Ultrasound in Med. & Biol., Vol. 42, No. 12, pp. 2812–2825, 2016

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2016.07.023

● *Original Contribution*

VISUALIZING THE VIBRATION OF LARYNGEAL TISSUE DURING PHONATION USING ULTRAFAST PLANE WAVE ULTRASONOGRAPHY

Bowen Jing, Shanshan Tang, Liang Wu, Supin Wang, and Mingxi Wan

Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China

(*Received* 7 *March* 2016; *revised* 12 *July* 2016; *in final form* 22 *July* 2016)

Abstract—Ultrafast plane wave ultrasonography is employed in this study to visualize the vibration of the larynx and quantify the vibration phase as well as the vibration amplitude of the laryngeal tissue. Ultrasonic images were obtained at 5000 to 10,000 frames/s in the coronal plane at the level of the glottis. Although the image quality degraded when the imaging mode was switched from conventional ultrasonography to ultrafast plane wave ultrasonography, certain anatomic structures such as the vocal folds, as well as the sub- and supraglottic structures, including the false vocal folds, can be identified in the ultrafast plane wave ultrasonic image. The periodic vibration of the vocal fold edge could be visualized in the recorded image sequence during phonation. Furthermore, a motion estimation method was used to quantify the displacement of laryngeal tissue from hundreds of frames of ultrasonic data acquired. Vibratory displacement waveforms of the sub- and supraglottic structures were successfully obtained at a high level of ultrasonic signal correlation. Moreover, statistically significant differences in vibration pattern between the sub- and supraglottic structures were found. Variation of vibration amplitude along the subglottic mucosal surface is significantly smaller than that along the supraglottic mucosal surface. Phase delay of vibration along the subglottic mucosal surface is significantly smaller than that along the supraglottic mucosal surface. (E-mail: mxwan@mail.xjtu.edu.cn or jingba@stu.xjtu.edu.cn) © **2016 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **High frame rate, Larynx, Motion estimation, Phonation, Ultrafast plane wave ultrasonography, Vibration.**

INTRODUCTION

The study of the physiology and pathology of phonation relies heavily on observation techniques for visualizing and recording the vibration of laryngeal tissue, especially the vocal fold (Doellinger et al. 2011; Krausert et al. 2011; Shaw and Deliyski 2008). To observe the vibration during phonation directly, high-frame-rate optical imaging techniques including high-speed video endoscopy (Booth and Childers 1979; Mergell et al. 2000; Tao et al. 2007), for which the frame rate is >2000 frames/s, and videokymography (Schutte et al. 1998; Svec and Schutte 1996), for which the frame rate reaches 8000 frames/s, have been widely applied for clinical and research purposes. However, when these techniques are used, only the suprasurface of the glottis is viewed at the horizontal plane when live human subjects are involved, because endoscopes are inserted via the oral or nasal cavity and placed above the glottis. Even if the newly developed technique based on optical coherence tomography (several tens of frames per second) is used, it is still difficult to image tissue more than 2 mm deep below the mucosal surface (Kobler et al. 2010; Yu et al. 2009). On the other hand, extensive studies (Alipour et al. 2000; Story and Titze 1995; Zhang 2009) on modeling the vibratory behavior of the multilayered structure of the vocal folds indicate that it is necessary to investigate how the structures beneath the mucosal surface are involved in the vibration under various physiologic and pathologic conditions. To observe and record the vibration of tissue beneath the surface without either interfering with voice production or exposing subjects to unnecessary radiation, B-mode ultrasonography has been introduced (Qin et al. 2011; Tang et al. 2013; Tsai et al. 2009).

It should be noted that there are phase differences between the vibrations of laryngeal tissues at different

Address correspondence to: Mingxi Wan, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: mxwan@mail.xjtu.edu.cn or jingba@stu.xjtu.edu.cn

locations during voice production, and the phase difference varies under different vibration patterns (Bailly et al. 2014; Nasri et al. 1996; Sakakibara et al. 2001; Story and Titze 1995; Titze et al. 1993). Therefore, quantifying the vibration phase is significant in the comprehensive investigation of vibration patterns of laryngeal tissue under various physiologic and pathologic conditions. However, little information on the vibration phase of laryngeal tissue has been obtained in the previous studies because of the limitation of the applied ultrasonic imaging technique. The technique used in previous studies is generally considered as conventional ultrasonography, in which each frame recorded is constructed by multiple scanlines acquired at different moments within one frame period (Tanter and Fink 2014). The major limitation of conventional ultrasonography is its low imaging frame rate. B-Mode ultrasonography used by Tsai et al. (2009) works at 40 frames/s, which is lower than the vibration frequency of the larynx, which ranges from 72 to 1000 Hz, as reported in a previous study (Deliyski et al. 2015). In our previous studies (Qin et al. 2011; Tang et al. 2013), by drastically narrowing the field of view (FOV) to 7.6 mm in width, the frame rate of ultrasonography was increased to 800 frames/s. However, it is still not high enough to provide sufficient sampling of the vibration phases. Moreover, quantification of the phase difference between tissues at different locations is also impeded by the non-negligible time lag between acquisitions of the scanlines.

Ultrafast plane wave ultrasonography has been successfully used for visualization and measurement of tissue vibration in the form of artificially induced shear waves (Sandrin et al. 1999, 2002) or natural waves such as the arterial pulse wave (Couade et al. 2011). The application of this technique for imaging of laryngeal vibration was reported in our preliminary study (Jing et al. 2015). The major advantage of this technique is that, without sacrificing the size of the FOV, the frame rate could reach thousands of frames per second, which is much higher than the frame rate of conventional ultrasonography. Thus, it ensures a sufficient sampling of the laryngeal vibration. Moreover, the acquisition lag between scanlines is eliminated, and a near-instantaneous snapshot of tissue movement can be obtained. The advantages mentioned above suggest that ultrafast plane wave ultrasonography could potentially be used to quantify the vibration phase as well as the amplitude of laryngeal tissue.

One potential problem of ultrafast plane wave ultrasonography is that the high frame rate ($\leq$10,000 frames/s) is usually achieved with a compromise of image quality, as reported in the previous study (Montaldo et al. 2009). Therefore, it is necessary to elucidate the feasibility of applying this technique for visualization and quantification of the vibration of the larynx. In the present study, ultrafast plane wave ultrasonography is employed to visualize the vibration of laryngeal tissue. The ultrafast plane wave imaging mode was programmed on an ultrasound scanner. The plane wave ultrasonic images of the larynx were acquired in the coronal plane at the same location as the images acquired using conventional ultrasonography in previous studies (Qin et al. 2011; Tang et al. 2013; Tsai et al. 2009). The difference between the images obtained with the two imaging techniques was addressed when we tried to identify certain laryngeal structures from the ultrafast plane wave ultrasonic images. As only one of the paired vocal folds can be visualized because the trachea lumen lies between the vocal folds, the vibratory status of the glottis cannot be directly visualized in the ultrasonic image. Therefore, the electroglottography (EGG) signal, which served as an indicator of the vibration status of the glottis, was acquired simultaneously. Because low echogenicity (*i.e.*, low signal-to-noise ratio) could result in poor estimation of tissue displacement, the echogenicity of different parts of the larynx was evaluated under different emitting frequencies. Furthermore, we tried to quantify the displacement of the vocal fold body and the vocal fold edge, as well as sub- and supraglottic structures, using an ultrasonic radiofrequency (RF) data–based motion estimation algorithm. Therefore, we could further obtain the amplitude and the phase difference of vibration of tissues at multiple locations in the larynx. Moreover, the difference in vibration patterns between different laryngeal structures was investigated in this study.

METHODS

Acquisition of ultrasonic images

The ultrasonic imaging probe used in the present study is a linear array transducer (L14-5/38, Ultrasonix, Richmond, BC, Canada) designed for imaging of small structures in the head and neck regions. The ultrafast plane wave ultrasonic imaging sequence was programmed to work on an ultrasound scanner (SonixTouch, Ultrasonix, Richmond, BC, Canada). Emission of an ultrasonic plane wave was achieved by triggering all the elements of the transducer array at the same time to insonify the whole FOV. In the meantime, ultrasound echoes were received by each element of the linear array transducer and digitized by a 128-channel parallel data receiver (SonixDAQ, Ultrasonix). In the full imaging depth, the gain of the data receiver remained fixed at the maximum. Then, the digitized RF ultrasound echo was used as the input of a delay-and-sum beamformer to generate beamformed RF data. The Hilbert transform was used to produce the envelope of the beamformed RF data.

Gray-scale ultrasonic images were obtained by adjusting the dynamic range of the envelope data.

