Research Article

Comparison of Videostroboscopy to Stroboscopy Derived From High-Speed Videoendoscopy for Evaluating Patients With Vocal Fold Mass Lesions

Maria E. Powell,^{a,b} Dimitar D. Deliyski,^c Robert E. Hillman,^{d,e,f} Steven M. Zeitels,^{d,e}
James A. Burns,^{d,e} and Daryush D. Mehta^{d,e,f}

Purpose: Videostroboscopy (VS) uses an indirect physiological signal to predict the phase of the vocal fold vibratory cycle for sampling. Simulated stroboscopy (SS) extracts the phase of the glottal cycle directly from the changing glottal area in the high-speed videoendoscopy (HSV) image sequence. The purpose of this study is to determine the reliability of SS relative to VS for clinical assessment of vocal fold vibratory function in patients with mass lesions.

Methods: VS and SS recordings were obtained from 28 patients with vocal fold mass lesions before and after phonosurgery and 17 controls who were vocally healthy. Two clinicians rated clinically relevant vocal fold vibratory features using both imaging techniques,

indicated their internal level of confidence in the accuracy of their ratings, and provided reasons for low or no confidence.

Results: SS had fewer asynchronous image sequences than VS. Vibratory outcomes were able to be computed for more patients using SS. In addition, raters demonstrated better interrater reliability and reported equal or higher levels of confidence using SS than VS.

Conclusion: Stroboscopic techniques on the basis of extracting the phase directly from the HSV image sequence are more reliable than acoustic-based VS. Findings suggest that SS derived from high-speed videoendoscopy is a promising improvement over current VS systems.

Laryngeal imaging is an invaluable component of the voice assessment protocol (Dejonckere et al., 2001). Information regarding vocal fold kinematics, which can only be obtained through laryngeal imaging, provides visual indicators of tissue health and function, which are critical for accurate diagnosis (Sataloff et al., 1988; Woo, Colton, Casper, & Brewer, 1991; Paul et al., 2013).

The clinically significant vibratory features obtained through laryngeal imaging are often used as outcome measures for determining the effectiveness of treatment (Behrman, 2005; Bonilha, Focht, & Martin-Harris, 2015). Therefore, imaging tools to assess vocal fold vibratory function are indispensable to the laryngologist or voice clinician.

The vocal folds vibrate too fast for human perception to appreciate without the aid of technology. There are two approaches for addressing this issue. The first approach is to capture the true vibratory function using high-speed imaging with significantly reduced playback rates that the human eye can appreciate (Farnsworth, 1940; Eysholdt, Tigges, Wittenberg, & Pröschel, 1996; Hertegård, Larsson, & Wittenberg, 2003). As long as capture rates are at least 4,000 frames per second (fps), this approach provides an accurate representation of the true vibratory cycle (Deliyski, Powell, Zacharias, Gerlach, & de Alarcon, 2015). This technology, although historically extremely valuable for increasing understanding of vocal fold vibratory function, has been limited to research labs.

The second approach is to take advantage of the quasi-periodic nature of vocal fold vibration (Titze, 1994)

^aDepartment of Communication Sciences and Disorders, University of Cincinnati, OH

^bLaryngeal Biology Laboratory, Vanderbilt University, Nashville, TN

^cDepartment of Communicative Sciences and Disorders, Michigan State University, East Lansing

^dCenter for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston

^eDepartment of Surgery, Harvard Medical School, Boston, MA

^fDepartment of Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA

Correspondence to Maria E. Powell: maria.e.powell@vanderbilt.edu

Editor: Krista Wilkinson

Associate Editor: Preeti Sivasankar

Received May 5, 2015

Revision received December 31, 2015

Accepted March 31, 2016

DOI: 10.1044/2016_AJSLP-15-0050

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

and only sample consecutive phases of vibration across multiple vibratory cycles using stroboscopic principles (described by Hillman & Mehta [2010]). Today, videostroboscopy (VS) systems use an indirect acoustic or electroglottographic signal to predict the phase for sampling. The resulting image sequences provide an estimate of vocal fold vibration in real time. This method has several clinical advantages. First, these VS systems can record long phonation samples, allowing clinicians to collect a full clinical protocol, including a variety of pitches and intensities, in a single recording. Second, data storage and retrieval procedures have been streamlined, providing immediate access to the recording for playback. Third, the real-time video can be played back with synchronous audio, which allows clinicians to refine judgments about the normality of the vibratory function (Mehta & Hillman, 2012). The practicality of VS as an alternative to high-speed photography for functionally evaluating vocal fold vibration made the technology much more accessible to clinicians (Yanagisawa, Casuccio, & Suzuki, 1981), which in turn facilitated the establishment of normative data (Hirano & Bless, 1993) and development of well-researched clinical protocols (Dejonckere et al., 2001). Since the early 1990s, VS has been the gold standard for evaluation with widespread clinical implementation.

Despite these well-documented advantages of VS, there are also well-known limitations in the sampling methodology that diminish its clinical value for evaluating patients with even moderately perturbed acoustic signals. VS relies on the acoustic signal from a contact microphone for frequency extraction, which is then used to predict in real time the next phase of the glottal cycle to be sampled. For VS, the phase delay between frames is 18 degrees of the fundamental frequency of the preceding acoustic cycles. Although frames are not sampled from consecutive cycles (i.e., multiple vibratory cycles are skipped between frames), when these frames are played back in sequence, these images present a slow-motion estimate of the underlying vibratory function. (See Hillman & Mehta [2010], for a more detailed discussion of the principles of stroboscopy.) If the acoustic signal is sufficiently perturbed, as is often the case for patients with moderate or severe dysphonia, then the fundamental frequency cannot be extracted, and the next phase of the glottal cycle cannot be accurately predicted, resulting in asynchronous image sequences that cannot be interpreted. Previous studies have reported that between 17% to 63% of patient recordings could not be assessed due to the inability of the strobe to synchronize to the fundamental frequency of the acoustic signal (Woo et al., 1991; Patel, Dailey, & Bless, 2008). The failure of VS to provide interpretable data for patients with voice disorders represents a significant clinical concern. Despite this well-documented limitation of VS, no major technological advances in the sampling methodology have been made since its clinical implementation.

Within the past decade, high-speed videoendoscopy (HSV) has gained renewed research interest, with an emphasis on addressing many of the methodological, technical, and practical challenges that have limited the implementation

of HSV in clinical settings. Technological advancements have made color imaging with improved spatial resolution available (as demonstrated in Mehta et al. [2012]). However, data storage and retrieval, as well as the slow process of data analysis, remain significant issues (Deliyski et al., 2008). Another practical limitation for clinics is the lack of cost justification for purchasing both a VS and high-speed system. Comparison studies report varied interpretations as to the clinical value of HSV apart from VS. Some conclude that HSV is purely supplemental to VS (Mendelsohn, Remacle, Courey, Gerhard, & Postma, 2013); others believe that if the technological challenges listed previously can be overcome, then it will supplant VS (Olthoff, Woywood, & Kruse, 2007); and still others envision systems that integrate both technologies into a single unit (Mehta & Hillman, 2012).

Stroboscopy derived from HSV was developed and first reported in 2005 and then briefly described in later publications (Deliyski, 2010; Deliyski et al., 2008; Deliyski, Shaw, Martin-Harris, & Gerlach, 2005). The technique is based on the same stroboscopic principles of VS, but rather than relying on the indirect acoustic or electroglottographic signal for phase sampling, simulated stroboscopy (SS) extracts the stroboscopic image sequence from the HSV recording after the fact by estimating the fundamental frequency of vibration directly from the changing glottal area over time. Recently, the validity of this technique was established on the basis of strong to very strong correlations for visual-perceptual ratings of vibratory function to the high-speed video from which the SS image sequence was derived (Powell, Deliyski, Mehta, & Hillman, 2015).

