Biomedical Optics EXPRESS

Synchronized, concurrent optical coherence tomography and videostroboscopy for monitoring vocal fold morphology and kinematics

GOPI MAGULURI,[1] DARYUSH MEHTA,[2] JAMES KOBLER,[2] JESUNG PARK,[1] AND NICUSOR IFTIMIA[1,*]

[1] *Physical Sciences Inc., Andover, MA 01810, USA*
[2] *Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114, USA*
[*] *iftimia@psicorp.com*

Abstract: Voice disorders affect a large number of adults in the United States, and their clinical evaluation heavily relies on laryngeal videostroboscopy, which captures the medial-lateral and anterior-posterior motion of the vocal folds using stroboscopic sampling. However, videostroboscopy does not provide direct visualization of the superior-inferior movement of the vocal folds, which yields important clinical insight. In this paper, we present a novel technology that complements videostroboscopic findings by adding the ability to image the coronal plane and visualize the superior-inferior movement of the vocal folds. The technology is based on optical coherence tomography, which is combined with videostroboscopy within the same endoscopic probe to provide spatially and temporally co-registered images of the mucosal wave motion, as well as vocal folds subsurface morphology. We demonstrate the capability of the rigid endoscopic probe, in a benchtop setting, to characterize the complex movement and subsurface structure of the aerodynamically driven excised larynx models within the 50 to 200 Hz phonation range. Our preliminary results encourage future development of this technology with the goal of its use for *in vivo* laryngeal imaging.

© 2019 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Voice disorders affect approximately 30% of adults in the United States at some point in their lives [1–3] and include a broad spectrum of diagnostic categories, including phonotraumatic vocal hyperfunction (organic vocal fold lesions such as nodules, polyps), non-phonotraumatic vocal hyperfunction (muscle tension dysphonia in the absence of lesions), laryngeal paralysis/paresis, and vocal fold scar.

Laryngologists and speech-language pathologists heavily rely on laryngeal videostroboscopy (VS) for the clinical assessment of voice disorders. VS captures medial-lateral and anterior-posterior motion of the vocal folds using stroboscopic sampling [4–7], enabling clinicians to observe many salient features in real time, such as left-right symmetry and amplitude, which are difficult to evaluate at standard video rates [8]. In addition to clinical diagnosis, VS is also used to evaluate the surface motion of the vocal folds before and after surgical intervention [9]. These capabilities, combined with cost affordability, have promoted VS as the ‘gold standard’ clinical tool for diagnosing voice disorders. However, VS has limited capabilities. It only provides the assessment of vocal fold motion in the medial-lateral dimension, with only an indirect appreciation of superior-inferior, or vertical, tissue motion in the coronal plane. However, accurate measurement of vertical tissue motion is considered very significant for understanding vocal fold function [10]. Furthermore, VS is not capable of subsurface visualization, which is

#360886 https://doi.org/10.1364/BOE.10.004450
Journal © 2019 Received 26 Feb 2019; revised 17 Jul 2019; accepted 24 Jul 2019; published 6 Aug 2019

important in highlighting epithelial and subepithelial tissue abnormalities (e.g., scarring, nodules, polyps, cancer lesions, etc.) that occur in the superficial lamina propria layer of the vocal folds.

Besides VS, high-speed videoendoscopy (HSV) at 1000 frames/sec or higher is used in research studies, as well in some clinics, due to its temporal sampling advantages over VS, particularly when phonation is aperiodic [11–14]. However, besides the high cost of the camera, the analysis of HSV images can be time-consuming and cumbersome. As a result, HSV clinical adoption has proven to be fairly limited.

To compensate for VS limitations, scientists have considered the use of optical coherence tomography (OCT) as an adjunct tool to VS, capable of providing both superior-inferior surface motion and subsurface information [15]. OCT penetrates axially (into the tissue in the coronal plane) to depths up to 2 mm, depending on the used wavelength and tissue light scattering properties [16–18], and thus can resolve important pathological structures such as scar tissue, nodules, polyps, cysts, dysplasia, and cancer [19–21]. OCT recordings of 3D vocal fold kinematics through a rigid endoscope have been performed before with swept-source (SS) based systems with fast image acquisition rates [22,23]. However, even the high speed OCT systems (>100kHz) lacked temporal co-registration with vocal fold movement and exhibited signal-to-noise degradation when the beam was scanned across the moving vocal folds, and thereby, suffered from aliasing and motion blur [23–25].

Previous efforts to complement VS also included other imaging modalities. Ultrasound imaging has been used to trace tissue particles in motion pictures along the coronal plane with good penetration, but it suffered from poor spatial resolution [26,27]. Optical imaging based on structured illumination has been used to measure vocal fold kinematics in both excised and *in vivo* experimental configurations [28–31]. Whereas this method allowed for fast imaging rates (~4000 fps), it suffered from poor spatial resolution (10 points per line) and significant computational overhead for vocal fold surface reconstruction. A depth-kymography technique was used to capture vocal fold surface motion in 3D, but required a long processing time (15–20 minutes), and provided low resolution (~100 µm) and limited subsurface information [32,33].

In this paper we present the design of a transoral, rigid laryngeal endoscopic probe that integrates OCT and VS imaging within the same optical path. Preliminary testing of the OCT/VS imaging system on calf and human excised larynges has demonstrated the capability of this technology to provide spatially and temporally co-registered OCT and VS images during phonation. These images enabled the reconstruction of 3D vocal fold kinematics, as well as the visualization of their subsurface morphology.

2. Materials and methods

2.1. Instrument description

A portable instrument with a common OCT/VS optical path probe was designed and assembled. The instrument contains five subsystems: Imaging Probe, OCT Unit, Strobe Unit, Data Acquisition & Processing Unit, and Display (see schematic in Fig. 1).

The Imaging Probe uses a dichroic mirror to combine the OCT and VS channels through the same distal-end optics, consisting of a set of relay lenses, an imaging objective, and a beam-deflection mirror. The OCT channel includes a fiber collimator, a folding mirror, a beam-divergence adjusting lens, and a galvanometer-based scanning engine consisting of 2 galvanometers (Model 6200, Cambridge Technology, Bedford, MA). The VS channel uses a folding mirror, a focus adjustment lens, and a camera objective lens system. The parfocality of the OCT and VS channels is enabled by the focus adjustment lenses used in the video channel. The imaging probe consists of a 3D printed case ~ (1.5” × 5” × 6.5”) enclosing the optics module, with the endoscope attached that is of ~7” in length and 18 mm in diameter.