It should be noted here that each frame of ultrasonic images was acquired after a single emission of ultrasonic plane waves. The imaging frame rate is higher than the minimum frame rate requirement (4000 frames/s) for voice assessment (Deliyski et al. 2015). As the inter-frame displacement of vibrating laryngeal tissue could be as large as 0.2 mm (5000 frames/s) because of the high velocity of the vibration (*i.e.*, above 1 m/s), coherent plane wave compounding could result in additional severe motion blurring of ultrasonic images, as described in a previous study (Denarie et al. 2013). Thus, no coherent compounding method was used in the present study. The parameter settings of the ultrasonic imaging system are listed in Table 1.

Experiment setup

The probe was placed on the neck skin surface at the level of the glottis. Images of the larynx were obtained in the coronal plane so that the FOV could cover the glottis as well as sub- and supraglottic structures, including the false vocal folds (Fig. 1a). One problem that may arise in the use of ultrasonography for laryngeal imaging is the degree to which the ultrasound beam is reflected by the thyroid cartilage. This effect is a function of the degree to which the cartilage in question resembles bone. Ossification of the thyroid cartilage is commonly observed in older adults (Bozzato et al. 2007; Hu et al. 2010). Actually, we also conducted a pilot study on a 54-year-old man with significant ossification to test the degree to which our method is affected by ossification. The results confirmed that ultrasonography using our technique is negatively affected by ossification. Therefore, to minimize the likelihood of ossification, all participants recruited for this study were younger than 30 years of age. Ten volunteers (4 women and 6 men, aged 23 to 30 years) without any history of laryngeal disorders were recruited into the experiment.

Table 1. Parameter settings of the ultrasonic imaging system

Parameter	Setting
Probe	Linear array transducer
Number of elements	128
Pitch, mm	0.3048
Emitting frequency, MHz	5, 8, 10, 13.3
Probe center frequency, MHz	7.2
Frame rate, frames/s	5000, 10,000
Number of pulse cycles	1
Sampling rate, MHz	80
Beamformer	Delay and sum (Hanning window apodization)
F number for dynamic receive	2
Dimensions of field of view, mm	
Width	38
Depth	25

During the experiment, participants were instructed to breathe quietly for a short period and then to begin producing sustained/u/at a comfortable pitch and loudness for several seconds. The ultrasonic images were acquired at 5000 frames/s at the same time to record the onset of laryngeal vibration. In an additional experiment, the frame rate was increased to 10,000 frames/s.

The electroglottograph, an instrument used to non-invasively register glottal vibratory movements by measuring the electrical impedance variations across the glottis (Childers et al. 1986), was employed in this study as a reference. The EGG (Model EG2-PC, Glottal Enterprises, Syracuse, NY, USA) signal was acquired at 50 kHz using a data acquisition module (GAGE 14100, Gage Applied Technologies, Montreal, QC, Canada). Acquisitions of ultrasonic images and EGG signals were started simultaneously using the same trigger pulse to activate both the ultrasonic scanner and the data acquisition module. This study was approved by the Biomedical Ethics Committee of Xi'an Jiaotong University, and signed informed consent forms were obtained from all participants.

Quantification of vibration amplitude and phase difference

First, a motion estimation algorithm (Walker and Trahey 1995) based on normalized cross-correlation was used to estimate inter-frame displacement of sub- and supraglottic tissues from consecutive frames of beamformed RF data. The size of the sliding kernel window is 0.29 mm in the horizontal direction (axis Z in Fig. 1a) and 0.3 mm in the vertical direction (axis X in Fig. 1a). The size of the post-frame searching window is 0.58 mm in the horizontal direction and 0.3 mm in the vertical direction. The results of measurement of sub- and supraglottic vibrations using this set of kernels are illustrated in Figures 2–6. The estimation is carried out every 0.1 mm along the horizontal direction and every 0.3 mm along the vertical direction. The RF signal segments are upsampled by a factor of 20 in the horizontal direction before calculation of the normalized cross-correlation function. Note that only the tissue displacement in the direction of the ultrasound beam is estimated using this algorithm. This means only the vibration along the horizontal direction (indicated by the axis Z in Fig. 1a) is measured.

Furthermore, the sliding kernel window and post-frame searching window used for motion estimation were expanded to 0.58 mm in the horizontal direction, 0.9 mm in the vertical direction and 1.45 mm in the

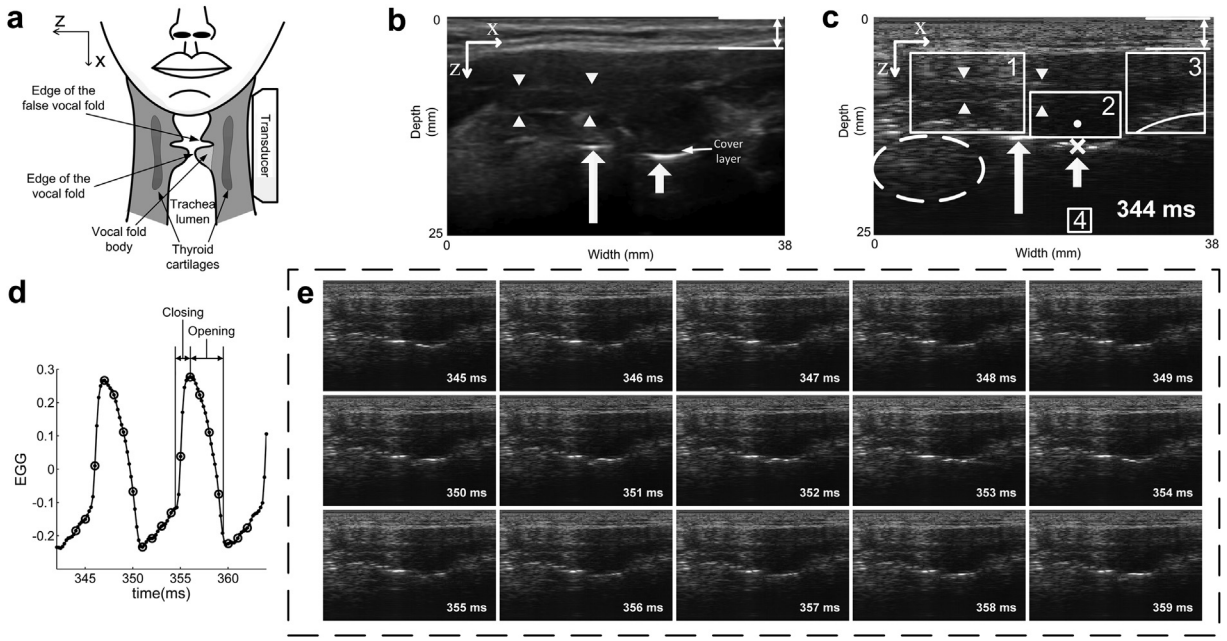

Fig. 1. Setup for imaging of the larynx and the acquired ultrasonic images. (a) Relative position of the ultrasound transducer and laryngeal structures. The horizontal and vertical directions are referenced to the subject. The horizontal direction is indicated by the z-axis. The vertical direction is indicated by the x-axis. (b) Frame of conventional B-mode ultrasonic images of the region under investigation. The location of the hypo-echoic cover layer is pointed out in the image. (c) Identification of the laryngeal structures in the ultrafast plane wave ultrasonic image. Four *arrowheads* in (b) and (c) indicate the location of the thyroid cartilage. The *long arrow* and *short arrow* point to the edges of the false vocal fold and the vocal fold, respectively. The *double-headed arrow* points to the neck skin. Reverberation artifacts in the trachea lumen are indicated by the *dashed ellipse*. The supraglottic region, the vocal fold body, the subglottic region and the region where the noise floor is estimated are indicated by white rectangles numbered 1, 2, 3 and 4 respectively in (c). The curve near the bottom of the region 3 indicates the location of the subglottic mucosal surface. The *white cross mark* in (c) is at the vocal fold edge. The *white round dot* right above the cross mark is in the vocal fold body. (d) Short segment of electroglottography (EGG). (e) Simultaneously acquired ultrasonic images (decimated to 1000 frames/s). The closing and opening phases are pointed out in (d). The *black dots* in (d) indicate the moment of acquisition of ultrasonic images, and the *black circles* indicate the moment of acquisition of the images in (e).