One impetus for developing SS was to address the high prevalence of asynchronicity, resulting from inappropriate phase selection, when using VS. Because VS relies on the fundamental frequency extraction of the acoustic signal to predict the next phase for sampling, VS is essentially a hybrid of acoustic analysis and laryngeal imaging. Thus, any breakdown in the analysis of the fundamental frequency of the acoustic signal would necessarily translate to a breakdown in meaningful phase selection for laryngeal imaging. Bonilha and Deliyski (2008) compared acoustic signals to the corresponding HSV image sequences and found that highly perturbed acoustic signals were not always associated with visible variations in glottal period; rather, local irregularities within the glottal cycle (such as irregularities in glottal width, phase symmetry, mucus, and so on) could be factors contributing to period perturbations of the acoustic signal (see Deliyski [2010] for further discussion). This finding suggests that reliance on the indirect acoustic signal may not be a reliable method for accurate phase sampling to synchronize to the underlying true vibratory function. In contrast, SS is derived from the superset of the full HSV recording on the basis of temporal analysis of the images. In theory, using the glottal area waveform to determine the next phase of the vibratory cycle for sampling may be a more reliable methodology for creating synchronous stroboscopic image sequences.

The purpose of this study was to investigate the reliability of SS relative to acoustic-based VS to evaluate clinically relevant ratings of vocal fold vibratory function. It is hypothesized that SS will provide more instances where the image sequence synchronizes to the underlying true vibratory cycle than acoustic-based VS, allowing for outcome measures to be computed in more patients and for more vibratory features than VS. It is subsequently hypothesized that raters will demonstrate better interrater reliability using SS than VS because the presence of asynchronous image sequences would likely affect raters' abilities to accurately and confidently assess clinically significant vibratory features.

Method

Participants

A total of 28 adults (15 men and 13 women aged 17 to 90 years) with vocal fold mass lesions recruited from the Massachusetts General Hospital Voice Center (Boston, MA) and 17 controls who were vocally healthy (7 men and 10 women aged 24 to 65 years) recruited from Charlotte Eye, Ear, Nose and Throat Associates (Charlotte, NC) participated in this study. All participants with voice disorders were recorded twice: once prior to microlaryngeal surgery to remove either unilateral or bilateral vocal fold mass lesions and then again during a follow-up evaluation approximately 3.5 weeks following surgery. Patients were evaluated using VS as part of the clinical protocol and then consented for the study if they met the inclusion criteria. After consent, they were recorded using HSV. Participants had time to rest between endoscopies and would not be expected to be overly affected by fatigue or discomfort during the HSV evaluation. The controls who were vocally healthy were recorded once using both imaging modalities in a single session—first with VS, followed by HSV. At the time of evaluation, control participants were assessed by a licensed speech language pathologist with expertise in voice and determined to be vocally healthy if they received a score within the normal limits on the Consensus Auditory-Perceptual Evaluation of Voice assessment scale (Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009) and presented with normal anatomy and vibratory function during endoscopy (Roy et al., 2013).

Instrumentation and Postprocessing

VS Recordings

VS examinations were performed using a KayPENTAX digital stroboscopy system (RLS 9100B, PENTAX Medical, Montvale, NJ) coupled to a handheld 70° transoral rigid endoscope (Model 9106, PENTAX Medical) and a 120 W xenon light source. The spatial resolution of the VS data was the National Television System Committee standard of 720 horizontal $\times$ 480 vertical pixels (Figure 1, left). All VS recordings were manually reviewed, and a representative 2- to 4-s segment of habitual pitch (excluding onset and offset of phonation) was selected from each recording.

Each VS sample yielded three to six vibratory cycles for evaluation.

HSV Recordings

Participants were instructed to produce a sustained phonation at their habitual pitch and loudness. HSV examinations were performed using a color high-speed video camera (Phantom v7.3, Vision Research, Inc., Wayne, NJ) that was suspended from a crane using a ball joint and gimbal frame. This configuration balanced the weight of the HSV camera while providing clinicians with a full range of motion to maneuver the camera around the axis of the endoscope for optimal visualization of the larynx. The camera lens was coupled to a 70° transoral rigid endoscope (JEDMED, St. Louis, MO) and a 300 W xenon light source containing three glass infrared filters for thermal energy reduction. The video sampling rate was set to 6250 fps for patients and 5512.5 fps for controls with maximum integration time. The spatial resolution for high-speed images was 320 horizontal $\times$ 352 vertical pixels (Figure 1, right). Due to data storage limitations, the full recording time for all HSV samples was 1.5 s.

SS Recordings Derived From HSV

The full 1.5-s HSV recordings were subjected to automated temporal segmentation that calculated the fundamental frequency of vocal fold vibration from the glottal area waveform (GAW) on the basis of the second central moment of intensity of the HSV image (Deliyski et al., 2008). The fundamental frequency of the GAW was then used to select the phase delay for image sampling from the HSV recording. For each SS frame, the phase delay was selected to be 18° of the GAW fundamental frequency of that frame. For frames where a fundamental frequency could not be defined, the phase delay was 0° compared with the previous SS frame. Preliminary recommendations from the American Speech-Language-Hearing Association Instrumental Voice Assessment Protocol ad hoc committee state that at least three vibratory cycles are needed to accurately assess vibratory features (see <http://www.asha.org/About/governance/committees/Active-Ad-Hoc-Committees>). However, the maximum length of the full HSV recordings was 1.5 s. If the typical VS sampling rate of 30 fps were applied, less than three vibratory cycles would be extracted from the 1.5-s HSV recording. To produce a comparable number of cycles in the SS recordings as the VS recordings, the sampling rate was increased to 80 fps. This increase in sampling rate produced approximately 120 frames, which, when played back at 30 fps, resulted in 4 s of SS data (consistent with VS recordings). These 4-s SS recordings were manually trimmed to 2 to 4 s to (a) exclude any onsets or offsets of phonation and (b) limit the number of vibratory cycles presented between three and six cycles—equal to those presented by VS. Although synchronous video with audio playback is a distinct advantage of VS and is possible with SS, the sampling rate of the SS recordings in the current study prevented synchronous audio playback.

Figure 1. Screen captured from videostroboscopy (left) and simulated stroboscopy (right) for a 40-year-old man with bilateral vocal fold polyps.

Therefore, both VS and SS recordings were presented without audio.

Visual-Perceptual Experiments

Raters

Nawka and Konerding (2012) reported that two to three raters were optimal to test reliability measures. In addition, they found that experience was a significant factor in reliability outcomes up to 2 years; after that, experience contributed negligible benefit in terms of reliability. To minimize variability associated with differences in clinical training (Deliyski et al., 2015), all raters in the current study were certified speech-language pathologists. Thus, three expert voice clinicians, each with between 2 and 10 years of experience rating VS in high-volume voice clinics, participated in this study.

Feature-Specific Scales

Raters conducted visual-perceptual ratings of mucosal wave, amplitude of vibration, phase asymmetry, vocal fold edge, and synchronicity using feature-specific scales. These scales were tailored to assess specific vibratory features common to patients with vocal fold mass lesions while providing adequate precision to allow for comparison of pre- and postsurgery ratings. Similar to the Stroboscopy Evaluation Rating Form (SERF; Poburka, 1999), amplitude and mucosal wave were rated in percentages of the width of each vocal fold; however, there was concern that a scale with 20 percentage-point (pp) increments may not be adequate precision to allow for comparison of pre- to postsurgery ratings. Therefore, visual analog scales were used, and the precision was increased to increments of 1% with anchors placed every 15 pp.