The OCT Unit is based on the swept-source OCT approach that offers the advantage of a negligible washout of the fringe signal compared to a spectral-domain approach [34]. A

Biomedical Optics EXPRESS

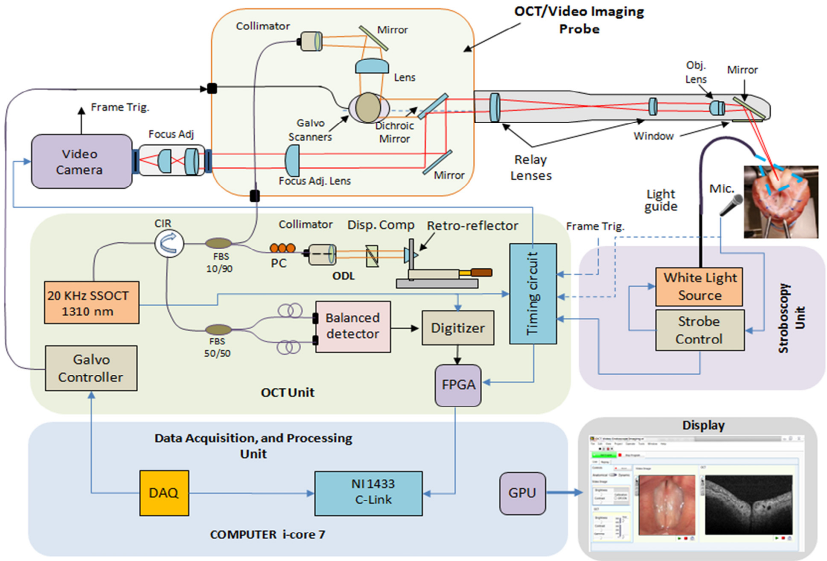


Fig. 1. Schematic of the OCT/VS system; Abbreviations: Cir- Circulator; FPGA- Field Programmable Gated Array; Mic- Microphone; FBS- fiber beam splitter; GPU- Graphical Processing Unit; PC- Polarization Controller; DAQ- Data Acquisition Card; ODL- Optical Delay Line; Disp- Dispersion.

broadband wavelength-sweeping laser source (Model HSL-2000, Santec, Inc.) with 1310 nm central wavelength, a 3-dB bandwidth of 100 nm, a 20 kHz sweeping frequency, and an average power of 12 mW is used as illumination source. This low illumination provides minimal risk with no discomfort or pain to subjects and does not exceed the ANSI standard for safe human use [35]. The near-infrared light is passed through a fiber circulator and then split by a 1 × 2 (90:10) single-mode coupler, such that 90% of the light goes to the probe and 10% to the optical delay line. The light beam going into the endoscopic probe is scanned by two galvanometers (6200H, Cambridge Technology) to generate volumetric OCT images. OCT spectral interferograms generated by the reflected light from both arms of the fiber interferometer are sent to a balanced photodetector. The signal from the balanced detector is digitized using a 100 MHz digitizer and is fed to a custom-built field-programmable gate array (FPGA) module. A custom-built timing circuit is used to synchronize the OCT data acquisition with the VS unit. The timing sequence in the circuit uses the strobe trigger from the Strobe Unit to control the frame rate of the video camera and the streaming of the OCT data to the FPGA board.

The Data Acquisition & Processing Unit consists of a computer equipped with a data acquisition card, a frame-grabber, and a graphics processing unit (GPU), which enables real-time data processing and display. Depth-resolved structural images are generated in real time and displayed at various rates (between 10 and 40 Hz) depending on the number of pixels per frame.

As shown in Fig. 2, both the OCT and the Data Acquisition & Processing units are enclosed within a custom-made 19” rack. The OCT imaging beam is directed at an angle of 70 degrees from the distal end of the probe, and the imaging plane is formed at a distance of 60 mm from the probe tip to meet the clinical requirements for endoscopic laryngeal imaging.

2.2. Benchtop setup for excised larynx models

Five fresh calf larynges, obtained from a research tissue provider, Research87 (Boylston, MA), as well as two previously frozen human larynx specimens, which were obtained from autopsies

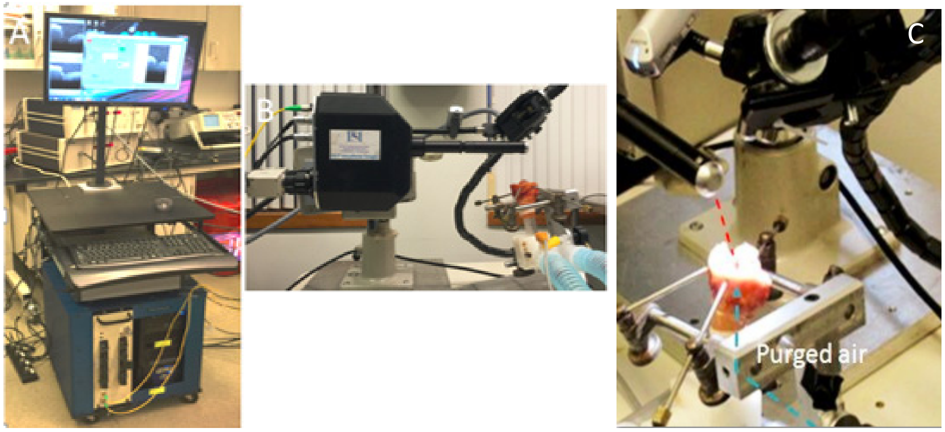


Fig. 2. Photographs of the instrument and tissue specimen fixtures. A- OCT instrumentation unit; B- OCT/video imaging probe; C- Tissue specimen fixture.

performed in the Massachusetts General Hospital Department of Pathology, as approved by the Partners Institutional Review Board were used in this study. The excised larynx specimens were mounted in a custom holder by placing the trachea over a plastic tube, and the cricoid and thyroid cartilages were secured with four corkscrew-tipped rods that were mounted on a rigid frame surrounding the specimen. Tissues above the true vocal folds were partially resected and/or retracted with stay sutures to provide a clear view of the superior surface of the true vocal folds. A suture through the paired arytenoid cartilages was used to adduct the vocal processes into a phonatory posture. Warm (37 °C) humidified air was generated using a ConchaTherm (model 380-55, Respiratory Care Inc., Arlington Heights, IL) and used to drive phonation. Airflow was controlled with a pressure regulator, and subglottal pressure was monitored using a pressure transducer placed about 10 cm below the vocal folds. The phonatory acoustic signal was captured with a microphone that was placed 10 cm above the glottis. The subglottal pressure signal was passed through a custom zero-crossing detector circuit to generate a synchronization TTL pulse for each glottal cycle.

