Visualizing the mechanical wave of vocal fold tissue during phonation using electroglottogram-triggered ultrasonography

Bowen Jing, Zhengtong Ge, Liang Wu, Supin Wang, and Mingxi Wan[a)]
Department of Biomedical Engineering, The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an, Shaanxi, 710049, People's Republic of China
jingba@stu.xjtu.edu.cn, gezhengtong@stu.xjtu.edu.cn, liangwu@xjtu.edu.cn, spwang@mail.xjtu.edu.cn, mxwan@mail.xjtu.edu.cn

Abstract: In order to investigate the vibration pattern, especially the vibrational phase of tissue beneath the vocal fold mucosa, an imaging method called electroglottogram-triggered ultrasonography is proposed. The ultrasonic images of the vocal fold vibration are obtained in the coronal plane from five adult subjects during phonation. The velocity of the vocal fold tissue beneath the mucosal surface is obtained by using a motion estimation method. The results show that the vibration phase difference between tissues at different locations beneath the vocal fold mucosa results in a mechanical wave traveling upward at a speed of 720 to 1826 mm/s.

[CCC]
Date Received: March 25, 2018 **Date Accepted:** May 9, 2018

1. Introduction

The phase difference of the vocal fold vibration results in the wave-like motion of the vocal fold tissue especially on the vocal fold mucosal surface (Titze *et al.*, 1993; Zhang, 2016). Observation and quantification of this wave-like motion is essential for study of the physiology and pathology of voice production as well as assessment of vocal function (Wittenberg *et al.*, 2000). However, due to the limited imaging depth, only the mucosal surface can be observed by using widely applied optical imaging techniques such as the high speed video-endoscope. It is unachievable to investigate the vibration characteristics including the vibration phase of the tissue layers deep beneath the mucosa.

The image quality of the medical ultrasonic imaging techniques including the conventional B mode and the color Doppler imaging mode allows the researcher to visualize the vibration of the vocal fold tissue deep beneath the mucosa (Shau *et al.*, 2001; Qin *et al.*, 2011; Tang *et al.*, 2013). However, these conventional ultrasonic imaging techniques are all based on the focused beam line-by-line scanning scheme. It means that the low imaging frame rate (less than 1000 frames/s) and the time lag between scanlines make it hardly achievable to use these techniques to investigate the vibration pattern, especially the vibrational phase difference between tissues at different locations beneath the vocal fold mucosa. As vibration frequency of the vocal folds could range from 72 to 1000 Hz, the minimum frame rate required is 4000 frames/s to provide sufficient sampling of the vibrational phases (Deliyski *et al.*, 2015). Therefore, ultrafast plane wave ultrasonography (5000 frame/s) has been introduced to record the vibration (Jing *et al.*, 2016). However, the poor image quality, which is because of the unfocused emitting ultrasound beam and the low echogenicity of the vocal fold tissue, makes it hardly achievable to use this technique to observe and quantify the vibration of the vocal fold tissue (Jing *et al.*, 2016).

In order to investigate the vibration pattern, especially the vibrational phase difference between tissues at different locations beneath the vocal fold mucosa, an imaging method called electroglottogram (EGG)-triggered ultrasonography is proposed. The focused beam-emitting scheme, the image quality of which has been shown sufficient for imaging of vocal fold tissue in previous reports (Qin *et al.*, 2011; Tang *et al.*, 2013), is employed in the proposed imaging method. More importantly, an EGG-triggered line-by-line scanning scheme is employed to sample the vocal fold

[a)]Author to whom correspondence should be addressed.

tissue vibration at 4000 Hz and obtain the column-wise reconstructed ultrasonic images of the tissue. Moreover, a motion estimation algorithm is used for quantifying the tissue velocity and visualizing the motion of the vocal fold tissue.