The SERF rates phase symmetry on the basis of the percentage of the whole recording during which the vocal fold was symmetrical. Given the limited number of cycles presented at playback, this feature was altered to describe the severity of the left-right phase asymmetry during the

three to six vibratory cycles using a visual analog scale. For vocal fold edge, the smoothness and straightness scales from the SERF were collapsed into one visual analog scale that had clear, anatomically based definitions specific to patients with vocal fold mass lesions for each level within the scale. As stated previously, asynchronicity of the image sequences may not necessarily indicate period irregularities or aperiodicity of the vibratory cycle. Therefore, in lieu of the terms *periodicity* or *regularity*, the term *asynchronicity* was used. Ratings for this feature were categorical: completely synchronous, intermittently asynchronous, or completely asynchronous. See Table 1 for more details regarding feature scales.

For each rating, participants were also instructed to indicate their level of confidence in their rating on a 4-point scale: (1) *very confident*, (2) *confident*, (3) *not confident*, and (4) *cannot rate*. If raters indicated low or no confidence, they were instructed to also indicate one or more reasons for their lack of confidence from the following list:

1. Dark image obscures feature
2. Image quality obscures feature
3. Asynchronicity obscures feature
4. Scope angle obscures/distorts feature
5. Anatomical structure obscures feature
6. Mucus obscures feature
7. Multiple vibratory segments complicate rating
8. Other: Please explain

Consensus Training

The three experienced voice clinicians participated in approximately 3 hours of consensus training prior to executing the experimental ratings. This training provided raters with the opportunity to familiarize themselves with using the scales and form a consensus as to the thresholds for each level of the feature-specific scale on the basis of objective, anatomical markers. At least six VS and six SS

Table 1. Vibratory features were rated on the basis of feature definitions and scale levels.

Feature	Definition	Rating scale	Range
Mucosal wave	Vocal fold deformation along the lateral plane during the opening phase	75-point visual analog scale in % with six anchoring points Absent 20% 35% 50% 65% ≥80%	15 percentage points
Amplitude	Maximum lateral excursion of each vocal fold from most medial position	75-point visual analog scale in % with six anchoring points Absent 20% 35% 50% 65% ≥80%	15 percentage points
Left-Right phase asymmetry	Phase difference between left and right vocal fold vibration	100-point visual analog scale with four anchoring points Absent: Left and right vocal folds reach both maximal amplitude and the midline at the same time. Mild: One vocal fold lags slightly behind the other either during the lateral-to-medial or medial-to-lateral transition. Moderate: One vocal fold has already transitioned and begun to move medially before the other vocal fold has hit maximal amplitude. Severe: One vocal fold has hit midline while the other vocal fold is at maximal amplitude.	33 points
Synchronicity	Continuity of the images within the simulated cycle	3-point scale Completely synchronous: Continuous movement showing the expected phases of vocal fold vibration is present for the entire recording. Intermittently asynchronous: Continuous movement is appreciable at times, but there are one or more instances of asynchronicity noted. Completely asynchronous: Continuous movement is not appreciable.	1 point
Vocal fold edge	Irregularity of the vibrating vocal fold edges	100-point visual analog scale with four anchoring points Regular: Vocal fold edge is straight with a sharp superior edge. Mildly irregular: Vocal fold edge may be rounded or a mild swelling may be present. A sulcus may be present on the medial edge of the fold. Glottal closure may still be complete or a small anterior or posterior gap may be present. Moderately irregular: Vocal fold edge is clearly nonlinear, but does not impede the vibratory function of the contralateral fold. Glottal closure may form an hourglass configuration. Severely irregular: Vocal fold edge is nonlinear with a large protrusion or multiple areas of irregularity. The lesion may impede the vibratory function of the contralateral fold.	33 points

recordings were discussed for each vibratory feature. Samples used to form a consensus were taken from the pool of recordings that were collected for but unused in the experimental portion of the current study. Participants were asked to rate the feature individually, and then the ratings were discussed collectively to form a consensus. Once a consensus was formed, representative images of each recording along with the agreed-on rating for each vibratory feature were compiled into an experimental workbook. This experimental workbook as well as the recordings used for consensus training were available to raters as reference throughout the duration of the study. Raters

were encouraged to calibrate their ratings to the group using these references prior to beginning a new experimental set.

Experimental Ratings

VS and SS recordings were rated for 28 patients (28 presurgery recordings, 28 postsurgery recordings) and 17 controls for a total of 73 recordings. These recordings were deidentified, randomized, and compiled into two experimental sets that were used to evaluate each vibratory parameter. A 10% redundancy was added (eight recordings), resulting in 81 deidentified and randomized recordings for

each experimental set. Experiments were organized in a randomized block design. Raters were instructed to rate all vibratory features within a single experimental set before beginning the next experimental set to maintain consistency.

VS and SS images differed in spatial resolution (Figure 1); therefore, to control for the size of the vocal folds, raters were instructed to zoom the VS images by 200% and the SS images by 300% during playback. For this reason, raters were not blinded to which imaging technique they were rating; however, they were blinded to whether the recording was presurgery, postsurgery, or control. Both VS and SS recordings were defaulted to a playback rate of 30 fps and were looped for continuous playback; however, raters were able to slow down the recordings or advance frame by frame as they wished. Using custom-designed software (Alvin2; Hillenbrand & Gayvert, 2005), raters were asked to move a tick mark along a scroll bar to indicate their rating on the visual analog scale (e.g., see Figure 2). Backtracking was permitted. Raters were encouraged to take breaks as necessary to minimize fatigue.

Statistical Analysis

Intra- and interrater agreement were calculated as raw percent agreements, where ratings were considered in agreement if they fell within $\pm \frac{1}{2}$ the range of one level of the feature-specific scale. (See Figure 2 for examples.) Interrater reliability was calculated using Spearman's correlation coefficient and the intraclass correlation coefficient (ICC). A threshold of reliability was established a priori for each vibratory parameter on the basis of percent agreement between raters. Vibratory features that did not achieve at least 60% direct agreement between raters for both imaging

techniques would be automatically excluded from additional analysis. Group means were compared using a linear mixed-effects model. The proportions of synchronicity between imaging techniques were analyzed using chi-square analysis, and the proportions of the raters' levels of confidence across techniques were analyzed using chi-square analysis or the Fisher exact test for samples with cell frequency counts less than 5.

Results

One rater was unable to complete the protocol for both VS and SS samples due to the time commitment required to complete the protocols; therefore, her ratings were excluded from the analysis. This resulted in a total of 2,592 ratings being included for analysis: eight vibratory features rated for 81 recordings by two raters and for two different imaging techniques. Within this sample, 256 ratings were used for the purpose of evaluating reliability only. Rater 1 completed the VS ratings in approximately 6.5 hours and the SS ratings in approximately 6.0 hours. Rater 2 completed the VS ratings in approximately 9.1 hours and the SS ratings in approximately 7.8 hours.

Of the 1,296 VS observations, 153 (12%) were unable to be rated for at least one vibratory feature. Of the 1,296 SS observations, 51 (4%) were unable to be rated for at least one vibratory feature. Table 2 shows the breakdown of observations that could not be rated for each vibratory feature.

Methodological Reliability

Asynchronicity, which precluded ratings of vibratory features, was present in more VS than SS recordings. Chi-square analysis indicated statistically significant differences

Figure 2. Custom-designed software for data collection. Raters used the visual analog scale (left) to indicate the maximum amplitude of vibration. They also indicated their level of internal confidence in the accuracy of their rating using radio buttons (right) and described their reasons for low confidence using the text boxes. A priori established range for agreement: $\pm \frac{1}{2}$ scalar level. The ranges of error overlap in Observation 1, therefore, would be considered in agreement. In Observation 2, the ranges of error do not overlap. Therefore, these ratings would not be considered in agreement. Note Rater 2's lack of confidence with Reason 4 (scope angle obscures/distorts feature) indicated.