The instrument was first tested in a morphological scanning (M-scan) mode, where the vocal folds were at rest (non-vibratory state). A raster scan (C-scan) consisting of 512 B-scans was obtained for both human and calf excised larynx specimens. The OCT unit yielded subsurface images of the vocal folds with 10 µm superior-inferior resolution and 40 µm medial-lateral resolution, which was mainly dictated by the relatively long imaging plane (~60 mm away from probe tip) that is typical during clinical practice.

2.3. OCT-VS synchronization

Stroboscopy is commonly used to observe vocal fold vibration at typically 0.5 to 2 periods per second, enabling real-time examination of vocal fold motion in the clinic [36]. A microphone, placed near the larynx, recorded an acoustic signal that was down-sampled to generate a low-frequency strobe illumination trigger per video field (see Fig. 3(A)). The trigger was incremented relative to the phase of the glottal cycle for each video field such that vocal fold oscillation appears slowed down. This is represented as the “reconstructed glottal cycle” in Fig. 3(B). Synchronization of OCT with VS was then achieved by triggering B-scan OCT acquisitions simultaneously with each strobe flash by using the synchronization pulse.

In this way, the VS images and the OCT B-scans were acquired simultaneously during a reduced speed glottal cycle (see Figs. 3(C)–3(E)). VS acquires top-down two-dimensional (2D)

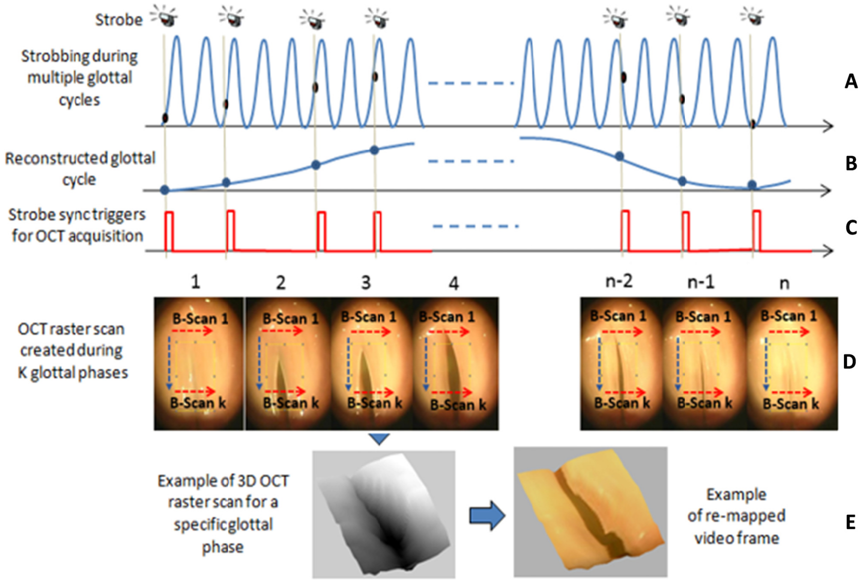

Strobe
Strobing during multiple glottal cycles
A
Reconstructed glottal cycle
B
Strobe sync triggers for OCT acquisition
C
1
2
3
4
n-2
n-1
n
OCT raster scan created during K glottal phases
B-Scan 1
B-Scan k
D
Example of 3D OCT raster scan for a specific glottal phase
Example of re-mapped video frame
E

Fig. 3. OCT synchronization scheme for stroboscopic imaging. Data was captured from previously frozen human larynx

images of the vocal folds, while OCT acquires cross-sectional information in the third, vertical dimension. The plane of the OCT image can be incremented for each new reconstructed glottal cycle along the anterior-posterior axis of the vocal folds to capture multiple slices and generate a 3D image, which can be remapped on the video image to generate a high-resolution 3D video of vocal fold movement (Fig. 3(E) illustrates one such 3D image). The duration of the B-scan is determined by the number and rate of A-lines captured. A relatively sparse A-line density (16 A-lines per scan using a 20 kHz OCT system) was used in order to capture the essential features of the mucosal wave without motion blur artifacts during a B-scan. Higher-density scans could be performed with faster swept-source OCT systems, at the expense of a significant increase in instrument cost.

2.4. OCT-VS image processing and visualization

The 3D rendering of the glottal phases, similar to the one shown in Fig. 3(E), was made possible by overlaying the OCT images on the VS images. This allowed to reconstruct a videoendoscopic movie that showed both the medial-lateral and superior-inferior motion of the VFs. An image processing and visualization algorithm was developed and implemented in MATLAB (The Math Works, Inc.) for this purpose. The main image processing steps are illustrated in the flow chart as shown in Fig. 4. Sparse OCT B-scans, corresponding to several phases of a glottal cycle were acquired at time intervals t_0 to t_n (glottal phases) and repeated at several vertical locations in subsequent phonation cycles, forming a 4D raster (C-scans collected at different time intervals). The coarse resolution of the B-scans was dictated by the number of the A-lines that can be collected within a glottal cycle, for a specific phonation frequency. For a swept-source sweeping frequency of 20 kHz, we were able to collect 10 B-scans associated to 10 phases of a glottal cycle repeating it subsequently for 10 vertical positions of the VF opening. For each B-scan, the vertical profile of the mucosal wave was obtained by segmenting each image and detecting the VF surface.