2. Method

The EGG-triggered ultrasonography is based on the triggered line-by-line scanning scheme [Fig. 1(a)]. The field of view is divided into multiple columns, each of which is insonified and scanned by using a focused narrow beam to obtain a scanline of ultrasound radio frequency data. The ultrasound probe used is a linear array transducer (L14-5/38, Ultrasonix, Richmond, BC, Canada). A single-cycle ultrasound pulse with a center frequency of 10 MHz is used to insonify the tissue. According to the dimension of the transducer array used, the vertical resolution could reach 0.3 mm. According to the emitting ultrasound frequency, the horizontal resolution could reach 0.08 mm. The trigger pulse is generated by using an oscilloscope (TDS724C, Tektronix, Beaverton, OR) thresholding the periodic EGG signal at the closing phase (the rising edge) of each vibration cycle [Fig. 1(a)]. Upon receiving the trigger pulse, an arbitrary waveform generator (DG1022U, RIGOL Technology, Beijing, China) sends a pulse train to trigger a medical ultrasonic scanner (SonixTouch, Ultrasonix) to scan a column of tissue at 4000 Hz. Once the scan of the current column is finished and the next electroglottographic trigger pulse arrives, the scanner starts to scan the next neighboring column. This procedure continues until the data of the whole field of view are obtained. It should be noted that the waveform generator and the ultrasound scanner do not respond to trigger pulses during scanning until the scan of the current column is finished. Therefore, only the trigger pulses to which the scanner responds are valid and illustrated in Fig. 1(a). Images of the entire field of view can be reconstructed by re-grouping the scanlines with respect to vibration phases according to the EGG [Fig. 1(a)]. The frame rate of the reconstructed images is 4000 frames/s as the scanline data are obtained at 4000 Hz. However, since the entire field of view is divided into 128 columns and each column is scanned for 11 ms that is longer than a vocal fold vibration cycle, it costs at least 1408 ms to obtain the image of the entire field of view.

The image of the vocal fold is obtained in the coronal plane [Fig. 1(b)]. By manually adjusting the position and orientation of the ultrasound probe, the position of the imaging plane is chosen approximately at the mid portion along the longitudinal direction of the vocal fold. During imaging, 5 volunteers (3 males and 2 females of age from 23 to 28 with no vocal disorder) are asked to maintain a steady-state phonation of vowel /u:/ at their comfort pitch and loudness for 3 s in order to record a quasi-periodic vibration of the vocal folds and also ensure there is sufficient time for a scan to be finished. The normalized cross correlation method, which has been widely used for tissue displacement estimation in ultrasonic elastography, is used to estimate the vocal fold tissue velocity along the horizontal direction (Pinton *et al.*, 2006). The reason that only the horizontal velocity is obtained in the current study is that the vertical velocity is too low to obtain using this motion estimation algorithm under current limited vertical imaging resolution.

3. Results

3.1 Ultrasonic images of vocal fold vibration

The images obtained by using the EGG-triggered ultrasonography is shown in Fig. 2(a). In order to show the images across the whole vibration cycle at a limited page length, the frame rate of the images in Fig. 2(a) has been decimated from 4000 to 1000 frames/s. By comparing the images with the anatomy of the larynx [Fig. 1(b)], edges (mucosal

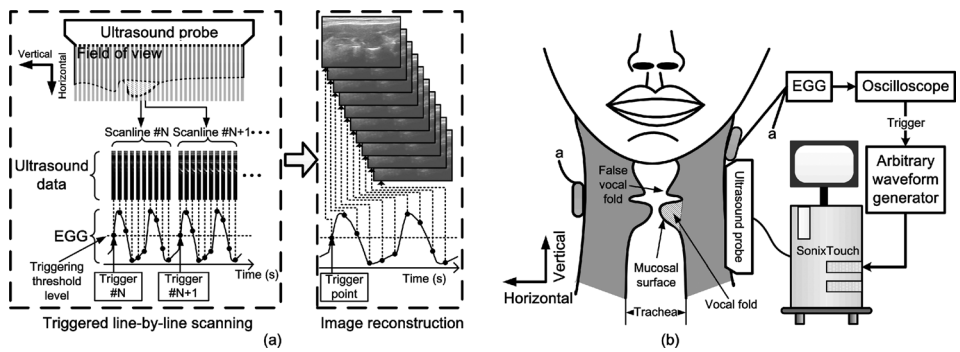

Fig. 1. (a) Triggered line-by-line scanning scheme and (b) the imaging system with an illustration of the vocal folds in the coronal plane.

