 2007;117:1123–1126.
15. Woo P. Objective measures of laryngeal imaging: what have we learned since Dr. Paul Moore. *J Voice* 2014;28:69–81.
16. Patel R, Dubrovskiy D, Döllinger M. Characterizing vibratory kinematics in children and adults with high-speed digital imaging. *J Speech Lang Hear Res* 2014;57:S674–S686.
17. Yamauchi A, Yokonishi H, Imagawa H, Sakakibara KI, Nito T, Tayama N. Quantitative analysis of vocal fold vibration in vocal fold paralysis with the use of high-speed digital imaging. *J Voice* 2016;30:766.e13–766.e22.
18. Kunduk M, Döllinger M, McWhorter A, Lohscheller J. Assessment of the variability of vocal fold dynamics with and between recordings with high-speed imaging and by phonovibrogram. *Laryngoscope* 2010;120:981–987.
19. Schützenberger A, Kunduk M, Döllinger M, et al. Laryngeal high-speed videoendoscopy: sensitivity of objective parameters towards recording frame rate. *Biomed Res Int* 2016;2016:4575437.
20. Schlegel P, Semmler M, Kunduk M, Döllinger M, Bohr C, Schützenberger A. Influence of analyzed sequence length on parameters in laryngeal high-speed videoendoscopy. *Appl Sci* 2018;8:2666.
21. Lohscheller J, Toy H, Rosanowski F, Eysholdt U, Döllinger M. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med Imag Anal* 2007;11:400–413.
22. Schlegel P, Kunduk M, Stingl M, et al. Influence of spatial camera resolution in high-speed videoendoscopy on laryngeal parameters. *PLoS One* 2019;14:e0215168.
23. Buder EH. Acoustic analysis of voice quality: a tabulation of algorithms 1902–1990. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement*. San Diego, CA: Singular Publishing; 2000:119–244.
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–163.
25. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall; 2000.
26. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice* 2006;20:527–544.
27. Misono S, Merati AL, Eadie TL. Developing auditory-perceptual judgment reliability in otolaryngology residents. *J Voice* 2012;26:358–364.
28. Brinca L, Batista AP, Tavares AI, Pinto PN, Araújo L. The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *J Voice* 2015;29:776.e7–e14.
29. Baijens L, Barikroo A, Pilz W. Intrarater and interrater reliability for measurements in videofluoroscopy of swallowing. *Eur J Radiol* 2013;82: 1683–1895.
30. Konradsen H, Trosborg I, Christensen L, Pedersen PU. Evaluation of interrater reliability assessing oral health in acute care settings. *Int J Nurs Pract* 2014;20:258–264.
31. Gloger O, Lehnert B, Schrade A, Völzke H. Fully automated glottis segmentation in endoscopic videos using local color and shape features of glottal regions. *IEEE Trans Biomed Eng* 2015;62:795–806.
32. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
33. Gomez P, Semmler M, Schützenberger A, Bohr C, Döllinger M. Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Med Biol Eng Comput* 2019;57:1451–1463.
34. Gomez P, Schützenberger A, Semmler M, Döllinger M. Laryngeal pressure estimation with a recurrent neural network. *IEEE J Transl Eng Health Med* 2018;7:2000111.