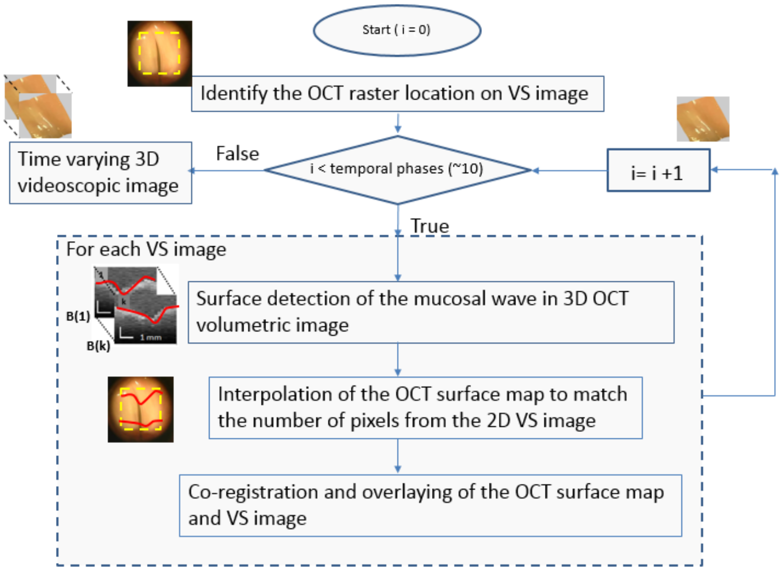

Start (i = 0)
Identify the OCT raster location on VS image
i < temporal phases (~10)
False
Time varying 3D videoscopic image
i= i +1
True
For each VS image
B(1)
B(k)
1 mm
Surface detection of the mucosal wave in 3D OCT volumetric image
Interpolation of the OCT surface map to match the number of pixels from the 2D VS image
Co-registration and overlaying of the OCT surface map and VS image

Fig. 4. Flow chart for image processing and visualization

The main steps for data processing were: (1) surface detection of the mucosal wave in the 3D OCT volumetric image; (2) interpolation of the OCT surface map to match the number of pixels from the 2D VS image, and (3) overlaying of the OCT surface map on the VS image.

In the first step, tissue surface formed from B-scans at each vertical step (e.g. at t_0) is detected. The detection of the surface of mucosal wave is performed by calculating the maximum derivative of the slope of each A-line in the OCT B-scan image. It is to be noted that each OCT scan is sparser (16 A-lines *10 B-scans) than the VS image, which has 400 *800 pixels spanning 5 mm *10 mm. An OCT surface map is generated by assembling the surfaces of the OCT B-scans images in a 3D volume. Linear interpolation is then used to smooth the surface of the reconstructed mucosal wave and match it to the number of pixels of the 2D VS images. This routine is repeated for subsequent glottal phases (t_1 to t_n). The superimposed OCT/VS images were rendered using ImageJ (1.8.0, NIH, Bethesda, MD) for optimal display.

2.5. Resolution

The medial-lateral resolution l_s (number of A-lines per B-scan) is defined by the following relation:

$$l_s \sim A_r/(f_0 * \phi_n) \tag{1}$$

where A_r is the A-line rate of the swept source system, f_0 is the fundamental frequency of phonation, and ϕ_n is the number of phases desired per strobe cycle, equivalent to the strobe flashes utilized. With a 20 kHz A-line rate, for a 100 Hz fundamental frequency and a desired number of 10 phases, we were able to record 16 A-lines per B-scan spanning 5 mm, which provided a medial-lateral resolution of 312 µm per pixel. We collected 10 B-scans in vertical direction spanning 10 mm that provided an anterior-posterior resolution of 1 mm. It can be noted that this resolution is easily scalable with longer durations of phonation at relatively constant fundamental frequency as each step is based on an independent reconstructed glottal cycle. The superior-inferior (axial) resolution of the OCT system was measured to be ~12 µm, dictated by the light source bandwidth. We were able to record 10 phases for a phonation frequency of 100 Hz, giving us a temporal resolution of 1 millisecond.

Biomedical Optics EXPRESS

3. Results

The instrument was used to collect high-resolution images in the morphological M-scan mode with the vocal folds at rest, and lower lateral resolution in the dynamic OCT-VS mode. The results from the excised larynx experiments are reported here.

3.1. Morphological M-scan mode

With the vocal folds at rest, high-density cross-sectional OCT scans with 1024 A-lines per B-scan were collected. As shown in Fig. 5, the epithelium, superficial lamina propria, and small blood vessels were well resolved, although specimen quality was not optimal for optical imaging due to being frozen and thawed.

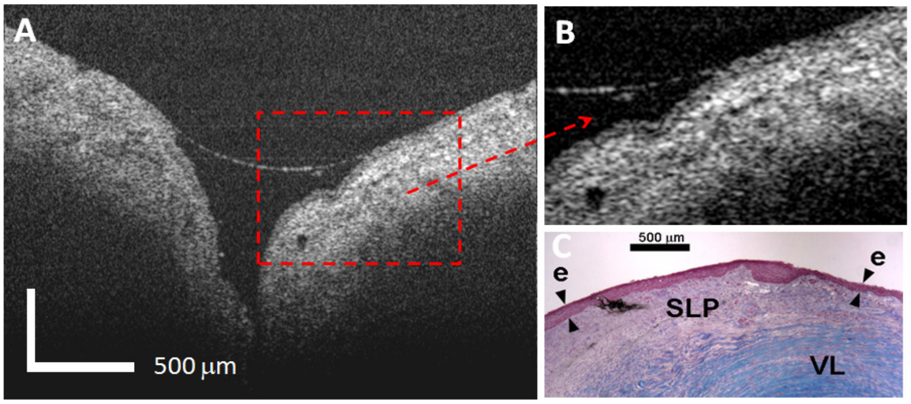


Fig. 5. Example of a cross-sectional OCT image of human excised vocal folds. (A) large-scale image; (B) 2.5X magnified area; (C) example of histology (from a different specimen). Legend: e-epithelium; SLP-Superficial lamina propria; VL-vocal ligament