Table 2. Number of recordings that were unable to be rated for each vibratory feature.

Vibratory feature	VS, <i>n</i> = 1,296		SS, <i>n</i> = 1,296	
	Count	%	Count	%
Left mucosal wave	32	2.5	10	0.8
Right mucosal wave	31	2.4	9	0.7
Left amplitude	13	1.0	4	0.3
Right amplitude	9	0.7	1	0.1
Left-Right phase asymmetry	43	3.3	17	1.3
Synchronicity	2	0.2	2	0.2
Left Vocal fold edge	12	0.9	4	0.3
Right Vocal fold edge	11	0.9	4	0.3
Total	153	11.9	51	4.0

Note. VS = videostroboscopy; SS = simulated stroboscopy.

between VS and SS for ratings of synchronicity. For VS, 48% of observations were rated as completely synchronous, 34% were rated intermittently asynchronous, and 16% were rated completely asynchronous. For SS, 77% of observations were rated as completely synchronous, 15% were rated intermittently asynchronous, and 6% were rated completely asynchronous (Table 3). One patient recording could not be rated for synchronicity using both VS and SS due to a lack of visible vocal fold vibration.

Rater Reliability

Analysis of rater agreement and reliability excluded any observations that could not be rated. Therefore, sample sizes varied between imaging techniques (VS and SS) for each vibratory feature.

Rater Agreement

Intrarater agreement within $\pm 1/2$ level of the scale for SS was near perfect, ranging from 88% to 100% (average 98%) for Rater 1, and 100% agreement for Rater 2. Intrarater agreement for VS was slightly lower, ranging from 50% to 100% (average 85%) for Rater 1 and 83% to 100% (average 98%) for Rater 2. Table 4 details the feature-specific intrarater agreement results.

Table 3. Number of recordings rated as completely synchronous, intermittently asynchronous, or completely asynchronous for each imaging technique.

Level of synchronicity	VS		SS		<i>p</i> value
	Count	%	Count	%	
Completely synchronous	70	48	113	77	<.0001*
Intermittently asynchronous	50	34	33	15	<.0001*
Completely asynchronous	24	16	9	6	<.0001*
Cannot rate	2	1	2	1	—

Note. Ratings from both raters are combined (73 recordings rated by two raters) for the videostroboscopy (VS) and simulated stroboscopy (SS) samples.

*Findings from chi-square analysis are considered statistically significant at $p < .05$.

Interrater agreement using SS (Table 5) was also strong to near perfect for all parameters, ranging from 81% to 99%. Percent agreement for VS was also strong to near perfect for all parameters except mucosal wave, resulting in 61% and 60% for left and right mucosal wave, respectively.

Rater Reliability

For interrater reliability (Table 5), Spearman correlations were strong to very strong, ranging from 0.70 to 0.86 for VS and 0.73 to 0.89 for SS. ICC coefficients for both imaging techniques were significantly lower than the Spearman correlations, such that the upper end of the ICC 95% confidence interval fell below the Spearman coefficient for the same vibratory feature. Left-right phase asymmetry showed the lowest level of reliability for both imaging techniques on the basis of the Spearman coefficients and ICCs.

Rater Confidence

Initial analysis of confidence ratings for amplitude, mucosal wave, asymmetry, and edge showed very few instances of “very confident” from either rater; Rater 1 indicated “very confident” in 0.8% of observations (8/1,022), and Rater 2 used “very confident” in 0.4% of observations (4/1,022). Thus, the rating categories of “very confident” and “confident” were collapsed into one. Table 6 shows the results of chi-square analysis or Fisher’s exact test for each vibratory feature. Rater 1 showed statistically significant differences for left and right amplitude, left-right phase asymmetry, and right vocal fold edge. Differences in Rater 2’s levels of confidence were not found to be statistically significant across imaging techniques for any vibratory feature.

Further investigation into the sources of low confidence or inability to rate vibratory features altogether are detailed in Table 7. Percentages were calculated as a proportion of the total number of recordings that showed either low confidence or could not be rated for that feature. For VS, the majority of observations that resulted in low or no rater confidence for mucosal wave, amplitude, and left-right phase asymmetry were attributed to the presence of asynchronous image sequences. For vocal fold edge, low confidence was most often attributed to poor image quality (43%), and no confidence was attributed to asynchronicity (53%).

For SS, reasons for low or no rater confidence were more varied. For mucosal wave, low rater confidence was most often attributed to poor image quality (36%) or anatomical structures that obscured the feature (37%), whereas inability to rate the feature was attributed to asynchronicity (63%). Reasons for low or no confidence using SS for amplitude and left-right phase asymmetry ratings were typically attributed to scope angle (32%) or asynchronicity (24%). Low rater confidence for vocal fold edge was most frequently attributed to image quality (43%), and no confidence was attributed to the placement of other anatomical structures that prevented visualization of the vocal fold edge (38%).

Table 4. Intrarater agreement was calculated as percentage of agreement within $\pm \frac{1}{2}$ the range of one level of the scale (e.g., ratings of mucosal wave must have been within 15 percentage points [pp] of each other [the range of one level of the scale] to be considered in agreement).

Vibratory feature	Range	Rater	VS		SS	
			<i>n</i> agree	% agree	<i>n</i> agree	% agree
Left mucosal wave	15 pp	Rater 1	3/4	75	8/8	100
		Rater 2	6/6	100	8/8	100
Right mucosal wave	15 pp	Rater 1	2/4	50	7/8	88
		Rater 2	5/6	83	8/8	100
Left amplitude	15 pp	Rater 1	6/6	100	8/8	100
		Rater 2	7/7	100	8/8	100
Right amplitude	15 pp	Rater 1	4/4	100	8/8	100
		Rater 2	8/8	100	8/8	100
Left-Right phase asymmetry	33 pts	Rater 1	2/3	67	8/8	100
		Rater 2	3/3	100	8/8	100
Left vocal fold edge	33 pts	Rater 1	7/7	100	8/8	100
		Rater 2	7/7	100	8/8	100
Right vocal fold edge	33 pts	Rater 1	7/7	100	8/8	100
		Rater 2	7/7	100	8/8	100

Note. VS = videostroboscopy; SS = simulated stroboscopy.

Outcome Measures

Some recordings that had intermittent asynchronicity could still be rated for certain vibratory features; therefore, additional analysis was conducted to compare the number of patients (of 28) who could not be rated either presurgery or postsurgery specifically due to the presence of asynchronous image sequences (image sequences that were unable to synchronize to the underlying vibratory cycle). Results varied based on the feature being evaluated (Table 8). Left-right phase asymmetry showed the greatest percentage of patients for which a change in function (presurgery to

postsurgery) could not be computed, at 43% and 29% for VS and SS, respectively. Vocal fold edge showed the smallest percentage of patients for which presurgery and/or postsurgery function could not be computed due to asynchronicity: 7% and 11% for ratings of left and right vocal fold edge, respectively, using VS. Vocal fold edge could be rated for all patients using SS.

For recordings that could be rated, the difference in group means were reported for both imaging techniques for mucosal wave, amplitude, and vocal fold edge (Table 9). Left-right phase asymmetry was excluded from analysis due to low rater reliability judging this feature. Using VS, controls

Table 5. Interrater agreement and reliability using videostroboscopy and simulated stroboscopy.