horizontal direction, 1.8 mm in the vertical direction, respectively, to track and estimate the vibration displacement of the vocal fold edge and body. As the vibration velocity of the vocal folds could reach 1.34 m/s (Doellinger and Berry 2006), the windows were so expanded in case there was any large out-of-window displacements causing erroneous estimation. The results of measurement of vocal fold vibration using this set of kernels are illustrated in Figures 7 and 8. The RF signal segments are upsampled by a factor of 20 in the horizontal direction and 10 in the vertical direction before calculation of the normalized cross-correlation function. The motion estimation method used in this study is based on the basic principle that inter-frame tissue displacement is estimated by searching for the best-matched pair of signal segments between two neighboring frames. Severe signal decorrelation is usually one of the major causes of poor estimation of tissue displacement (Ramamurthy and Trahey 1991). Therefore, to monitor this decorrelation effect, the normalized cross-correlation coefficient between the displaced and best-matched reference signal segments, which serves as an indicator of signal decorrelation, is also provided in the results. Moreover, large inter-frame displacement of vocal fold tissue, which is due to the high velocity (*i.e.*, above 1 m/s) of the vibration of the vocal fold, could potentially lead to severe signal decorrelation and therefore compromise the performance of the motion estimation method. To minimize the likelihood of severe signal decorrelation during the estimation of vocal fold displacement, an additional experiment was conducted to minimize the inter-frame displacement of fast vibrating vocal fold tissue by increasing the frame rate to 10,000 frames/s. Because of the limitation of the current hardware, the frame rate was not tripled to 15,000 frames/s. Thus, we were able to test if the estimation of vocal fold displacement was improved by significantly increasing the imaging frame rate.

The displacement waveform is obtained by cumulatively summing the inter-frame displacement with respect to time. Before quantification of vibration amplitude and

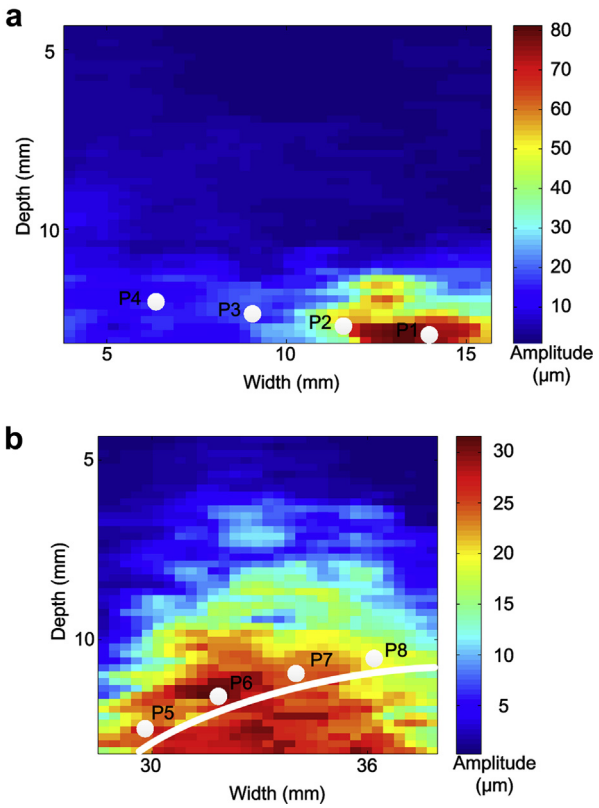

Fig. 2. Vibration amplitude of tissue in the (a) supraglottic region (region 1 in Fig. 1c) and (b) subglottal region (region 3 in Fig. 1c). The *white curve* in (b) indicates the position of the subglottic mucosal surface. Locations of interest are indicated by P1, P2, P3, P4, P5, P6, P7 and P8 in (a) and (b).

the phase difference, a high-pass zero-phase digital filter (−30 dB at 10 Hz, −1 dB at 55 Hz) was employed to remove the low-frequency drift from the vibration displacement waveform. After the vibration displacement waveform is obtained, the vibration amplitude of laryngeal tissue is also obtained. The phase difference between the vibrations of tissues at two neighboring locations is quantified using a time delay estimator based on normalized cross-correlation analysis. There is no upsampling of the displacement waveform before the normalized cross-correlation analysis. The displacement waveform segment, including the onset of the vibration, is used for the normalized cross-correlation analysis to determine the time delay (*i.e.*, the phase delay).

RESULTS

Imaging of the larynx

Images of the larynx obtained during phonation using ultrafast plane wave ultrasonography are provided in Figure 1c and e. Figure 1b is a frame of the conventional ultrasonic image acquired at the same anatomic location. Identification of laryngeal structures has been achieved using conventional ultrasonography in previous studies (Hu et al. 2010; Tsai et al. 2009). Although it was previously claimed (Montaldo et al. 2009; Tanter and Fink 2014) that the image quality of the technique used in the present study is poorer than that of conventional ultrasonography, the results indicate that certain laryngeal structures still can be identified in the ultrafast plane wave ultrasonic image (Fig. 1c). The neck skin is identified as an echogenic multilayer structure in the ultrasonic image. The outer and inner boundaries of the thyroid cartilage that lies below the neck skin are identified as two echogenic lines in the images. More importantly, the images obtained in this study (Fig. 1c) indicate that, for our study participants, the ultrasound beam was transmitted effectively through the thyroid cartilage, as the laryngeal structures below the cartilage are visible. The edges of the vocal fold and the false vocal fold can be identified as two short hyperechoic curves, which is consistent with the results obtained using conventional ultrasonography (Hu et al. 2010). The strong echo intensity at the edges and reverberation artifacts near the edges indicate that the ultrasound beam is fully reflected at the air–mucosa interface.

As illustrated in Figure 1c, apart from the hyperechoic vocal fold edge, the vocal fold body appears to be a hypo-echoic structure. It should be noted that the body layer and vocal fold body are two different terms. In this article, the vocal fold body refers to the whole body of the vocal fold apart from the edge. The body layer, which is a universally accepted term used in Hirano's cover-body model of the vocal fold, refers specifically to the structure that lies below the cover layer (Hirano 1974). In conventional ultrasonic images (Fig. 1b), the cover layer, although not very clear, can be distinguished from the body layer and appears to be a thin hypo-echoic layer above the hyper-echoic edge of the vocal fold. On the other hand, in ultrafast plane wave ultrasonic images, the body layer and cover layer are not distinguishable from each other. Despite this, it can be seen in Figure 1c that the supraglottic tissue including the false vocal fold and the subglottic tissue appear echoic in contrast to the hypo-echoic vocal fold body. This is consistent with the results obtained using conventional ultrasonography (Hu et al. 2010).

The signal-to-noise ratio (SNR) was introduced to evaluate the echogenicity of different parts of the larynx. The SNR was calculated from the RF data without applying time-gain-control or log scaling. It was calculated as the ratio of the signal magnitude in the region of interest (ROI) to the magnitude of the noise floor. The noise floor was estimated in the anechoic region, where there is no scatterer (region 4 in Fig. 1c). SNRs at the vocal fold body as well as sub- and supraglottic regions are illustrated in Figure 9. There was no significant