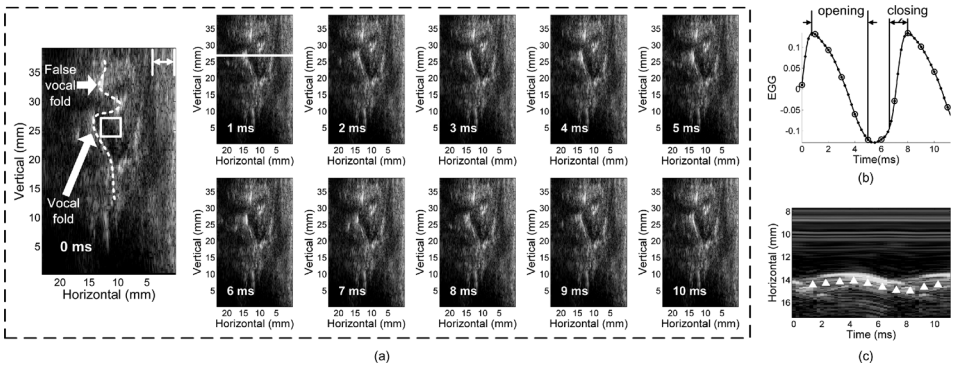

Fig. 2. (a) The ultrasonic images of the vocal fold during phonation (down-sampled to 1000 frames/s). The long and short arrows in the image captured at 0 ms point out the edges (mucosal surface) of the vocal fold and the false vocal fold, respectively. The dashed curve indicates the profile of the airway. The white rectangle indicates the ROI where the vocal fold tissue velocity is quantified and shown in Fig. 3(a). The double-headed arrow indicates the layered structure of the neck skin. The horizontal line in the image captured at 1 ms indicates the position where the M-mode image in (c) is generated. (b) The simultaneously recorded EGG signal. The opening and closing phase of the vibration is identified in (b). The black dots on the EGG waveform indicate the acquisition moment of ultrasonic images (4000 frames/s) while the black circles indicate the acquisition moment of the images in (a). (c) The M-mode image of the vibrating vocal fold. The white arrowheads indicate the mucosal surface of the vocal fold during vibration.

surface) of the vocal fold and the false vocal fold can be identified as two short hyperechoic (i.e., strong ultrasound echo intensity) curves in 0 ms image in Fig. 2(a), which is consistent with the results obtained in previous studies (Hu *et al.*, 2010; Jing *et al.*, 2016). The strong echo intensity at the edges and reverberation artifacts near the edges indicates that the ultrasound beam is fully reflected at the air-mucosa interface. It also indicates that the ultrasound beam is transmitted effectively through the thyroid cartilage as the laryngeal structures below the cartilage are visible. Besides, only one of the paired vocal folds is visible in the images since the ultrasound cannot be transmitted through the air between the paired vocal folds.

The fundamental frequency of the voice recorded is from 213 to 263 Hz for female subjects and from 118 to 156 Hz for males. The images in Fig. 2(a) are captured when phonation occurs, as is indicated by the EGG waveform shown in Fig. 2(b). The opening and closing phase of the glottis could be identified in the EGG waveform. The black dots (i.e., the moments of image acquisition) in the EGG waveform indicate that the rapidly-closing phase of the glottis, which is the steep rising edge in the EGG waveform, is sufficiently sampled by the EGG-triggered ultrasonography working at 4000 frames/s. This is consistent with the recommendation that the frame rate should be high enough to ensure the acquisition of at least 16 frames per vibration cycle to warrant sufficient representation of the glottal phases (Deliyski *et al.*, 2015). Moreover, the spatial location, brightness, and shape of the echoic speckles at the edge of the vocal fold changes constantly during phonation [Fig. 2(a)]. In order to give a clearer view of the vibration, the M-mode image of the vibrating vocal fold is shown in Fig. 2(c). The M-mode image, which is available in most medical ultrasound scanners, is obtained by grouping scanline data of a single tissue column into a single two-dimensional image. The edge of the vocal fold can be identified (pointed out by a series of white arrowheads). In Fig. 2(c), the spatial location of the vocal fold edge changes, which forms a hyperechoic curve. This result further indicates that the vibration of the vocal fold can be visualized using the proposed imaging method.

3.2 Visualizing the mechanical wave of vocal fold tissue

Although Fig. 2(c) indicates that the vocal fold edge (hyperechoic mucosal surface) is vibrating during phonation, it is unknown how the vocal fold tissue beneath the mucosal surface is involved in the vibration. In the present study, the velocity of the vocal fold tissue which lies more than 1 mm beneath the vocal fold mucosal surface [region of interest (ROI) indicated by the white rectangle in 0 ms image in Fig. 2(a)] is estimated by using the normalized cross correlation method. The ROI is a fixed rectangle region manually selected at 1 mm beneath the hyperechoic mucosal surface during the glottal open phase. However, as the glottis closes, the spatially fixed ROI turns to be more than 1 mm beneath the mucosal surface.