Vibratory feature	% Agreement			Spearman <i>p</i>	Intraclass correlation	
	<i>n</i> agree	% agree	Range		ICC	95% CI
Videostroboscopy						
Left mucosal wave	36/59	61	15 pts	0.72*	0.31	0.10–0.49
Right mucosal wave	36/60	60	15 pts	0.84*	0.55	0.37–0.68
Left amplitude	70/73	96	15 pts	0.84*	0.57	0.41–0.69
Right amplitude	69/72	96	15 pts	0.86*	0.64	0.50–0.74
Left-Right phase asymmetry	48/56	86	33 pts	0.70*	0.33	0.10–0.51
Left vocal fold edge	70/75	93	33 pts	0.75*	0.58	0.42–0.71
Right vocal fold edge	69/75	92	33 pts	0.72*	0.51	0.34–0.64
Simulated stroboscopy						
Left mucosal wave	60/74	81	15 pts	0.89*	0.73	0.61–0.81
Right mucosal wave	64/73	88	15 pts	0.89*	0.79	0.69–0.85
Left amplitude	76/77	99	15 pts	0.80*	0.51	0.34–0.64
Right amplitude	78/80	98	15 pts	0.79*	0.62	0.48–0.72
Left-Right phase asymmetry	58/70	83	33 pts	0.73*	0.35	0.12–0.50
Left vocal fold edge	78/79	99	33 pts	0.79*	0.55	0.39–0.67
Right vocal fold edge	76/78	97	33 pts	0.86*	0.52	0.35–0.65

Note. Agreement is reported as the number and percentage of direct agreement within the range of one level of the feature-specific scale. Reliability is reported by Spearman correlations, which are statistically significant at $p < .0001$ (*), and the intraclass correlation (ICC) as well as the 95% confidence interval (CI) for the ICC. pts = points.

Table 6. Comparison of each rater's level of confidence for each vibratory feature by imaging technique.

Vibratory feature	Level of confidence	Rater 1		Rater 2	
		VS, <i>n</i> = 73	SS, <i>n</i> = 73	VS, <i>n</i> = 73	SS, <i>n</i> = 73
		% (count)	% (count)	% (count)	% (count)
Left mucosal wave	Confident	45 (33)	55 (40)	44 (32)	52 (38)
	Not confident	32 (23)	36 (26)	44 (32)	44 (32)
	Cannot rate	23 (17)	10 (7)	12 (9)	4 (3)
	Chi-square adj <i>p</i>	<i>p</i> = .0761		<i>p</i> = .1690	
Right mucosal wave	Confident	48 (35)	62 (45)	53 (39)	63 (46)
	Not confident	32 (23)	30 (22)	33 (24)	33 (24)
	Cannot rate	21 (15)	8 (6)	14 (10)	4 (3)
	Chi-square adj <i>p</i>	<i>p</i> = .0721		<i>p</i> = .1026	
Left amplitude	Confident	58 (42)	81 (60)	52 (38)	63 (46)
	Not Confident	34 (25)	18 (13)	41 (30)	33 (24)
	Cannot Rate	8 (6)	1 (1)	7 (5)	4 (3)
	Fisher exact adj <i>p</i>	<i>p</i> = .0038*		<i>p</i> = .4688	
Right amplitude	Confident	56 (41)	79 (58)	56 (41)	74 (54)
	Not Confident	36 (26)	19 (14)	38 (28)	23 (17)
	Cannot Rate	8 (6)	1 (1)	5 (4)	3 (2)
	Fisher exact adj <i>p</i>	<i>p</i> = .0049*		<i>p</i> = .0780	
Left-Right phase asymmetry	Confident	42 (31)	79 (58)	44 (32)	63 (46)
	Not confident	32 (23)	10 (7)	36 (26)	25 (18)
	Cannot rate	26 (19)	11 (8)	21 (15)	12 (9)
	Chi square adj <i>p</i>	<i>p</i> < .0001*		<i>p</i> = .0638	
Left vocal fold edge	Confident	60 (44)	77 (56)	71 (52)	81 (59)
	Not confident	33 (24)	21 (15)	22 (16)	16 (12)
	Cannot rate	7 (5)	3 (2)	7 (5)	3 (2)
	Fisher exact adj <i>p</i>	<i>p</i> = .0907		<i>p</i> = .3232	
Right vocal fold edge	Confident	56 (41)	78 (57)	67 (49)	79 (58)
	Not confident	38 (28)	19 (14)	26 (19)	18 (13)
	Cannot rate	5 (4)	3 (2)	7 (5)	3 (2)
	Fisher exact adj <i>p</i>	<i>p</i> = .0165*		<i>p</i> = .2285	

Note. Samples with more than five observations in each cell were analyzed using chi-square analysis, and samples with fewer than five observations in each cell were analyzed using Fisher's exact test. VS = videostroboscopy; SS = simulated stroboscopy; adj = adjusted.

*Findings are considered statistically significant at *p* < .05.

were statistically different from patients (both pre- and post-surgical assessments) for all vibratory features except left mucosal wave. Presurgery patients were statistically different from postsurgical VS assessment for left vocal fold edge only. Using SS, controls were statistically different from presurgery patients for all vibratory features. Presurgery patients were statistically different from postsurgical SS assessment for every vibratory feature except right mucosal wave. Controls were statistically different from postsurgery patients for all SS vibratory features except left and right amplitude.

Discussion

This study sought to determine whether SS improved on acoustic-based VS in terms of methodological reliability, rater reliability, rater confidence, and ability to compute outcome measures on the basis of visual-perceptual ratings.

Methodological Reliability

The limitation of VS to produce synchronous image sequences for some patients with voice disorders is well

known (Patel et al., 2008; Deliyski, 2010). In the current study, SS was able to produce more synchronous image sequences than VS. Of the recordings, 77% were rated as completely synchronous using SS compared with 48% using VS. In addition, only 6% of the recordings were rated as completely asynchronous using SS compared with 16% of VS recordings (Table 3). These findings suggest that SS may be more methodologically reliable than VS for extracting synchronous image sequences; however, it is important to note that, in the current study, the sampling methodology of SS differed from VS in three ways.

First, the signal used to determine the phase of the glottal cycle for sampling differed. VS used the fundamental frequency of the acoustic signal captured by a contact microphone at the participant's neck, whereas SS used the fundamental frequency calculated from the glottal area waveform. Current findings support the theory that extracting the phase of the glottal cycle directly from the HSV image is a more physiologically relevant sampling methodology and may be less sensitive to intracycle perturbations that may affect the periodicity of the acoustic signal (Bonilha & Deliyski, 2008).

Table 7. Reasons for low confidence or inability to rate vibratory features are categorized based on whether they are technology-based reasons or anatomy/physiology-based reasons. Percentages were calculated out of the total number for a particular feature that showed low or no confidence. For example: using VS, raters indicated they were “Not Confident” for 102 ratings, 9 of which (or 9%) were attributed to “dark image obscures feature”. Raters were allowed to indicate more than one reason for low or no confidence. Therefore, percentages may add up to greater than 100%.