3.2. Stroboscopic OCT-VS mode

Synchronized VS-OCT imaging was performed with the vocal folds vibrating at fundamental frequencies ranging from 50 to 200 Hz (80 to 200 Hz for the human larynx specimens and from 50 to 160 Hz for the calf specimens). An example of ten phases of vocal fold motion with associated sample OCT depth information is shown in Fig. 6 for a calf specimen phonated at 100 Hz. In our system, the strobe flash frequency determined the temporal and spatial resolution of the OCT images. The strobe triggers, at about 60 Hz, made the cyclical motion appear to be 0.5 Hz. Hence, it took 2 seconds to capture one full glottal cycle. During this visualized cycle, OCT has captured one axial cross-section (B-scan) per trigger in medial-lateral direction. For each temporal phase imaged by VS, the number of A-lines across the vocal folds (B-scan) was determined by the fundamental frequency of phonation and the speed of the OCT source as dictated by Eq. (1). Here, for 20 kHz A-line rate and 100 Hz fundamental frequency, we acquired 16 A-lines (one B-scan) to generate 10 phases per reconstructed glottal cycle.

Since stroboscopy enabled the examination of multiple phonatory cycles at a reduced speed, it was possible to collect OCT B-scans synchronously with VS images and build up 3D maps of vocal fold kinematics. Figure 7 shows examples of co-registered OCT-VS frames at 3 different phases, each one with corresponding OCT B-scan at the center, OCT 3D interpolated surface of 10 B-scans, and reconstructed OCT-VS superimposed images. For simplicity, three distinct temporal phases of a phonatory cycle are shown in the left column, when the vocal folds are closed (A_1), half open (B_1) and fully open (C_1). The second column from the left shows the corresponding OCT B-scans (A_2, B_2, C_2) at the center location of the C scan, with the red line

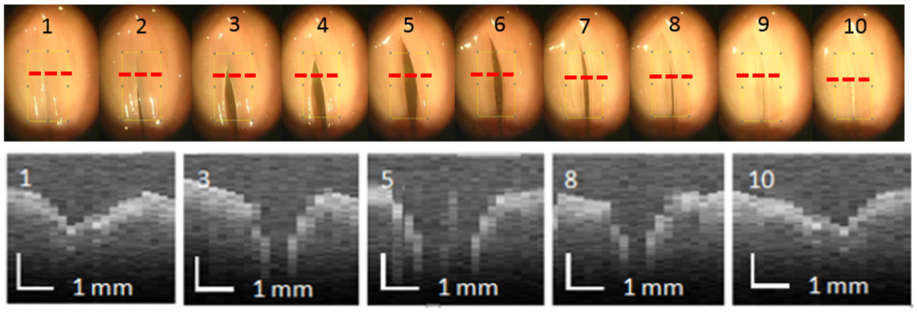


Fig. 6. Example of synchronized OCT-VS imaging of vocal folds showing 10 temporal phases (only 5 representative OCT B-scans are shown here)

tracing the vocal fold surface. The third column from the left shows the OCT 3D surface images (A_3, B_3, C_3), which were obtained by interpolating 10 slices of the OCT C-scan taken across the anterior-posterior direction. In the fourth column of Fig. 7 are shown the VS images re-mapped based on the OCT 3D surface data, creating a pseudo-3D video stroboscopic image. Due to the low resolution of the OCT scan, an interpolation between OCT scans at various lateral positions has been made, to generate an image with the same number of pixels as the VS image. As a result, some artifacts can be noticed in the reconstructed image.

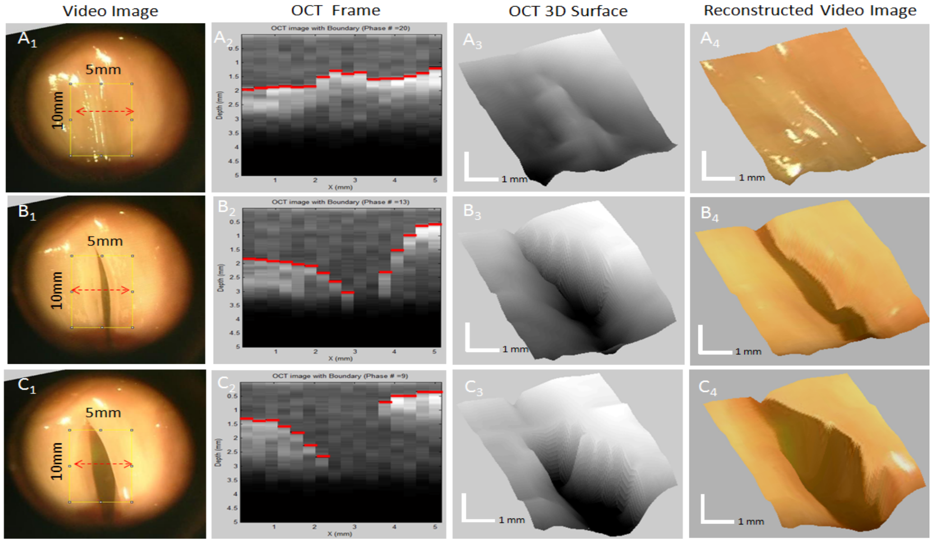
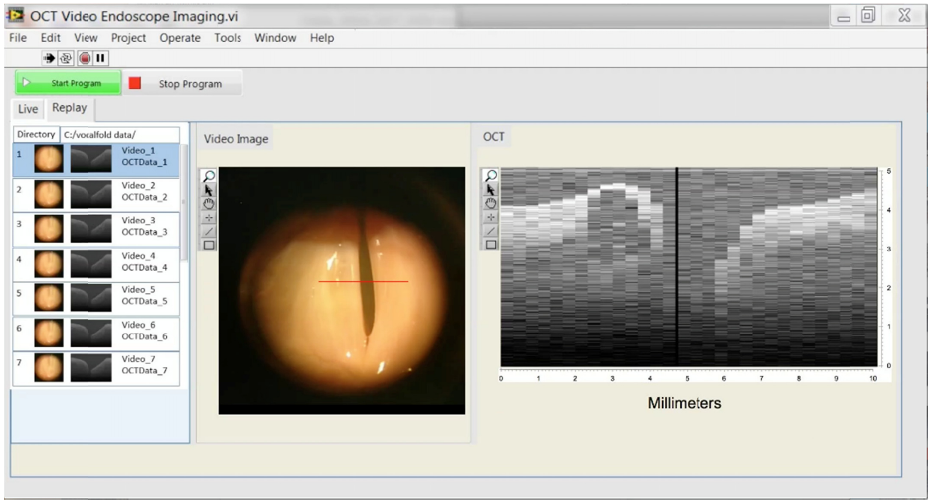


Fig. 7. Example of synchronized OCT-VS imaging of calf vocal folds. A1-B1-C1: VS images for 3 different openings of the VFs; A2-B2-C2: Corresponding cross-sectional OCT images at the middle of the opening, as indicated by red arrows; A3-B3-C3: Reconstructed OCT images from the interpolation of 10 B-scans; A4-B4-C4: VS images reconstructed based on the OCT images.