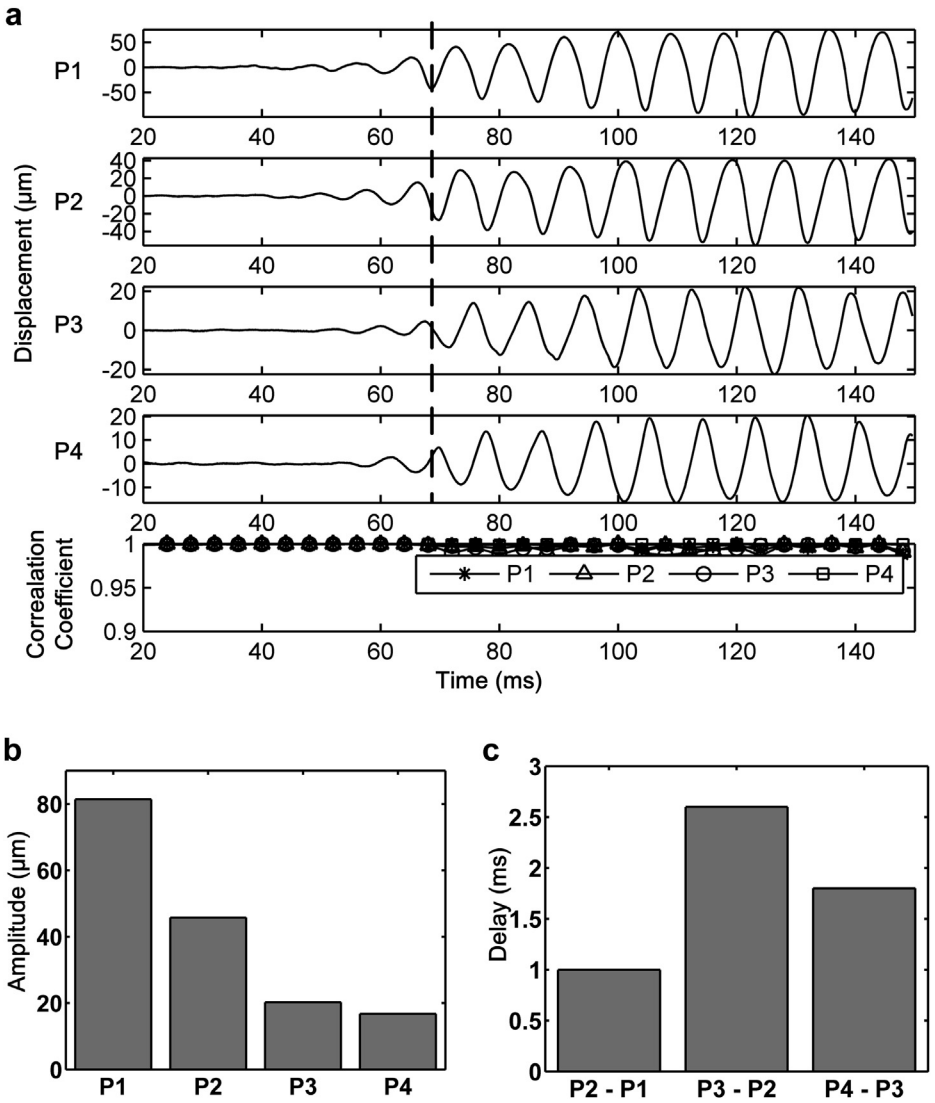

Fig. 3. Vibration of the supraglottic tissue. (a) Vibration waveforms of tissue at locations P1, P2, P3 and P4 and the corresponding correlation coefficients. (b) Amplitude of the vibration. (c) Phase delay between neighboring locations.

difference in SNR between emitting frequencies of 5 and 8 MHz (repeated-measures analysis of variance [ANOVA], $p = 0.236$). However, the SNR obtained at 8 MHz was significantly higher than that at 10 and 13.3 MHz, respectively (*post hoc* test, $p < 0.001$). Moreover, the SNR at the vocal fold body was significantly lower than that of the sub- and supraglottic regions, respectively (*post hoc test*, $p < 0.001$), which agrees with the earlier observation that the sub- and supraglottic tissues appear echoic in contrast to the hypo-echoic vocal fold body.

The ultrasonic images in Figure 1c and e were captured when phonation occurred, as indicated by the EGG waveform in Figure 1d. It can be observed that the spatial location, brightness and shape of the echoic speckles at the edge of the vocal fold change periodically during phonation. The opening and closing phases of the glottis could be identified in the EGG waveform. As indicated by the black dots (*i.e.*, moments of frame acquisition) in the EGG waveform, the high-frequency fast closing phase of the vibration, which is the steep rising edge in the EGG waveform, is well sampled by ultrafast plane wave ultrasonography working at 5000 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, even at low phonation frequency (*i.e.*, mean value at 125 Hz for men and 250 Hz for women) during normal vocal gestures (Deliyski 2010). From the end of the opening phase to the beginning of closing phase (350–354 ms), the edge of the vocal fold appears to be bright hyper-echoic speckle.

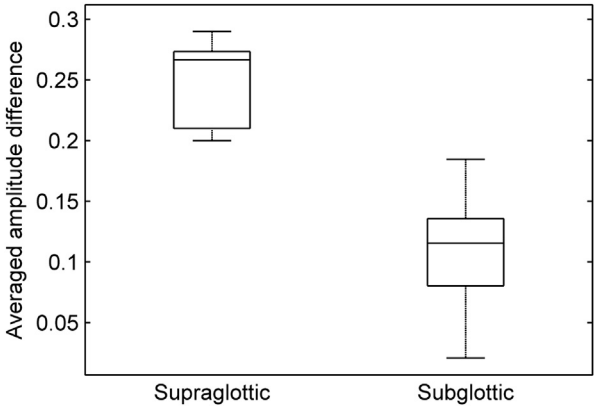

Fig. 4. Averaged amplitude difference between P1, P2, P3 and P4 in the supraglottic region and averaged amplitude difference between P5, P6, P7 and P8 in the subglottic region. Data are from 10 independent observations of the subjects.

This indicates that ultrasound is mostly reflected at the air–mucosal interface when the contact area between the pair of vocal folds drastically decreases in the opening phase. From the beginning of the closing phase to the end of the opening phase (355–358 ms), the speckle at the vocal fold edge appears dimmer than that before the closing phase, indicating that ultrasound is less reflected back to the transducer and penetrates through to the other side of the glottis when the paired vocal folds contact each other during the closing phase.

Vibration of the supraglottic tissue

The motion estimation algorithm was used to quantify the displacement of tissues at four locations, P1, P2, P3 and P4 (Fig. 2a). These four locations are near the mucosal surface of the supraglottic tissue. The distance between these locations is about 2.4 mm. Displacement of the tissue was obtained at a high level of RF signal correlation as the correlation coefficient is >0.95, which indicates that there is no severe decorrelation causing poor estimation of tissue vibration displacement. Actually, the transition from static state to stationary vibration during the onset of phonation can be visualized in the displacement waveforms (Fig. 3a), which indicates that the vibration displacement waveform of the supraglottic tissue was successfully quantified. The vibration amplitude is gradually growing to its maximum in the transition.

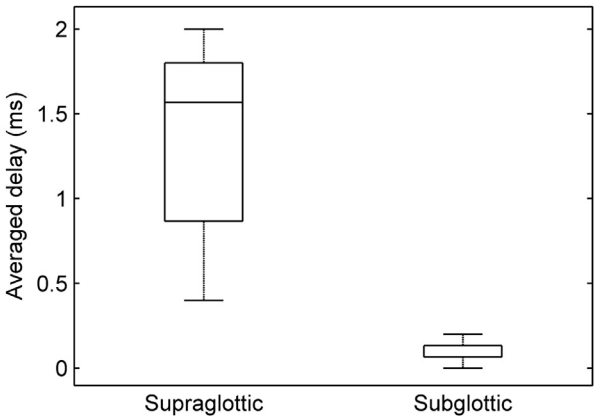

Fig. 5. Averaged delay between P1, P2, P3 and P4 in the supraglottic region and averaged delay between P5, P6, P7 and P8 in the subglottic region. Data are from 10 independent observations of the subjects.

The amplitude of the quasi-periodic vibration of the supraglottic tissue is illustrated in Figure 2a, where it can be seen that the vibration amplitude of the tissue in the small area at the bottom right is much larger than that at the remaining locations in the region. In Figure 3b, the vibration amplitude gradually decreases from location P1 to location P4. The same pattern is observed for all subjects. To quantify this variation in vibration amplitude, the averaged amplitude difference between P1, P2, P3 and P4 in the subglottic region, for all subjects, is calculated according to the equation (Fig. 4)

$$\begin{aligned}\text{Amp\_diff\_averaged} = \\ [|\text{Amp}_1-\text{Amp}_2|+|\text{Amp}_2-\text{Amp}_3|+ \\ |\text{Amp}_3-\text{Amp}_4|]/\text{Amp}_{\text{max}}/3\end{aligned} \quad (1)$$

where Amp_1, Amp_2, Amp_3 and Amp_4 denote the vibration amplitudes of tissues at locations P1, P2, P3 and P4 respectively. The vertical lines denote the absolute value. Amp_{max} denotes the maximum vibration amplitude of supraglottic tissue among locations P1 to P4. Amp_diff_averaged is the averaged amplitude difference between P1, P2, P3 and P4 in the subglottic region. In the comparison of vibration patterns between sub- and supraglottic tissues, the relative difference rather than the absolute difference is investigated. Thus, the averaged amplitude difference illustrated in Figure 4 was obtained by normalizing the absolute difference between locations by the maximum vibration amplitude. The result indicates that the averaged difference in vibration amplitudes between neighboring locations in the supraglottic region is >20% of the maximum vibration amplitude for all subjects (Fig. 4).

