

Dynamic vocal fold imaging with combined optical coherence tomography/high-speed video endoscopy

Nicusor Iftimia¹, Gopi Maguluri¹, Ernest Chang¹, Jesung Ppark¹, James Kobler² and Daryush Mehta²

1. *Physical Sciences Inc., 20 New England Business Center, Andover, MA 01810*

2. *Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, 55 Fruit Street, Boston MA 02114*

1. Introduction

Voice disorders due to trauma (e.g., intubation injuries, vocal abuse) and disease (e.g., dysplasia, cancer, recurrent respiratory papillomatosis, nodules, polyps, and scar) are currently evaluated in otolaryngology clinics using endoscopic imaging techniques such as videostroboscopy [1] or high-speed videoendoscopy (HSV) [2]. Clinicians couple these visual observations of vocal fold tissue motion with auditory-perceptual judgments of voice quality as part of a comprehensive assessment of the health and function of the larynx during phonation [3]. Endoscopic imaging, however, only provides two-dimensional (2D) spatial information and thus only quantifies the lateral tissue motion, critically lacking vertical axis information. Various high-speed imaging techniques have attempted to capture vocal fold surface motion during phonation in three spatial dimensions [4–6]; however, they either lack adequate spatial resolution or have not been validated *in vivo*. In this paper, we present a dual modality imaging approach, where optical coherence tomography (OCT) imaging augments HSV by providing the missing depth information. OCT quantitatively measures the vertical motion of the vocal fold surface during phonation with micron scale resolution. Furthermore, it also provides the subsurface structural morphology to depth to at least 1.5 mm. Therefore, the combination of these two modalities within the same instrument seems to be a suitable approach for examining the pathology of the vocal folds.

2. Methods

A common optical imaging path swept source OCT/high speed video (SSOCT/HSV) endoscopy instrument was developed and used on an *ex vivo* study on excised animal tissue specimens. A simplified schematic of the instrument is shown in Figure 1.

The OCT component of the system uses a swept source (SS) approach. The light source (Santec) operates at a scan rate of 20 kHz, while providing a broad spectrum light with a 3dB bandwidth of 100 nm at a center wavelength of 1310 nm. This enables subsurface tissue imaging with an axial resolution better than 10 μm . A fiber optic interferometer and a balance detector are used to generate interference fringes, by combining the retro-reflected light from the imaged sample with that from a mirror, placed in the reference arm of the interferometer. A constructive interference occurs when the path-length difference between the two arms of the interferometer is within the coherence range of the light source. Each wavelength sweep of the light source is thus used to generate what is called an OCT A-scan. An individual A-scan is thus used to generate a sample reflectivity profile. The A-scan signals from the balanced detector are fed to a custom built FPGA module that acquires data when receives A-scan triggers from a custom-designed trigger circuit. These triggers are sent only when a signal from the vocal fold pressure sensor are detected, such that each full movement cycle of the vocal folds, also called phonation cycle, is digitized on a reasonable number of points ($N > 10$). The same triggers are fed to the HSV camera, enabling a perfect temporal correlation of the OCT and HSV data. The digitized signals are sent through a camera link interface to a frame grabber (NI

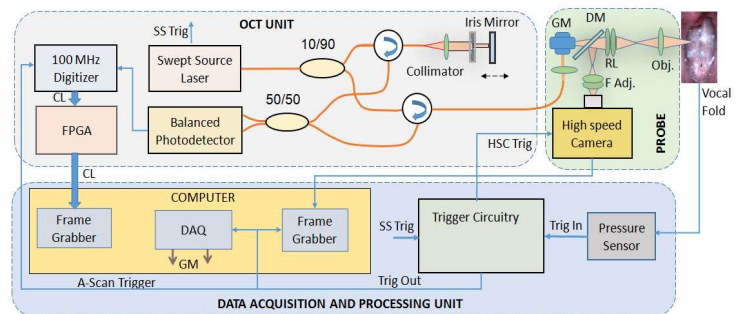


Figure 1. Simplified schematic of the SSOCT/HSV instrument

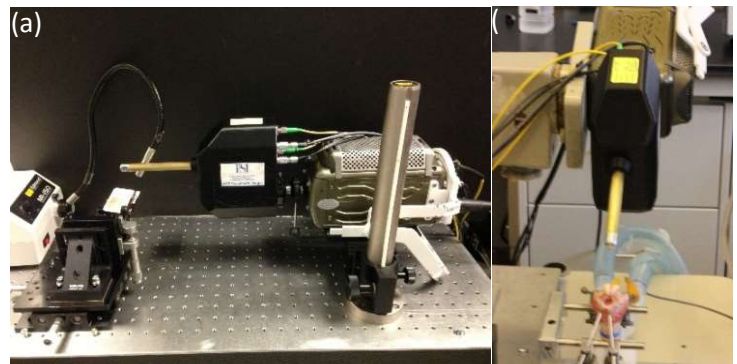


Fig. 2. OCT-HSV instrument. Setup showing (a) the OCT probe head mounted on the high-speed camera for co-linear imaging through a rigid endoscope and (b) the *ex vivo* calf larynx being imaged.