The subsurface morphology was recovered using the high-density M-scan mode, while the vocal folds were at rest, as shown in Fig. 5. However, subsurface information can be obtained during phonation at a lower resolution, as shown in Fig. 8 (see Visualization 1). The right vocal fold (left side of image) was injected sub-epithelially with a stiff hyaluronic acid-based gel (Restylane). To improve resolution, the two vocal folds were imaged separately and then combined during post-processing. Each A-line in the OCT image was expanded horizontally to restore the natural aspect ratio. The injected gel can be seen in the OCT image on the left side as a low-reflectance, circular zone, approximately 1.25 mm in diameter. In contrast, the gel injection site is not easily visible in the VS video images. The presence of the gel induced asymmetry of the mucosal wave in vertical axis, is clearly apparent in the OCT video but more challenging to observe in the VS data. In this particular case, the fundamental phonation frequency was ~50 Hz, which allowed the system to capture 20 phases per cycle and thus 16 A-lines per B-scan on each of the vocal folds.



Directory C:/vocalfold data/

Fig. 8. Mock-up of graphical user interface in stroboscopic OCT-VS mode. The video and OCT data were collected from a phonating human ex-vivo larynx and were combined using post-processing to simulate the display interface [Visualization 1]

Besides the amplitude of the vocal fold vibration, velocity measurements were derived as well from the recorded data set, as shown in Fig. 9. The plots in Fig. 9(a) show the surface contour of the vocal folds across the medial-lateral dimension. The surface contour was measured manually by placing each B-scan image on a high resolution grid for five equidistant phases (captured OCT B-scans). With the known phonation frequency of 50 Hz, for the 20 phases captured, the elapsed time between each phase was estimated to be 1.05 milliseconds. These data were used to calculate the maximum velocity in the vertical (z) dimension at which each lateral position of the vocal fold surface moved (see Fig. 9(b)). As it can be observed, the maximum velocity varied across the vocal folds since the mucosal wave speed varied across the vocal fold surface. The opening of the vocal folds vibrated at the highest speed (~100 mm/sec) with over 1 mm amplitude. The speed/amplitude decreased while the mucosal wave traveled laterally from the midline. However, OCT could not detect the position of the surface at the opening, and thus the velocity could not be evaluated at the midline.

The measured amplitudes and velocities may be compared with those obtained for human vocal folds in *in vivo* (~1–2 mm and ~2 m/sec, respectively) [31] and excised (~2 mm and ~1 m/s, respectively) [29]. However, the human excised larynx exhibited an induced reduced amplitude, presumably significantly affecting the normal kinematics of the contralateral vocal fold that, while mechanically normal, came into contact with a vocal fold of atypical stiffness. As acknowledged in [31], these types of measurements are not widely available in the literature and may be significantly affected by the type of larynx (*in vivo*, excised, healthy/pathological, etc.), imaging technology (structured illumination, etc.), and specific tissue tracking method (suture flesh points, medial edge tracking, surface edge tracking, etc.). In addition, a comprehensive

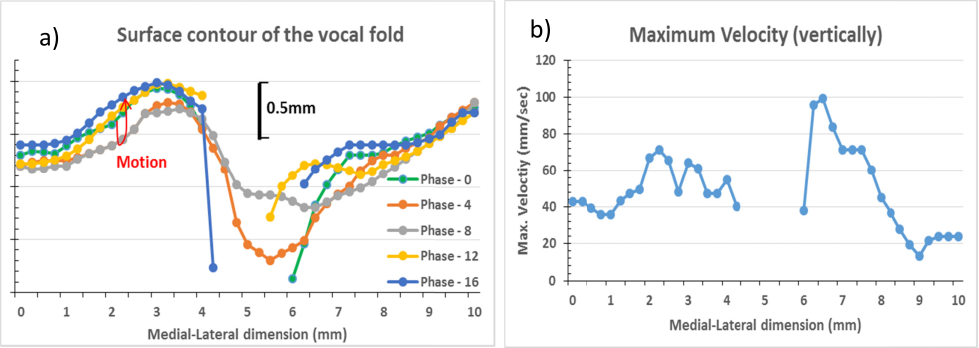


Fig. 9. a) Illustration of vocal fold position for five different phases in the middle of the opening, as derived from the OCT B-scans. b) Vocal fold maximum velocity in the superior-inferior dimension at points in the medial-lateral dimension, as derived from the surface position at different time intervals.

reporting of phonatory characteristics is needed for fair cross-study comparisons, including subglottal pressure, sound pressure level, and fundamental frequency. The current study presented a proof-of-concept technology that can be used in the future for more detailed studies of vocal fold kinematics.

4. Discussion and conclusions

This paper demonstrates the feasibility of combined OCT/VS used for vocal folds imaging during phonation, enabling the clinician to quantify the amplitude of the mucosal wave, as well as its temporal and spatial irregularities during a clinical exam. To our knowledge, this is the first demonstration of combined OCT/VS within the same optical path, providing spatial co-registration and temporal synchronization of the recordings.

Previous efforts imaging vocal folds with OCT imaging systems (including polarization-sensitive OCT) with catheter-based systems have shown high-quality images with useful subsurface information [17,19–21]. These studies, however, were performed with the imaging probe in contact with the vocal folds, and thus were not suitable for dynamic imaging. More recent efforts using high speed OCT technique [23] have demonstrated the dynamic imaging capability, but significant aliasing was still noted.