Reason for Low Rater Confidence	Rater “Not Confident”				Rater “Cannot Rate”			
	VS		SS		VS		SS	
	Count	%	Count	%	Count	%	Count	%
Technology-Based Reasons								
Mucosal Wave								
Dark image obscures feature	9/102	9	8/104	8	0/51	0	0/19	0
Image quality obscures feature	24/102	24	37/104	36	2/51	4	2/19	11
Asynchronicity obscures feature	44/102	43	9/104	9	46/51	90	12/19	63
Scope angle obscures/distorts feature	7/102	7	7/104	7	1/51	2	1/19	5
Amplitude								
Dark image obscures feature	0/109	0	0/68	0	0/21	0	0/7	0
Image quality obscures feature	8/109	7	4/68	6	0/21	0	0/7	0
Asynchronicity obscures feature	56/109	51	21/68	31	15/21	71	4/7	57
Scope angle obscures/distorts feature	30/109	28	7/68	10	2/21	10	0/7	0
Left-Right Phase Asymmetry								
Dark image obscures feature	1/49	2	0/25	0	0/34	0	0/17	0
Image quality obscures feature	0/49	0	2/25	8	0/34	0	1/17	6
Asynchronicity obscures feature	31/49	63	6/25	24	30/34	88	13/17	76
Scope angle obscures/distorts feature	22/49	45	8/25	32	2/34	6	0/17	0
Vocal Fold Edge								
Dark image obscures feature	0/87	0	2/54	4	0/19	0	0/8	0
Image quality obscures feature	37/87	43	23/54	43	0/19	0	0/8	0
Asynchronicity obscures feature	27/87	31	4/54	7	10/19	53	0/8	0
Scope angle obscures/distorts feature	7/87	8	6/54	11	1/19	5	0/8	0
Anatomy/Physiology-Based Reasons								
Mucosal Wave								
Anatomical structure obscures feature	19/102	19	38/104	37	8/51	16	5/19	26
Mucus	10/102	10	17/104	16	0/51	0	0/19	0
Multiple vibratory segments complicate rating	0/102	0	2/104	2	0/51	0	2/19	11
Other	15/102	15	13/104	13	5/51	10	3/19	16
Amplitude								
Anatomical structure obscures feature	21/109	19	20/68	29	7/21	33	2/7	29
Mucus	2/109	2	10/68	15	0/21	0	0/7	0
Multiple vibratory segments complicate rating	0/109	0	0/68	0	0/21	0	0/7	0
Other	10/109	9	12/68	18	3/21	14	1/7	14
Left-Right Phase Asymmetry								
Anatomical structure obscures feature	4/49	8	0/25	0	5/34	15	4/17	24
Mucus	0/49	0	2/25	8	0/34	0	0/17	0
Multiple vibratory segments complicate rating	0/49	0	3/25	12	1/34	3	2/17	12
Other	0/49	0	7/25	28	5/34	15	4/17	24
Vocal Fold Edge								
Anatomical structure obscures feature	18/87	21	2/54	4	8/19	42	3/8	38
Mucus	12/87	14	14/54	26	0/19	0	0/8	0
Multiple vibratory segments complicate rating	0/87	0	2/54	4	0/19	0	0/8	0
Other	1/87	1	13/54	24	5/19	26	5/8	63

Second, the sampling rates for VS and SS differed. Although both techniques used an 18° phase delay, VS was sampled at 30 fps and SS was sampled at 80 fps. The difference in sampling rates was necessary to provide the same number of vibratory cycles during playback as VS from only 1.5 s of HSV data. However, the higher sampling rate of SS may inherently produce more synchronous image sequences because fewer vibratory cycles are skipped between frames when compared with VS. Future studies should compare VS and SS data collected at the same sampling rate.

Third, VS sampling is predicted in real time, whereas SS sampling is extracted from the full HSV recording on

the basis of image analysis during postprocessing. Because SS is created after the fact, it allows for the use of more complex algorithms. In particular, the phase of the SS cycle is identified on the basis of information from current, preceding, and following vibratory cycles, whereas the phase of the VS cycle is predicted on the basis of preceding cycles alone. This offline processing of SS has a second advantage over VS in that, if the acoustic signal is perturbed such that the VS system is unable to predict the correct phase for sampling, interpretable information is not collected. In contrast, SS is derived from the captured HSV recording; therefore, even if SS sampling returns an asynchronous

Table 8. Number of patients (of 28 possible) and percentage for each imaging technique where change in function due to surgical intervention was not able to be rated due to asynchronous image sequences that prevented visualization of vibratory features during either the presurgery or postsurgery evaluations.

Vibratory feature	VS, <i>n</i> = 28		SS, <i>n</i> = 28	
	Count	%	Count	%
Left mucosal wave	10	36	6	21
Right mucosal wave	10	36	4	14
Left amplitude	4	14	2	7
Right amplitude	5	18	2	7
Left-Right phase asymmetry	12	43	8	29
Left vocal fold edge	2	7	0	0
Right vocal fold edge	3	11	0	0

Note. VS = videostroboscopy; SS = simulated stroboscopy.

image sequence, the true vibratory cycle can be analyzed from the underlying HSV recording. Further research is needed to determine the relative contributions of these three methodological differences to the improvement of SS over VS.

Rater Reliability

Only two raters completed the entire protocol. Although Nawka and Konerding (2012) report that adequate reliability measures can be obtained from two raters, increasing the number of raters in future studies would improve the

generalizability of findings. In the current study, VS and SS show similar levels of interrater reliability for amplitude, left-right phase asymmetry, and vocal fold edge; however, SS demonstrated significantly higher levels of interrater reliability than VS for mucosal wave, as determined by the fact that there was no overlap in the ICC 95% confidence intervals for both left and right mucosal wave between techniques. ICC for left-right phase asymmetry was 0.33 for VS and 0.35 for SS, suggesting that left-right phase asymmetry is not reliably rated using stroboscopy, regardless of the signal used to determine phase selection. These findings are in accordance with other reports of low reliability for rating asymmetry using stroboscopic techniques (Bonilha, Deliyski, & Gerlach, 2008; Kelley, Colton, Casper Paseman, & Brewer, 2011; Kendall, 2009; Nawka & Konerding, 2012). Given the prevalence of left-right phase asymmetries in individuals who are vocally healthy (Bonilha et al., 2008) and the fact that the magnitude of the asymmetry appears to be diagnostically significant for individuals with voice disorders (Bonilha, Deliyski, Whiteside, & Gerlach, 2012), it is important that this feature be reliably rated. Findings from the current study, as well as reports from Bonilha et al. (2008) and Kendall (2009), which acknowledge VS's sensitivity to vocal fold asymmetries but question whether asymmetry can be specifically differentiated from other vibratory irregularities, suggest that stroboscopic technologies are inappropriate for the evaluation of left-right phase asymmetry. Instead, vocal fold asymmetry should be rated using high-speed imaging techniques (Nawka & Konerding, 2012).

Table 9. Statistical comparison of differences in group means for each vibratory feature using videostroboscopy and simulated stroboscopy. Differences are statistically significant when *p* are less than .05 (*).

Group comparisons	Videostroboscopy			Simulated stroboscopy		
	Difference in means	SEM	<i>p</i>	Difference in means	SEM	<i>p</i>
Left mucosal wave						
Control vs. pre-surgery	15	7	.213	34	6	<.001*
Pre- vs. post-surgery	-4	5	.962	-13	4	.007*
Post-surgery vs. control	11	7	.565	21	6	.019*
Right mucosal wave						
Control vs. pre-surgery	23	7	.007*	39	6	<.001*
Pre- vs. post-surgery	-4	4	.927	-8	3	<.001*
Post-surgery vs. control	19	7	.049*	31	6	.103
Left amplitude						
Control vs. pre-surgery	12	2	<.001*	12	2	<.001*
Pre- vs. post-surgery	-5	2	.075	-7	2	<.001*
Post-surgery vs. control	8	2	.027*	5	2	.298
Right amplitude						
Control vs. pre-surgery	13	2	<.001*	10	2	<.001*
Pre- vs. post-surgery	-3	1	.151	-4	1	.002*
Post-surgery vs. control	9	2	<.001*	5	2	.101
Left vocal fold edge						
Control vs. pre-surgery	-36	5	<.001*	-32	4	<.001*
Pre- vs. post-surgery	10	3	.026*	13	3	<.001*
Post-surgery vs. control	-26	5	<.001*	-19	4	<.001*
Right vocal fold edge						
Control vs. pre-surgery	-3	5	<.001*	-30	4	<.001*
Pre- vs. post-surgery	6	3	.408	13	2	<.001*
Post-surgery vs. control	-25	5	<.001*	-18	4	<.001*

Rater Confidence

In general, raters indicated equal or higher levels of confidence in the accuracy of their ratings when using SS for every vibratory feature than when using acoustic-based VS (Table 6). In addition, VS resulted in more instances of vibratory features that could not be rated for every vibratory feature than SS. Although confidence ratings were not statistically significant for Rater 2, Rater 1 did show statistically significant differences in confidence using SS compared with VS for amplitude, left-right phase asymmetry, and right vocal fold edge. Although internal perceptions of confidence are highly subjective and findings are likely to vary depending on the individual, these initial results suggest that clinicians may easily adapt to the new technology with as much confidence in their clinical ratings as they had with VS. These findings should be considered preliminary because data were compiled from only two raters.