As indicated by the *vertical dashed line* in Figure 3a, there are phase differences between the vibrations of the tissues at different locations. The quantitative results indicate that the vibration of the tissue at P1 is earlier than that of the tissues at P2, P3 and P4 (Fig. 3c). A pattern is observed in all subjects: the vibration phase of the tissue closer to the right boundary of the supraglottic region occurs earlier than that of the tissue farther from the right boundary. The averaged phase delay between neighboring locations (*i.e.*, P1, P2, P3

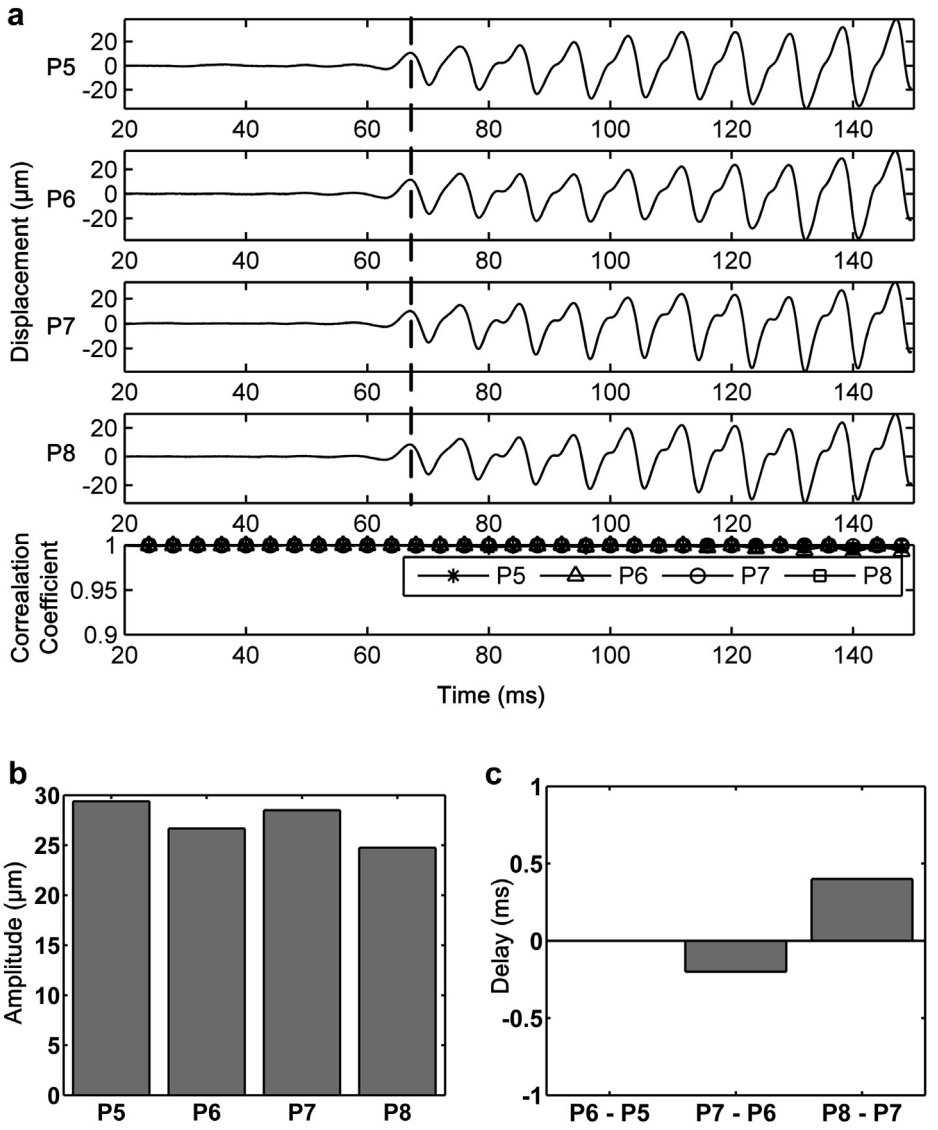

Fig. 6. Vibration of the subglottic tissue. (a) Vibration waveforms of tissue at locations P5, P6, P7 and P8 and the corresponding correlation coefficients. (b) Amplitude of the vibration. (c) Phase delay between neighboring locations.

and P4) in the supraglottic region is calculated for all subjects according to the equation (Fig. 5)

$$\text{Averaged\_delay} = [|\text{Delay}_{P2-P1}| + |\text{Delay}_{P3-P2}| + |\text{Delay}_{P4-P3}|]/3 \quad (2)$$

where Delay_{P2-P1}, Delay_{P3-P2} and Delay_{P4-P3} denote the vibration phase delays between the neighboring locations. The averaged phase delay in the supraglottic region could reach 2 ms (Fig. 5).

Vibration of the subglottic tissue

The motion estimation algorithm was used to quantify the displacement of tissue at four locations, P5, P6, P7 and P8 (Fig. 2b). These four locations are near the mucosal surface of the subglottic tissue. The distance between these locations is about 2.4 mm. The displacement of the tissue was obtained at a high level of RF signal correlation, as the correlation coefficient is still high above 0.95. There is no severe decorrelation causing poor estimation of tissue vibration displacement, similar to the quantification of supraglottic vibration displacement. The transition from static to stationary vibration is observed in the vibration waveforms in Figure 6a.

The amplitude of vibration of subglottic tissue is illustrated in Figure 2b. It can be seen that the vibration amplitude of tissue nearer to the curved mucosal surface is larger than that of tissue farther away from the surface. Moreover, variation in vibration amplitude

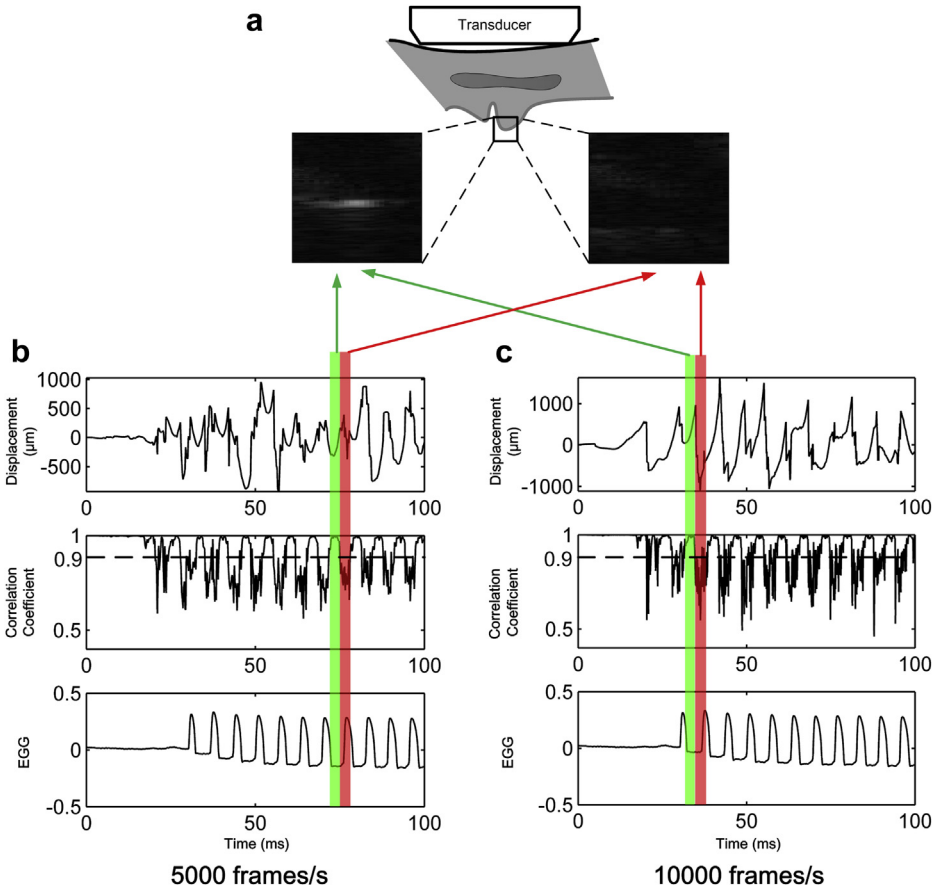

Fig. 7. Vibration of the vocal fold edge. (a) Two frames of images of the vocal fold edge acquired at two different periods during phonation. (b) Displacement waveform obtained at 5000 frames/s. (c) Displacement waveform obtained at 10,000 frames/s. The *green bands* in (b) and (c) indicate the period during which the hyper-echoic vocal fold edge was continuously tracked at high levels of signal correlation. The *red bands* in (b) and (c) indicate the period during which the vocal fold edge turns hypo-echoic and is unable to be tracked because of severe signal decorrelation.