The presented approach scores over previous efforts in terms of temporal synchronization with VS, which is currently the gold standard for laryngologists. It demonstrated the capability of providing co-registered visualization of the mucosal wave amplitude (using OCT), while maintaining the existing benefits of the VS for the diagnostic practice.

The system described in this paper has certain limitations, mainly related to the reduced speed of the used OCT swept source. By increasing the A-line rate of the swept source to 200 kHz or more, either the lateral resolution, as described in Eq. (1) or the phase resolution can be improved by 10 times.

The increase of the A-line rate could also reduce image blurring due to the reduced time used to produce a number of B-scans during the duration of the glottal cycle. During each phase, the vocal folds are still moving as the beam is scanned across them (B-scan), although the motion is limited to 1/nth of the laryngeal cycle for a desired ‘n’ number of phases to be captured. In terms of the imaging time, 10 B-scans were acquired in 2 sec (time for one reconstructed glottal cycle). Ten such axial movement scans separated by 1 mm in the orthogonal direction took 20 seconds.

Another limitation is that we are underutilizing the strobe flashes (strobe flashes allow to capture a maximum of 120 phases in 2 seconds in a single reconstructed glottal cycle). The

OCT system was too slow to capture a B-scan for each strobe frame. To achieve a more accurate temporal snapshot, a parallel OCT approach might allow the capture of one OCT cross-section per strobe flash [37–40]. Another key parameter requiring improvement is the diameter of the current probe head (~18 mm), which is almost twice as large as the current state of the art rigid laryngoscopes used for *in vivo* human vocal fold imaging [41].

Imaging range is another key feature for clinical imaging. To employ this approach in a clinical setting, a laser source with longer imaging range (at least 10 mm) is preferred to keep the vocal folds within the imaging range during phonation. A stabilization apparatus compensating unwanted relative motion between the probe head and vocal folds is potentially needed for clinical operation to keep the motion of the vocal folds within the imaging range of the OCT system. Therefore, future efforts will employ the use of a high-speed parallel OCT approach with synchronous video imaging, along with a stabilization apparatus which would enable clinicians to reliably monitor 3D vocal fold kinematics in real time.

Funding

National Institute on Deafness and Other Communication Disorders (1R43DC013743-01, 1R43DC017407-01).

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosures

The authors declare that there are no conflicts of interest related to the work presented in this article.

References

1. N. Bhattacharyya, “The prevalence of voice problems among adults in the United States,” Laryngoscope **124**(10), 2359–2362 (2014).
2. N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, “Voice disorders in the general population: prevalence, risk factors, and occupational impact,” Laryngoscope **115**(11), 1988–1995 (2005).
3. NIDCD, Statistics on Voice, Speech and Language, National Institute of Deafness and Other Communication Disorders (NIDCD), July 2016. https://www.nidcd.nih.gov/health/statistics/statistics-voice-speech-and-language
4. K. A. Kendall, “High-speed digital imaging of the larynx: recent advances,” Curr Opin Otolaryngol Head Neck Surg **20**(6), 466–471 (2012).
5. D. M. Bless, M. Hirano, and R. J. Feder, “Videostroboscopic evaluation of the larynx,” Ear Nose Throat J **66**, 289–296 (1987).
6. N. Roy, J. Barkmeier-Kraemer, T. Eadie, M. P. Sivasankar, D. Mehta, D. Paul, and R. Hillman, “Evidence-based clinical voice assessment: a systematic review,” Am J Speech Lang Pathol **22**(2), 212–226 (2013).
7. D. D. Mehta and R. E. Hillman, “Current role of stroboscopy in laryngeal imaging,” Curr Opin Otolaryngol Head Neck Surg **20**(6), 429–436 (2012).
8. B. J. Poburka, “A new stroboscopy rating form,” J Voice **13**(3), 403–413 (1999).
9. S. M. Zeitels, A. Blitzer, R. E. Hillman, and R. R. Anderson, “Foresight in laryngology and laryngeal surgery: a 2020 vision,” Ann. Otol., Rhinol., Laryngol. **116**(9_suppl), 2–16 (2007).
10. C. R. Krausert, A. E. Olszewski, L. N. Taylor, J. S. McMurray, S. H. Dailey, and J. J. Jiang, “Mucosal wave measurement and visualization techniques,” J Voice **25**(4), 395–405 (2011).
11. D. D. Deliyski, P. P. Petrushev, H. S. Bonilha, T. T. Gerlach, B. Martin-Harris, and R. E. Hillman, “Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution,” Folia Phoniatr. **60**(1), 33–44 (2008).
12. R. Patel, S. Dailey, and D. Bless, “Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders,” Ann. Otol., Rhinol., Laryngol. **117**(6), 413–424 (2008).
13. G. B. Kempster, B. R. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, “Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol,” Am J Speech Lang Pathol **18**(2), 124–132 (2009).