It should be noted that the velocity obtained here is only the horizontal component of the velocity vector. The velocity maps of tissue in the ROI, which are shown in Fig. 3(a), indicate that the maximum velocity of the tissue at the left part of the

ROI is higher than that of the tissue at the right part. This result is consistent with the common knowledge that the vibration of the tissue closer to the vibrating medial edge, which is the hyperechoic curve indicated by the long arrows on the left of the ROI as shown in the 0 ms image in Fig. 2(a), is faster than that of the tissue that is further away from the vibrating medial edge and closer to the fixed lateral edge. Besides, the maximum horizontal velocity of the tissue in the ROI could reach more than 100 mm/s during vibration.

More importantly, the results in Fig. 3(a) show that, near the left boundary of the ROI, there is a mechanical wave traveling upward (from the inferior to the superior part of the vocal fold). In order to give a clearer view of this mechanical wave, the spatio-temporal velocity map of a vertical column of tissue near the left boundary of the ROI is shown in Fig. 3(b). The color of pixels in Fig. 3(b) still indicates the velocity of the tissue in the same manner as Fig. 3(a) shows. The slant stripe pattern in Fig. 3(b) indicates that there is a vibration phase difference between the tissues at different vertical locations. The vibration of the tissue at the lower location is earlier than that of the tissue at the upper location. Besides, the vertical speed of the mechanical wave, which can be directly estimated from Fig. 3(b) by finding the zero-crossing points and using the time-of-flight method, is approximately 1826 mm/s. Preliminary results obtained from the five subjects show that the vertical speed of the mechanical wave, which travels at more than 1 mm beneath the vocal fold mucosal surface, ranges from 720 to 1826 mm/s during phonation. This speed is on the same order as the speed of the wave on the medial mucosal surface that reaches 1200 mm/s in a previous study using the excised human hemilarynx setup (Doellinger and Berry, 2006).

4. Discussion and conclusion

The *in vivo* results show that the vibration of the vocal fold can be identified in the high frame rate ultrasonic images obtained. However, it should be noted that these images are column-wise reconstructed ultrasonic images rather than instantaneous snapshots of the vocal folds. The velocity map of the vocal fold tissue beneath the mucosal surface, which is obtained by using the motion estimation method, shows that the vibration phase difference between tissues at different locations results in a mechanical wave traveling upward beneath the vocal fold mucosal surface. A previous study (Pinton *et al.*, 2006) has shown that the motion estimation algorithm used in the present study can be used to detect tissue inter-frame displacement smaller than 0.005 mm with an error smaller than 0.0005 mm at the present ultrasound data sampling rate (40 MHz), which indicates that it can be used to detect tissue velocity as low as 20 mm/s with an error smaller than 2 mm/s at a frame rate of 4000 Hz. The horizontal velocity of the tissue at more than 1 mm beneath the mucosal surface is much lower than the velocity of the vibrating medial mucosal surface which could reach 1340 mm/s as shown in a previous study (Doellinger and Berry, 2006). Besides, the results in the present study are obtained at the mid-coronal section along the longitudinal direction of the vocal fold.

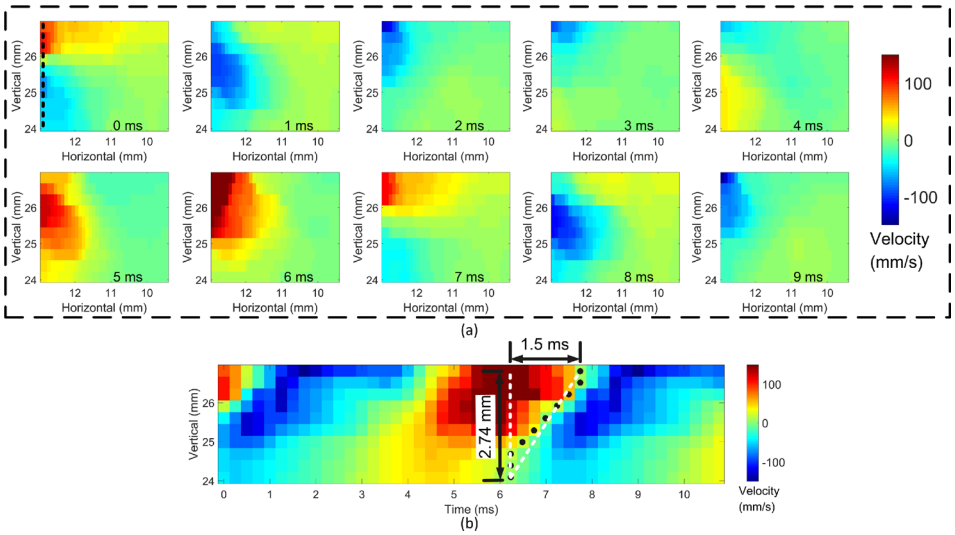

Fig. 3. (Color online) (a) A series of frames of velocity maps of the vocal fold tissue in the ROI. (b) The spatio-temporal velocity map of a vertical column of vocal fold tissue. The vertical dashed line in 0 ms image in (a) indicates the tissue where the spatio-temporal velocity map in (b) is generated. The black dots in (b) indicate the zero crossing points from which the speed of the mechanical wave is estimated using the time-of-flight method.