along the subglottic mucosal surface (Figs. 2b and 6b) appears to be smaller than that along the supraglottic mucosal surface (Figs. 2a and 3b). In addition, the vibration phase delay between neighboring locations at the subglottic region (Fig. 6c) is much smaller than that at the supraglottic region (Fig. 3c). Furthermore, a quantitative statistical comparison of vibration patterns (*i.e.*, amplitude variation and phase delay) between the sub- and supraglottic regions was carried out. The averaged amplitude difference between neighboring locations (P5, P6, P7 and P8) in the subglottic region was calculated for all subjects by replacing the supraglottic vibration amplitudes (*i.e.*, Amp_1, Amp_2, Amp_3, Amp_4 and Amp_{max}) in eqn (1) with the subglottic vibration amplitudes (*i.e.*, Amp_5, Amp_6, Amp_7, Amp_8 and Amp_{max}). The averaged phase delay between neighboring locations (*i.e.*, P5, P6, P7 and P8) in the subglottic region was calculated for all subjects by replacing the supraglottic delays (*i.e.*, $Delay_{P2-P1}$, $Delay_{P3-P2}$ and $Delay_{P4-P3}$) in eqn (2) with the subglottic delays (*i.e.*, $Delay_{P6-P5}$, $Delay_{P7-P6}$ and $Delay_{P8-P7}$). As the data (*i.e.*, either averaged amplitude difference or averaged phase delay) were obtained in two different regions (*i.e.*, sub- and supraglottic regions) in each subject, a paired-sample *t*-test was employed to compare the means of the data obtained in the two different regions. The data obtained from all subjects (Fig. 4) indicate that the averaged amplitude difference between the locations in the subglottic region is significantly smaller than that in the supraglottic region (paired sample *t*-test, $p < 0.001$). This result indicates that the variation in vibration amplitude along the subglottic mucosal surface is significantly smaller than that along the supraglottic mucosal surface. Moreover, data obtained from all subjects (Fig. 5) indicate that the averaged phase delay between the locations in the subglottic region is significantly smaller than the averaged phase delay between the locations in the supraglottic region (paired sample *t*-test, $p < 0.001$).

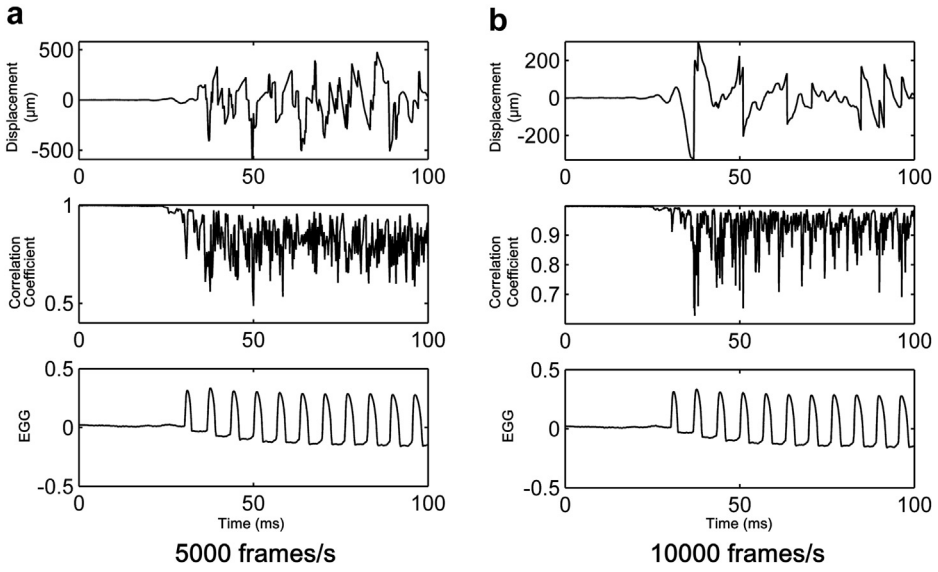

Fig. 8. Vibration of the vocal fold body. (a) Displacement waveform obtained at 5000 frames/s. (b) Displacement waveform obtained at 10,000 frames/s. The cross-correlation coefficient is provided in the middle row in (a) and (b). The electroglottography (EGG) waveform is shown in the bottom row in (a) and (b).

Vibration of the vocal folds

In our pilot experiment, the results of measurement of vocal fold vibration obtained using the set of small kernels were quite noisy because of the severe decorrelation issue. Considering that the inter-frame displacement of the vibrating vocal folds could reach 0.2 mm (5000 frames/s) because of the high velocity of the vibration (*i.e.*, above 1 m/s) (Doellinger and Berry 2006), the size of the kernel window could be potentially too small to see the large inter-frame displacements. Therefore, during estimation of the vibration displacement of vocal fold tissue, the kernel windows used in the motion estimation algorithm were expanded to prevent any large out-of-window displacement causing erroneous estimation and severe decorrelation. Although the windows were expanded in both the horizontal and vertical directions, only the displacement in the horizontal direction is shown in the results because of the limited lateral resolution. The ultrasonic data illustrated in Figures 7 and 8 were acquired at 10,000 frames/s using ultrafast plane wave ultrasonography. By decimating the data by a factor of 2, we obtained a set of data at a frame rate of 5000 frames/s. Therefore, both the original data (10,000 frames/s) and the decimated data (5000 frames/s) were obtained in a single acquisition from the subjects.

First, the motion estimation algorithm was used to quantify the displacement of the edge of the vocal fold (the location indicated by a *white cross mark* in Fig. 1c). The displacement waveform is illustrated in Figure 7b and c. A state of periodic vibration was established after phonation occurred, as indicated by the EGG waveform in Figure 7b and c. However, the displacement obtained by using the motion estimation algorithm appears chaotic and noisy. In addition, there was no significant improvement of the estimation of the vibration displacement when the imaging frame rate was increased

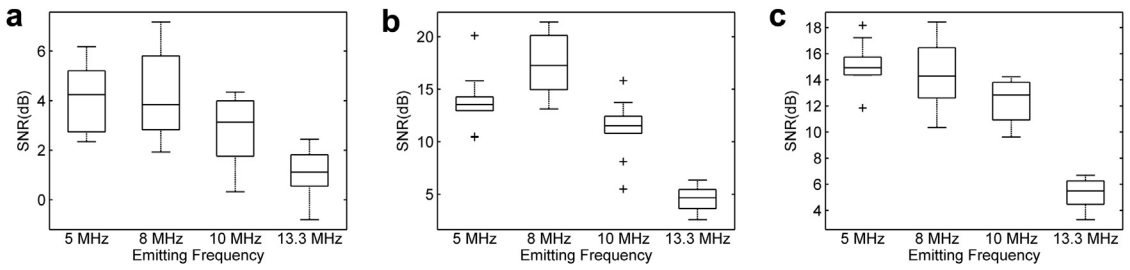

Fig. 9. Box plot of the range of echo signal-to-noise ratios at the (a) vocal fold body, (b) supraglottic region and (c) subglottic region. These regions are pointed out in Figure 1c. On each box, the central mark is the median. The edges of the box are the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually.

to 10,000 frames/s. As illustrated in the middle rows in Figure 7b and c, the decorrelation is so severe that the correlation coefficient drops to 0.5, which indicates that signal decorrelation causes the failure to continuously track the periodic oscillation of the vocal fold edge. However, there is a period in each vibration cycle when the correlation coefficient remains above 0.9 (indicated by the *dashed lines* in the middle rows in Fig. 7b and c), which indicates that there is no severe decorrelation causing erroneous estimation of displacement in this period of time. On further investigation, it can be seen that the vocal fold edge appears to be a hyper-echoic curve during the period when the correlation coefficient remains above 0.9 (indicated by the *green bands* in Fig. 7b and c). On the other hand, the echo intensity of the vocal fold edge drastically decreased during the period when the correlation coefficient dropped below 0.9 (indicated by the *red bands* in Fig. 7b and c). The vocal fold edge turned from a hyper-echoic structure into a hypo-echoic structure (Fig. 7a).

Furthermore, the motion estimation algorithm was also used to quantify the displacement of tissue at the vocal fold body (the location indicated by a *round dot* in Figure 1c, 3 mm above edge of the vocal fold). The displacement waveform is illustrated in Figure 8. As shown in the displacement waveform, the vocal fold tissue remained static before the onset of phonation. After phonation occurred, the motion estimation method failed to track the movement of the vocal fold body, which vibrated periodically, as indicated by the EGG waveform. As illustrated in the middle rows in Figure 8a and b, the signal decorrelation was so severe that the correlation coefficient dropped below 0.7, which indicates that signal decorrelation caused the failure to track the vibration of the vocal fold body. In addition, there was no significant improvement in estimation of the vibration displacement when the imaging frame rate was increased to 10,000 frames/s.