Reasons for poor rater confidence can be divided into two general categories: either technology-based challenges (dark image, image quality, asynchronicity, or scope angle obscures or distorts the feature) or anatomy- or physiology-based challenges (anatomical structure or mucus obscures feature, multiple vibratory segments complicate rating, or other). Raters more frequently cited technology-based reasons for decreased confidence or inability to rate when using VS compared with SS (see Table 7). Both VS and SS ratings were affected by the presence of asynchronous image sequences; however, VS ratings were affected both in greater proportions and greater numbers for every vibratory feature. SS ratings were equally or more affected by image quality than VS ratings, particularly for mucosal wave and left-right phase asymmetry. Because SS is derived from the underlying HSV sample, the spatial resolution and sample duration depend on the system's memory resources (Deliyski, 2010) and are currently lower than VS capabilities. In the current study, VS had a spatial resolution of 720×480 pixels compared with SS's spatial resolution of 320×352 pixels. As high-speed technology advances, it is possible that HSV and SS spatial resolution could equal that of current VS systems.

Another source for low rater reliability was scope angle, which was present using both techniques but in greater proportions using VS. A likely reason for this difference may be due to the fact that the VS camera was handheld, whereas the high-speed camera that captured the data for SS was suspended from a crane. It has been suggested that suspending the camera from a crane using a ball joint allows for the full range of motion for clinicians to be able to maneuver, but reduces endoscopic motion compensation and tilt (Deliyski et al., 2008); however, findings from the current study indicate that it may also reduce the distortion of clinical features due to scope angle. Further investigation is needed to determine the degree to which clinical ratings are affected by the angle of the scope because this would affect automated objective measures of vibratory function as well.

Anatomy- or physiology-based reasons for decreased confidence or inability to rate were more frequently cited

when using SS compared with VS. In particular, anatomical obstructions and the presence of mucus, which reduced rater confidence, were more prevalent in SS recordings than VS. Two factors may have contributed to this finding. When using VS, clinicians are able to (a) see the simulated slow-motion vibrations in real time and (b) record long phonation samples. These advantages provide clinicians with greater flexibility in terms of identifying issues and troubleshooting scope placement for optimal viewing. Clinicians can also see mucus pooling during vibration and have patients clear their throat to displace it. In the current study, when recordings were collected using HSV, clinicians were not able to see simulated slow-motion vibrations in real time, nor were they able to record extended phonation times; therefore, they were unable to troubleshoot scope placement to the same extent as when using VS. HSV clinical protocols will need to be developed to ensure optimal scope placement prior to recording HSV samples to minimize the presence of anatomical obstructions that preclude visualization of vibratory features such as amplitude or mucosal wave, which may be critical for accurate diagnosis. In the current study, the nonoptimized algorithm took approximately 3 minutes to derive one SS image sequence from 1.5 s of HSV data; thus, SS could be an effective means for previewing the recorded HSV sample prior to saving to ensure the best visualization of vocal fold vibratory function possible.

The presence of multiple vibratory segments, which complicate the rating, was listed once for VS samples compared with 11 times for SS samples. This finding highlights the value of using the GAW to determine the phase for sampling rather than the acoustic waveform. Vocal folds with multiple vibrating segments would likely produce highly perturbed acoustic signals. It is reasonable to assume that these samples would not be able to be rated using VS due to the presence of asynchronicity; however, nine of these samples could be rated using SS, albeit with reduced confidence (see Table 7).

Outcome Measures

For patients with mass lesions who undergo phonosurgery, change in function as a result of surgery can only be determined if presurgery and postsurgery ratings were obtained. Raters were able to compute outcomes in 96% of the samples using SS compared with 88% of VS samples. For both imaging techniques, amplitude and vocal fold edge were least affected by the presence of asynchronous image sequences, whereas left-right phase asymmetry and mucosal wave were the most affected (Table 8). Kelley et al. (2011) found redundancy in videostroboscopic vibratory parameters and suggested that vocal fold edge and amplitude of vibration provide the greatest information regarding vocal fold vibratory function. The current study suggests that these two parameters are also less affected by the presence of asynchronous image sequences than other vibratory features, further underscoring their value as clinically salient vibratory features.

A comparison of controls and pre- and postsurgery group means suggest that raters were able to see greater change in vibratory function following phonomicrosurgery using SS. In Table 9, statistically significant differences between vocally healthy controls and presurgery patients indicate that the technique was able to differentiate between patients and controls. The statistically significant differences between patients presurgery to postsurgery imply that postsurgery patients were functionally different from presurgery patients. The statistically significant differences between postsurgery patients and controls imply that following surgery, patients were still distinguishable from the controls. Using VS, raters were able to differentiate between controls and presurgery patients for all vibratory features except left mucosal wave. Although all features showed an improvement in ratings following surgery, only left vocal fold edge was statistically different (by an average of 10 points or approximately one third of a scalar level). Following surgery, patients were still statistically different from controls. Using SS, raters were able to differentiate between controls and presurgery patients for all vibratory features. As with VS, all features showed an improvement in ratings following surgery; however, these improvements were statistically significant for left mucosal wave with an average improvement of 13 pp, left and right amplitude (with an average improvement of 7 pp and 4 pp, respectively), and left and right vocal fold edge (with an average improvement of approximately 13 points, or approximately one third of a scalar level). Following surgery, patients were not statistically distinguishable from controls for left and right amplitude. These findings suggest that SS may be a more robust tool than VS for determining patient outcomes following phonomicrosurgery, particularly for amplitude.

Clinical Implications

VS and SS offer different clinical advantages. VS is not limited by bandwidth or memory that restrict recording length. This offers a distinct advantage over SS in that clinicians have the flexibility to thoroughly test the laryngeal mechanism using more extensive assessment protocols. In addition, VS displays the simulated slow-motion movement of the vocal folds in real time during capture, which provides clinicians with visual feedback necessary to fine-tune scope placement for optimal visualization of the vocal folds. Currently, VS systems have a superior spatial resolution, which may provide better visualization of mucosal wave.

The advantage of SS methodologically is that it is derived from the full HSV recording after the fact. This offline processing means that pitch extraction can be determined by both preceding and following cycles of the GAW, allowing for a more reliable algorithm for phase selection with fewer asynchronous image sequences, which in turn provides more interpretable data regarding vibratory function. The clinical result is that not only can outcomes be computed for more patients using SS (Table 8) but also those outcome measures may be more statistically significant using SS compared with VS (Table 9).

In addition, although SS is bound to the same technical limitations of available bandwidth and memory as HSV, it offers the unique advantages of being derived from an HSV recording. First, it provides clinicians with a means for quickly previewing the entire HSV recording, and should the SS image sequence not synchronize to the true vibratory cycle, the HSV recording can be accessed for further analysis. This option is particularly helpful for patients with moderate to severe dysphonia whose acoustic output would be too aperiodic for VS analysis. Second, SS can provide stroboscopic image sequences for even short phonation samples. VS requires at least 2 seconds of sustained phonation to provide an adequate number vibratory cycles for analysis; however, you can increase the sampling rate with SS, thereby obtaining at least three vibratory cycles from much shorter phonation samples. The trade-off is the loss of video with synchronous auditory playback: Shorter samples would require increased sampling rate with a slowed playback rate, effectively disrupting the alignment between video and audio playback.