1430), and then the acquired data are transferred to a graphic processing unit, which enables real time processing and display the OCT images. The HSV component of the system constitutes a Phantom v7.3 color camera (Vision Research Inc., Wayne, NJ), typically set to record at 5000 frames per second at a spatial resolution of 128 x 128 pixels. A photograph of the experimental setup for imaging airflow-driven phonation of an ex vivo calf larynx is shown in Figure 2. Special attention has been paid to the temporal synchronization of OCT and HSV data during a phonation cycle. Therefore, we used a pressure sensor which detects the motion of the vocal folds, such the OCT A-scans and HSV images are synchronously acquired over the entire period of a single vibratory cycle. For a typical phonating frequency of 100 to 200 Hz, at least 10 OCT scans are acquired per each HSV frame. The 3D OCT data set provides the necessary lateral and depth information needed to correlate the HSV frames and augment them with the depth info, such that a 3D video image is reconstructed.

3. Results

Measurements on vibrating calf vocal folds were performed combining OCT and HSV data. Both static anatomical images and dynamic images (10 phases during a phonation cycles) were recorded. Figure 3 shows the 3D reconstruction of the vocal fold surface at rest with an axial area of 10 mm x 10 mm and imaging depth of 3.5 mm, while Figure 4 shows the vertical position of the vocal fold surface during a phonation cycle. The depth information provided by OCT was used to reconstruct the HSV data in a 3D format. A representative 3D HSV frame is shown in Fig. 5. The 3D coordinates are used to provide a 3D surface contour for the HSV frames to enable 4D visualization of vocal fold kinematics.

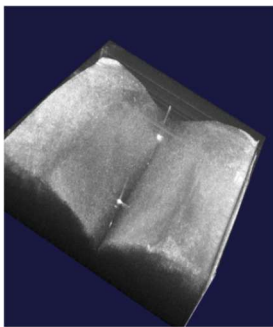


Fig. 3. 3D reconstruction of vocal fold integrating OCT and HSV images while the larynx is at rest.

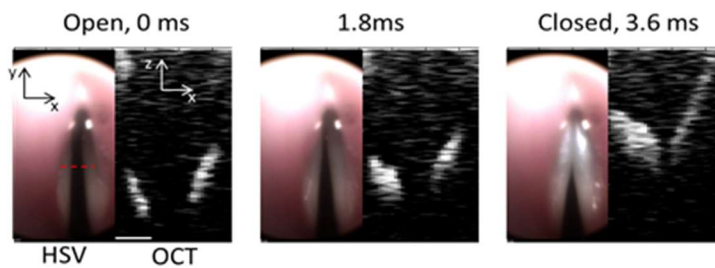


Fig. 4 Preliminary data displaying co-registered images acquired using swept-source OCT and HSV imaging modalities. The OCT beam scan path is indicated by a red dashed line. Scale bar: 1 mm.

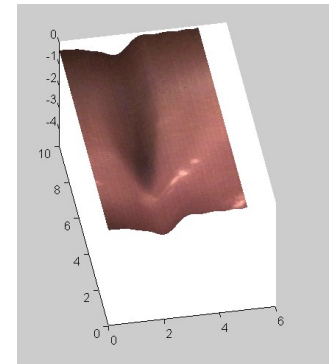


Fig.5. 3D HSV frame reconstructed based on OCT depth information

4. Conclusions

In conclusion, we have preliminarily tested a novel imaging modality combining optical coherence tomography (OCT) and high-speed videoendoscopy (HSV) to image laryngeal motion at high speeds and retrieve the 4D data showing the 3D movement of the vocal folds in time. This technology has the potential to enable a more in depth analysis of irregularities in vocal fold vibration. If successful, this technology will provide the otolaryngologists in the future with a novel tool for more reliable assessment of the vocal folds pathology and function.

5. References

- [1] D. D. Mehta and R.E. Hillman. "Current role of stroboscopy in laryngeal imaging," *Curr Opin Otolaryngol Head Neck Surg*, **20**, 429–436 (2012).
- [2] 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 Phoniatri Logop*, **60**, 33–44 (2008).
- [3] 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 Suppl 198**, 1–16 (2007).
- [4] C.G. Tsai, J.H. Chen, Y.W. Shau, T.Y. Hsiao, "Dynamic B-mode ultrasound imaging of vocal fold vibration during phonation," *Ultrasound Med Biol*, **35**, 1812–1818 (2009).
- [5] G. Luegmair, S. Kniesburges, M. Zimmermann, A. Sutor, U. Eysholdt, M. Doellinger, "Optical reconstruction of high-speed surface dynamics in an uncontrollable environment," *IEEE T Med Imaging*, **29**, 1979–1991 (2010).
- [6] D. E. Sommer, I.T. Tokuda, S.D. Peterson, K. Sakakibara, H. Imagawa, A. Yamauchi, T. Nito, T. Yamasoba, and N. Tayama, "Estimation of inferior-superior vocal fold kinematics from high-speed stereo endoscopic data in vivo," *J Acoust Soc Am*, **136**, 3290–3300 (2014).