DISCUSSION

In the present study, the vibration of the hyper-echoic vocal fold edge was visualized using ultrafast plane wave ultrasonography (Fig. 1e). On the other hand, in the ultrafast plane wave ultrasonic images, the body of the vocal fold appeared to be a hypo-echoic structure without any layered structure inside. Therefore, the vibration of the tissue in the vocal fold body was not visible. This does not agree with the results obtained using conventional ultrasonography. According to previous investigations using conventional ultrasonography (Qin et al. 2011; Tang et al. 2013; Tsai et al. 2009), the vocal fold can be divided into two distinguishable layers of different echo intensities in the image. Tsai et al. (2009) and Qin et al. (2011) proved that the echoic layer and hypo-echoic layer represent the body layer and cover layer, respectively. Moreover, the vibration of the body layer of the vocal fold can be visualized and quantified using conventional ultrasonography, as described in our previous studies (Qin et al. 2011; Tang et al. 2013). In the present study, the result that the layered structures of the vocal fold are not identifiable indicates the degradation of image contrast when the imaging mode was switched from conventional ultrasonography to ultrafast plane wave ultrasonography. This is not unexpected, as the previous investigation (Montaldo et al. 2009) revealed that the image contrast and SNR were lower in ultrafast plane wave ultrasonography than in conventional ultrasonography because of the lack of transmit focusing. We also investigated the echo intensity of the vocal fold tissue under different emitting frequencies. The results indicate that there is a significant decrease in echo intensity of the vocal fold tissue when the ultrasound emitting frequency is increased. This is probably attributable to the frequency-dependent attenuation effect of the tissue and the limited bandwidth of the ultrasonic transducer.

The severe signal decorrelation caused the failure to track the periodic oscillation of the vocal fold body in the present study. The decorrelation is usually caused by out-of-plane displacement, large inter-frame displacements induced by high motion rate, poor SNR and other factors. Although an increasing frame rate has been reported to be useful in reducing decorrelation caused by a high motion rate (Park et al. 2007; Wu et al. 2010), no significant improvement in the estimation of vibration displacement was obtained in this study when the imaging frame rate was increased to 10,000 frames/s. Another potential cause of the poor estimation of tissue displacement is the low SNR of the echo in the vocal fold body. According to a previous study, the performance of the normalized cross-correlation algorithm could drastically decrease at low SNR (Walker and Trahey 1995).

In our previous studies (Qin et al. 2011; Tang et al. 2013), we were unable to track the movement of the vocal fold edge. Although, in the present study, continuous tracking of the movement of the vocal fold edge during the whole vibration cycle was also unachievable, there was a period in each cycle during which the hyper-echoic vocal fold edge could be successfully and continuously tracked at a high level of signal correlation (Fig. 7). The glottis was open during this period, and there was little contact between the paired vocal folds, as the high echo intensity of the vocal fold edge indicated the strong reflection of ultrasound at the air–mucosa interface. However, when the echo intensity of the vocal fold edge drastically decreased, the

performance of motion estimation degraded causing the failure to continuously track the movement of the vocal fold edge. The decrease in echo intensity is most probably attributable to the contact of the paired vocal folds as illustrated in Figure 1. However, another potential cause of the decrease in echo intensity that should not be precluded is the anisotropic reflection of the ultrasound at the air–mucosa interface, of which the normal direction changes drastically during vibration.

One of the supraglottic structures to which researchers have been paying much attention is the false vocal folds. Although the echo intensity of the false vocal fold lying in the supraglottic region is significantly higher than that of the vocal fold body (Figs. 1c and 9), the vibration of the false vocal fold is not as obvious as the vibration of the vocal fold edge in the ultrasonic image sequence. By employing the motion estimation method, the vibration of the false vocal folds is successfully visualized in the displacement waveforms. The false vocal folds, which have been proved to act as an acoustical filter and an aerodynamic flow regulator (Alipour et al. 2007; Pepinsky 1942) during voice production, have usually been modeled as a rigid non-oscillatory structure in previous studies (Li et al. 2008; McGowan and Howe 2010; Zhang et al. 2002). However, the results of the present study reveal that the amplitude of the false vocal fold vibration could reach several tens of micrometers during normal phonation tasks (Fig. 3b). These results suggest that the false vocal folds actually act as a pliable structure during phonation, which is consistent with the conclusions of previous studies (Alipour and Finnegan 2013; Bailly et al. 2014). In addition, according to previous reports, the vibration amplitude of the false vocal folds could reach several millimeters during pathologic phonation (Lindestad et al. 2004; Nasri et al. 1996; Von Doersten et al. 1992). The results of the present study indicate that the false vocal fold can be identified as an echogenic structure in ultrafast plane wave images (Fig. 1c). Therefore, ultrafast plane wave ultrasonography, which is a non-invasive and well-tolerated medical imaging technique when compared with high-speed optical laryngeal endoscopy, can be employed as an alternative to detect and visualize the false vocal fold oscillation that accompanies certain types of vocal disorders. The phase delay of the tissue vibration illustrated in Figure 5 indicates that the vibration spreads along the mucosal surface of the false vocal folds and the supraglottic region in the form of a wave that travels at approximately 1.2 to 6 m/s. Observation of the false vocal fold mucosal traveling wave was first reported in a previous investigation in which the false vocal fold vibration was recorded in the horizontal plane from above the glottis using a stroboscopic endoscope (Nasri et al. 1996). Moreover, previous studies (Bailly et al. 2014; Sakakibara et al. 2001) in which the false vocal fold is modeled as two vertically spring-connected masses that vibrate at different phases reported that the traveling wave could be an indispensable component of the false vocal fold vibration. However, there is no previous report of measurement of the speed of this wave. To the best of our knowledge, this is the first time that the speed of the traveling wave in the false vocal folds has been quantified.

The result that the delay of the vibration between the tissues at neighboring locations near the subglottic mucosal surface is around zero (Fig. 6c) indicates that the subglottic tissues vibrate nearly synchronously during phonation. The results illustrated in Figures 4 and 5 also indicate that the vibration pattern of the subglottic tissue is significantly different from the vibration pattern of the supraglottic tissue. The statistically significant difference in vibration pattern between the sub- and supralgottic tissues indicates potential differences in tissue mechanical property. Moreover, the tissue mechanical property could vary over time, possibly as a result of pathologic changes such as stenosis (Wan et al. 2014) and carcinoma (Ophir et al. 2000). Therefore, the variation in the vibration pattern may be interpreted as a correlate of pathologic changes such as carcinoma (Sessions et al. 1975) in the subglottic region and upper trachea. This technique has potential value in the detection of pathologic changes in subglottic region and upper trachea.

CONCLUSIONS

In the present study, the feasibility and limitations of applying ultrafast plane wave ultrasonography for visualization and quantification of the phase and amplitude of the vibration of the laryngeal tissue were elucidated. This study reveals that laryngeal structures including the vocal folds, the thyroid cartilage and the sub- and supraglottic structures can be identified in ultrasonic images acquired using ultrafast plane wave ultrasonography. The difference in echogenicity between the vocal fold body and the sub- and supraglottic structures was also observed in the present study. The SNR of the ultrasound echo in the sub- and supraglottic regions was significantly higher than that in the vocal fold body. Moreover, the vibration of the vocal fold edge can be visualized using ultrafast plane wave ultrasonography.

Moreover, we found that displacement of the vibrating sub- and supraglottic structures could be achieved at a high level of RF signal correlation using a motion estimation algorithm. The transition from the static state to the stationary periodic vibration of the sub- and supraglottic tissues can be visualized in the displacement waveforms obtained. A statistically

significant difference in vibration pattern between the sub- and supraglottic structures was revealed in this study. Variation in vibration amplitude along the subglottic mucosal surface was significantly smaller than that along the supraglottic mucosal surface. The phase delay of vibration along the subglottic mucosal surface was significantly smaller than that along the supraglottic mucosal surface. Our next step will focus on the variation in vibration pattern of sub- and supraglottic tissues, especially the pliable false vocal folds, during different types of phonation tasks and vocal gestures. As the poor SNR of the hypo-echoic vocal fold body results in the limitation of this technique in continuously tracking movement of the vocal fold tissue, future work will also include improvement of the echo intensity of vocal fold tissue during imaging of laryngeal vibration.

Acknowledgments—This work was supported by the National Natural Science Foundations of China under Grant Nos. 61271087, 11274250, 11404256, and by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2016JQ2017.

REFERENCES

Alipour F, Berry DA, Titze IR. A finite-element model of vocal-fold vibration. J Acoust Soc Am 2000;108:3003–3012.

Alipour F, Jaiswal S, Finnegan E. Aerodynamic and acoustic effects of false vocal folds and epiglottis in excised larynx models. Ann Otol Rhinol Laryngol 2007;116:135–144.

Alipour F, Finnegan E. On the acoustic effects of the supraglottic structures in excised larynges. J Acoust Soc Am 2013;133:2984–2992.