It is reasonable to make a hypothesis that there is a variation of the vibration pattern along the longitudinal direction. This longitudinal variation remains to be investigated in the next step of work by using the imaging technique proposed in the present report.

In the present study, the measurement of the motion of the mucosal layer is not achievable using the proposed imaging method due to the strong reverberation artifact and severe decorrelation problem that have also been encountered in previous studies (Tang *et al.*, 2013; Jing *et al.*, 2016). Fortunately, this limitation has been overcome by using the array-based transmission ultrasonic glottography (Jing *et al.*, 2017). In future work, the EGG-triggered ultrasonography should be combined with the array-based transmission ultrasonic glottography for simultaneously visualizing and investigating the mechanical wave traveling on and beneath the mucosal layer to show a comprehensive view of vibration of the multi-layered vocal fold.

Another major limitation of the EGG-triggered ultrasonography is that the EGG-triggered line-by-line scanning scheme requires the subjects to produce a steady-state phonation during recording. Therefore, investigation of the unsteady phonation such as onset and offset is not achievable using the proposed imaging method. However, in the present study, the results revealing the vibration phase difference and the mechanical wave, which are obtained from steady-state phonation, indicate that the EGG-triggered ultrasonography provide a new angle of view for investigating the vibration pattern of the tissue beneath the vocal fold mucosal surface.

Acknowledgments

This work is supported by the National Natural Science Foundations of China under Grant Nos. 11404256 and 81771854, and by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2016JQ2017.

References and links

Deliyski, D. D., Powell, M. E. G., Zacharias, S. R. C., Gerlach, T. T., and de Alarcon, A. (**2015**). "Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment," Biomed. Signal Process. **17**, 21–28.

Doellinger, M., and Berry, D. A. (**2006**). "Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation," J. Voice **20**, 401–413.

Hu, Q., Zhu, S., Luo, F., Gao, Y., and Yang, X. (**2010**). "High-frequency sonographic measurements of true and false vocal cords," J. Ultrasound Med. **29**, 1023–1030.

Jing, B., Chigan, P., Ge, Z., Wu, L., Wang, S., and Wan, M. (**2017**). "Visualizing the movement of the contact between vocal folds during vibration by using array-based transmission ultrasonic glottography," J. Acoust. Soc. Am. **141**, 3312–3322.

Jing, B., Tang, S., Wu, L., Wang, S., and Wan, M. (**2016**). "Visualizing the vibration of laryngeal tissue during phonation using ultrafast plane wave ultrasonography," Ultrasound Med. Biol. **42**, 2812–2825.

Pinton, G. F., Dahl, J. J., and Trahey, G. E. (**2006**). "Rapid tracking of small displacements with ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1103–1117.

Qin, X., Wu, L., Jiang, H., Tang, S., Wang, S., and Wan, M. (**2011**). "Measuring body-cover vibration of vocal folds based on high-frame-rate ultrasonic imaging and high-speed video," IEEE Trans. Biomed. Eng. **58**, 2384–2390.

Shau, Y., Wang, C., Hsieh, F., and Hsiao, T. (**2001**). "Noninvasive assessment of vocal fold mucosal wave velocity using color Doppler imaging," Ultrasound Med. Biol. **27**, 1451–1460.

Tang, S., Zhang, Y., Qin, X., Wang, S., and Wan, M. (**2013**). "Measuring body layer vibration of vocal folds by high-frame-rate ultrasound synchronized with a modified electroglottograph," J. Acoust. Soc. Am. **134**, 528–538.

Titze, I. R., Jiang, J. J., and Hsiao, T. (**1993**). "Measurement of mucosal wave-propagation and vertical phase difference in vocal fold vibration," Ann. Oto. Rhinol. Laryn. **102**, 58–63.

Wittenberg, T., Tigges, M., Mergell, P., and Eysholdt, U. (**2000**). "Functional imaging of vocal fold vibration: Digital multislice high-speed kymography," J. Voice **14**, 422–442.

Zhang, Z. (**2016**). "Mechanics of human voice production and control," J. Acoust. Soc. Am. **140**, 2614–2635.