Conclusions

This study sought to determine whether SS improved on traditional VS in terms of rater reliability, rater confidence, and the overall number of patients that could be evaluated. Findings suggest that stroboscopy derived from HSV is a promising improvement over current acoustic-based VS systems. SS produced fewer asynchronous image sequences than VS, facilitating pre- and postoutcome measures in more patients than VS. Furthermore, SS is at least comparable with current VS technologies in terms of rater reliability, and raters indicated equal or increased confidence using SS over VS. SS improves on the current clinical gold standard in its ability to provide evaluation of more patients with vocal fold mass lesions and determine change in function before and after surgery. Future work should extend the current study to patients with other types of voice disorders and collect data from multiple raters to improve the generalizability of findings.

Acknowledgments

This study was fully supported by the National Institute on Deafness and Other Communication Disorders Grant R01-DC007640, "Efficacy of Laryngeal High-Speed Videoendoscopy", awarded to Dimitar Deliyski. Research was conducted in the Department of Communication Sciences and Disorders at the University of Cincinnati. Portions of this study have been presented at the Voice Foundation's 44th Annual Symposium: Care of the Professional Voice, Philadelphia, PA, May 2015. The authors thank the following researchers for their contributions to the execution of the study: Terri Gerlach, Stephanie Zacharias, Keiko Ishikawa, and Resmi Gupta.

References

- Behrman, A.** (2005). Common practices of voice therapists in the evaluation of patients. *Journal of Voice, 19*, 454-469.

- Bonilha, H. S., & Deliyski, D. D. (2008). Period and glottal width irregularities in vocally normal speakers. *Journal of Voice*, 22, 699–708.
- Bonilha, H. S., Deliyski, D. D., & Gerlach, T. T. (2008). Phase asymmetries in normophonic speakers. *American Journal of Speech-Language Pathology*, 17, 367–376.
- Bonilha, H. S., Deliyski, D. D., Whiteside, J. P., & Gerlach, T. T. (2012). Vocal fold phase asymmetries in patients with voice disorders: A study across visualization techniques. *American Journal of Speech-Language Pathology*, 21, 3–15.
- Bonilha, H. S., Focht, K. L., & Martin-Harris, B. (2015). Rater methodology for stroboscopy: A systematic review. *Journal of Voice*, 29, 101–108.
- Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., . . . Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology*, 258, 77–82.
- Deliyski, D. D. (2010). Laryngeal high-speed videoendoscopy. In K. Kendall & R. Leonard (Eds.), *Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging* (pp. 245–270). New York, NY: Thieme.
- Deliyski, D. D., Petrushev, P. P., Bonilha, H. S., Gerlach, T. T., Martin-Harris, B., & Hillman, R. E. (2008). Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. *Folia Phoniatrica et Logopaedica*, 60, 33–44.
- Deliyski, D. D., Powell, M. E. G., Zacharias, S. R. C., Gerlach, T. T., & de Alarcon, A. (2015). Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment. *Biomedical Signal Processing and Control*, 17, 21–28.
- Deliyski, D., Shaw, H., Martin-Harris, B., & Gerlach, T. (2005, June). *Facilitative playback techniques for laryngeal assessment via high-speed videoendoscopy*. Paper presented at the 34th Symposium Voice Found: Care of the Professional Voice, Philadelphia, PA.
- Eysholdt, U., Tigges, M., Wittenberg, T., & Pröschel, U. (1996). Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia phoniatrica et logopaedica*, 48, 163–170.
- Farnsworth, D. W. (1940). High-speed motion pictures of the human vocal cords. *Bell Laboratories Record*, 18, 203–208.
- Hertegård, S., Larsson, H., & Wittenberg, T. (2003). High-speed imaging: Applications and development. *Logopedics Phoniatrics Vocology*, 28, 133–139.
- Hillenbrand, J., & Gayvert, R. (2005). Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research*, 48, 45–60.
- Hillman, R., & Mehta, D. (2010). The science of stroboscopic imaging. In K. Kendall & R. Leonard (Eds.), *Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging* (pp. 245–270). New York, NY: Thieme.
- Hirano, M., & Bless, D. M. (1993). *Videostroboscopic examination of the larynx*. San Diego, CA: Singular.
- Kelley, R. T., Colton, R. H., Casper, J., Paseman, A., & Brewer, D. (2011). Evaluation of stroboscopic signs. *Journal of Voice*, 25, 490–495.
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18, 124–132.
- Kendall, K. A. (2009). High-speed laryngeal imaging compared with videostroboscopy in healthy subjects. *Archives of Otolaryngology–Head & Neck Surgery*, 135, 274–281.
- Mehta, D. D., & Hillman, R. E. (2012). Current role of stroboscopy in laryngeal imaging. *Current Opinion in Otolaryngology & Head & Neck Surgery*, 20, 429–436.
- Mehta, D. D., Zeitels, S. M., Burns, J. A., Friedman, A. D., Deliyski, D. D., & Hillman, R. E. (2012). High-speed videoendoscopic analysis of relationships between cepstral-based acoustic measures and voice production mechanisms in patients undergoing phonosurgery. *Annals of Otolaryngology, Rhinology & Laryngology*, 121, 341–347.
- Mendelsohn, A. H., Remacle, M., Courey, M. S., Gerhard, F., & Postma, G. N. (2013). The diagnostic role of high-speed vocal fold vibratory imaging. *Journal of Voice*, 27, 627–631.
- Nawka, T., & Konerding, U. (2012). The interrater reliability of stroboscopy evaluations. *Journal of Voice*, 26, 812.e1–812.e10.
- Olthoff, A., Woywood, C., & Kruse, E. (2007). Stroboscopy versus high-speed glottography: A comparative study. *Laryngoscope*, 117, 1123–1126.
- Patel, R., Dailey, S., & Bless, D. (2008). Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Annals of Otolaryngology, Rhinology & Laryngology*, 117, 413–424.
- Paul, B. C., Chen, S., Sridharan, S., Fang, Y., Amin, M. R., & Branski, R. C. (2013). Diagnostic accuracy of history, laryngoscopy, and stroboscopy. *Laryngoscope*, 123, 215–219.
- Poburka, B. J. (1999). A new stroboscopy rating form. *Journal of Voice*, 13, 403–413.
- Powell, M. E., Deliyski, D. D., Mehta, D. D., & Hillman, R. E. (2015, April). *Validation of stroboscopy derived from high-speed videoendoscopy*. Paper presented at the 11th International Advances in Quantitative Laryngology, Voice and Research, London, UK.
- Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M. P., Mehta, D., Paul, D., & Hillman, R. (2013). Evidence-based clinical voice assessment: A systematic review. *American Journal of Speech-Language Pathology*, 22, 212–226.
- Sataloff, R. T., Spiegel, J. R., Carroll, L. M., Schiebel, B. R., Darby, K. S., & Rulnick, R. (1988). Stroboscopy in professional voice users: Results and clinical value. *Journal of Voice*, 1, 359–364.
- Titze, I. R. (1994). *Workshop on acoustic voice analysis*. Denver, CO: National Center for Voice and Speech.
- Woo, P., Colton, R., Casper, J., & Brewer, D. (1991). Diagnostic value of stroboscopic examination in hoarse patients. *Journal of Voice*, 5, 231–238.
- Yanagisawa, E., Casuccio, J. R., & Suzuki, M. (1981). Video laryngoscopy using a rigid telescope and video home system color camera a useful office procedure. *Annals of Otolaryngology, Rhinology & Laryngology*, 90(4), 346–350.