Bailly L, Bernardoni NH, Muller F, Rohlfs AK, Hess M. Ventricular-fold dynamics in human phonation. J Speech Lang Hear Res 2014;57:1219–1242.

Booth JR, Childers DG. Automated-analysis of ultra high-speed laryngeal films. IEEE Trans Biomed Eng 1979;26:185–192.

Bozzato A, Zenk J, Gottwald F, Koch M, Iro H. Influence of thyroid cartilage ossification in laryngeal ultrasound. Laryngorhinootologie 2007;86:276–281.

Childers DG, Hicks DM, Moore GP, Alsaka YA. A model for vocal fold vibratory motion, contact area, and the electroglottogram. J Acoust Soc Am 1986;80:1309–1320.

Couade M, Pernot M, Messas E, Emmerich J, Hagege A, Fink M, Tanter M. Ultrafast imaging of the arterial pulse wave. IRBM 2011;32:106–108.

Deliyski DD. Laryngeal high-speed videoendoscopy. In: Kendall KA, Leonard RJ, (eds). Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging. New York: Thieme Medical; 2010. p. 243–270.

Deliyski DD, Powell MEG, Zacharias SRC, Gerlach TT, de Alarcon A. Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment. Biomed Signal Process Control 2015;17:21–28.

Denarie B, Tangen TA, Ekroll IK, Rolim N, Torp H, Bjastad T, Lovstakken L. Coherent plane wave compounding for very high frame rate ultrasonography of rapidly moving targets. IEEE Trans Med Imaging 2013;32:1265–1276.

Doellinger M, Berry DA. Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation. J Voice 2006;20:401–413.

Doellinger M, Kobler J, Berry DA, Mehta DD, Luegmair G, Bohr C. Experiments on analysing voice production: Excised (human, animal) and in vivo (animal) approaches. Curr Bioinform 2011;6:286–304.

Hirano M. Morphological structure of the vocal cord as a vibrator and its variations. Folia Phoniatr Logop 1974;26:89–94.

Hu QA, Zhu SY, Luo F, Gao Y, Yang XY. High-frequency sonographic measurements of true and false vocal cords. J Ultrasound Med 2010;29:1023–1030.

Jing BW, Tang SS, Wu L, Wang SP, Wan MX. High speed imaging and measurement of laryngeal vibration during phonation using ultrafast ultrasonography: A preliminary study. Proc IEEE Ultrason Symp 2015; http://dx.doi.org/10.1109/ULTSYM.2015.0259.

Kobler JB, Chang EW, Zeitels SM, Yun SH. Dynamic imaging of vocal fold oscillation with four-dimensional optical coherence tomography. Laryngoscope 2010;120:1354–1362.

Krausert CR, Olszewski AE, Taylor LN, McMurray JS, Dailey SH, Jiang JJ. Mucosal wave measurement and visualization techniques. J Voice 2011;25:395–405.

Li S, Wan MX, Wang SP. The effects of the false vocal fold gaps on intralaryngeal pressure distributions and their effects on phonation. Sci China C Life Sci 2008;51:1045–1051.

Lindestad PA, Blixt V, Pahlberg-Olsson J, Hammarberg B. Ventricular fold vibration in voice production: A high-speed imaging study with kymographic, acoustic and perceptual analyses of a voice patient and a vocally healthy subject. Logoped Phoniatr Vocol 2004;29:162–170.

McGowan RS, Howe MS. Influence of the ventricular folds on a voice source with specified vocal fold motion. J Acoust Soc Am 2010;127:1519–1527.

Mergell P, Herzel H, Titze IR. Irregular vocal-fold vibration: High-speed observation and modeling. J Acoust Soc Am 2000;108:2996–3002.

Montaldo G, Tanter M, Bercoff J, Benech N, Fink M. Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography. IEEE Trans Ultrason Ferroelectr Freq Control 2009;56:489–506.

Nasri S, Jasleen J, Gerratt BR, Sercarz JA, Wenokur R, Berke GS. Ventricular dysphonia: A case of false vocal fold mucosal traveling wave. Am J Otolaryngol 1996;17:427–431.

Ophir J, Garra B, Kallel F, Konofagou E, Krouskop T, Righetti R, Varghese T. Elastographic imaging. Ultrasound Med Biol 2000;26:S23–S29.

Park S, Aglyamov SR, Scott WG, Emelianov SY. Strain imaging using conventional and ultrafast ultrasound imaging: Numerical analysis. IEEE Trans Ultrason Ferroelectr Freq Control 2007;54:987–995.

Pepinsky A. The laryngeal ventricle considered as an acoustical filter. J Acoust Soc Am 1942;14:32–35.

Qin XL, Wu L, Jiang HJ, Tang SS, Wang SP, Wan MX. Measuring body-cover vibration of vocal folds based on high-frame-rate ultrasonic imaging and high-speed video. IEEE Trans Biomed Eng 2011;58:2384–2390.

Ramamurthy BS, Trahey GE. Potential and limitations of angle-independent flow detection algorithms using radio-frequency and detected echo signals. Ultrason Imaging 1991;13:252–268.

Sakakibara KI, Konishi T, Kondo K, Murano EZ, Kumada M, Imagawa H, Niimi S. Vocal fold and false vocal fold vibrations and synthesis of khoomei. In: Proceedings, International Computer Music Conference, Havana, Cuba, 2001:135–138.

Sandrin L, Catheline S, Tanter M, Hennequin X, Fink M. Time-resolved pulsed elastography with ultrafast ultrasonic imaging. Ultrason Imaging 1999;21:259–272.

Sandrin L, Tanter M, Catheline S, Fink M. Shear modulus imaging with 2-D transient elastography. IEEE Trans Ultrason Ferroelectr Freq Control 2002;49:426–435.

Schutte HK, Svec JG, Sram F. First results of clinical application of videokymography. Laryngoscope 1998;108:1206–1210.

Sessions DG, Ogura JH, Fried MP. Carcinoma of the subglottic area. Laryngoscope 1975;85:1417–1423.

Shaw HS, Deliyski DD. Mucosal wave: A normophonic study across visualization techniques. J Voice 2008;22:23–33.

Story BH, Titze IR. Voice simulation with a body-cover model of the vocal folds. J Acoust Soc Am 1995;97:1249–1260.

Svec JG, Schutte HK. Videokymography: High-speed line scanning of vocal fold vibration. J Voice 1996;10:201–205.

Tang SS, Zhang YY, Qin XL, Wang SP, Wan MX. Measuring body layer vibration of vocal folds by high-frame-rate ultrasound synchronized with a modified electroglottograph. J Acoust Soc Am 2013;134:528–538.

Tanter M, Fink M. Ultrafast imaging in biomedical ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 2014;61:102–119.

Tao C, Zhang Y, Jiang JJ. Extracting physiologically relevant parameters of vocal folds from high-speed video image series. IEEE Trans Biomed Eng 2007;54:794–801.

Titze IR, Jiang JJ, Hsiao TY. Measurement of mucosal wave-propagation and vertical phase difference in vocal fold vibration. Ann Otol Rhinol Laryngol 1993;102:58–63.

Tsai CG, Chen JH, Shau YW, Hsiao TY. Dynamic B-mode ultrasound imaging of vocal fold vibration during phonation. Ultrasound Med Biol 2009;35:1812–1818.

Von Doersten PG, Izdebski K, Ross JC, Cruz RM. Ventricular dysphonia: A profile of 40 cases. Laryngoscope 1992;102:1296–1301.

Walker WF, Trahey GE. A fundamental limit on delay estimation using partially correlated speckle signals. IEEE Trans Ultrason Ferroelectr Freq Control 1995;42:301–308.

Wan JJ, He FL, Zhao YF, Zhang HM, Zhou XD, Wan MX. Non-invasive vascular radial/circumferential strain imaging and wall shear rate estimation using video images of diagnostic ultrasound. Ultrasound Med Biol 2014;40:622–636.

Wu SY, Wang SL, Li PC. Tracking in high-frame-rate imaging. Ultrasonic Imaging 2010;32:1–15.

Yu LF, Liu GJ, Rubinstein M, Saidi A, Wong BJF, Chen ZP. Office-based dynamic imaging of vocal cords in awake patients with swept-source optical coherence tomography. J Biomed Opt 2009;14:064020.

Zhang C, Zhao W, Frankel SH, Mongeau L. Computational aeroacoustics of phonation: Part II, Effects of flow parameters and ventricular folds. J Acoust Soc Am 2002;112:2147–2154.

Zhang ZY. Characteristics of phonation onset in a two-layer vocal fold model. J Acoust Soc Am 2009;125:1091–1102.