14. D. D. Deliyski, R. E. Hillman, and D. D. Mehta, “Laryngeal High-Speed Videoendoscopy: Rationale and Recommendation for Accurate and Consistent Terminology,” J. Speech Hear. Res. **58**(5), 1488–1492 (2015).
15. J. B. Kobler, E. W. Chang, S. M. Zeitels, and S. H. Yun, “Dynamic imaging of vocal fold oscillation with four-dimensional optical coherence tomography,” Laryngoscope **120**(7), 1354–1362 (2010).
16. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, *et al.*, “Optical coherence tomography,” Science **254**(5035), 1178–1181 (1991).
17. J. A. Burns, S. M. Zeitels, R. R. Anderson, J. B. Kobler, M. C. Pierce, and J. F. de Boer, “Imaging the mucosa of the human vocal fold with optical coherence tomography,” Ann. Otol., Rhinol., Laryngol. **114**(9), 671–676 (2005).
18. G. J. Tearney, M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern, and J. G. Fujimoto, “In vivo endoscopic optical biopsy with optical coherence tomography,” Science **276**(5321), 2037–2039 (1997).
19. J. A. Burns, “Optical coherence tomography: imaging the larynx,” Curr Opin Otolaryngol Head Neck Surg **20**(6), 477–481 (2012).
20. K. H. Kim, J. A. Burns, J. J. Bernstein, G. N. Maguluri, B. H. Park, and J. F. de Boer, “In vivo 3D human vocal fold imaging with polarization sensitive optical coherence tomography and a MEMS scanning catheter,” Opt. Express **18**(14), 14644–14653 (2010).
21. B. J. Wong, R. P. Jackson, S. Guo, J. M. Ridgway, U. Mahmood, J. Su, T. Y. Shibuya, R. L. Crumley, M. Gu, W. B. Armstrong, and Z. Chen, “In vivo optical coherence tomography of the human larynx: normative and benign pathology in 82 patients,” Laryngoscope **115**(11), 1904–1911 (2005).
22. L. Yu, G. Liu, M. Rubinstein, A. Saidi, B. J. Wong, and Z. Chen, “Office-based dynamic imaging of vocal cords in awake patients with swept-source optical coherence tomography,” J. Biomed. Opt. **14**(6), 064020 (2009).
23. C. A. Coughlan, L. D. Chou, J. C. Jing, J. J. Chen, S. Rangarajan, T. H. Chang, G. K. Sharma, K. Cho, D. Lee, J. A. Goddard, Z. Chen, and B. J. Wong, “In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography,” Sci. Rep. **6**(1), 22792 (2016).
24. G. Liu, M. Rubinstein, A. Saidi, W. Qi, A. Foulad, B. Wong, and Z. Chen, “Imaging vibrating vocal folds with a high speed 1050 nm swept source OCT and ODT,” Opt. Express **19**(12), 11880–11889 (2011).
25. F. Benboujja, J. A. Garcia, K. Beaudette, M. Strupler, C. J. Hartnick, and C. Boudoux, “Intraoperative imaging of pediatric vocal fold lesions using optical coherence tomography,” J. Biomed. Opt. **21**(1), 016007 (2016).
26. B. Jing, Z. Ge, L. Wu, S. Wang, and M. Wan, “Visualizing the mechanical wave of vocal fold tissue during phonation using electroglottogram-triggered ultrasonography,” J. Acoust. Soc. Am. **143**(5), EL425–EL429 (2018).
27. B. Jing, P. Chigan, Z. Ge, L. Wu, S. Wang, and M. Wan, “Visualizing the movement of the contact between vocal folds during vibration by using array-based transmission ultrasonic glottography,” J. Acoust. Soc. Am. **141**(5), 3312–3322 (2017).
28. G. Luegmair, S. Kniesburges, M. Zimmermann, A. Sutor, U. Eysholdt, and M. Dollinger, “Optical reconstruction of high-speed surface dynamics in an uncontrollable environment,” IEEE Trans Med Imaging **29**(12), 1979–1991 (2010).
29. G. Luegmair, D. D. Mehta, J. B. Kobler, and M. Dollinger, “Three-Dimensional Optical Reconstruction of Vocal Fold Kinematics Using High-Speed Video With a Laser Projection System,” IEEE Trans Med Imaging **34**(12), 2572–2582 (2015).
30. M. Semmler, S. Kniesburges, V. Birk, A. Ziethe, R. Patel, and M. Dollinger, “3D Reconstruction of Human Laryngeal Dynamics Based on Endoscopic High-Speed Recordings,” IEEE Trans Med Imaging **35**(7), 1615–1624 (2016).
31. M. Semmler, M. Dollinger, R. R. Patel, A. Ziethe, and A. Schutzenberger, “Clinical relevance of endoscopic three-dimensional imaging for quantitative assessment of phonation,” Laryngoscope **128**(10), 2367–2374 (2018).
32. N. A. George, F. F. de Mul, Q. Qiu, G. Rakhorst, and H. K. Schutte, “Depth-kymography: high-speed calibrated 3D imaging of human vocal fold vibration dynamics,” Phys. Med. Biol. **53**(10), 2667–2675 (2008).
33. N. A. George, F. F. de Mul, Q. Qiu, G. Rakhorst, and H. K. Schutte, “New laryngoscope for quantitative high-speed imaging of human vocal folds vibration in the horizontal and vertical direction,” J. Biomed. Opt. **13**(6), 064024 (2008).
34. H. C. Hendargo, R. P. McNabb, A. H. Dhalla, N. Shepherd, and J. A. Izatt, “Doppler velocity detection limitations in spectrometer-based versus swept-source optical coherence tomography,” Biomed. Opt. Express **2**(8), 2175–2188 (2011).
35. American National Standards Institute (ANSI), American National Standard for the Safe Use of Lasers. American National Standard Institute, Inc., New York Standard Z136.1, 2000.
36. R. E. Hillman and D. D. Mehta, “The science of stroboscopic imaging,” in *Laryngeal Imaging: Indirect Laryngoscopy to High-Speed Digital Imaging*, K.A. Kendall and R. J. Leonard, eds. (Thieme Medical Publisher, Inc., 2010).
37. B. Grajciar, M. Pircher, A. Fercher, and R. Leitgeb, “Parallel Fourier domain optical coherence tomography for in vivo measurement of the human eye,” Opt. Express **13**(4), 1131–1137 (2005).
38. M. Mujat, N. V. Iftimia, R. D. Ferguson, and D. X. Hammer, “Swept-source parallel OCT,” Proc. SPIE **7168**, 71681E (2009).
39. D. J. Fechtig, T. Schmoll, B. Grajciar, W. Drexler, and R. A. Leitgeb, “Line-field parallel swept source interferometric imaging at up to 1 MHz,” Opt. Lett. **39**(18), 5333–5336 (2014).
40. D. J. Fechtig, B. Grajciar, T. Schmoll, C. Blatter, R. M. Werkmeister, W. Drexler, and R. A. Leitgeb, “Line-field parallel swept source MHz OCT for structural and functional retinal imaging,” Biomed. Opt. Express **6**(3), 716–735 (2015).
41. Olympus, “Laryngoscope,” http://medical.olympusamerica.com/products/rigid-laryngoscope.

65.215.34.66 {ts '2019-08-08 08:15:50'}