Measuring Body-Cover Vibration of Vocal Folds Based on High-Frame-Rate Ultrasonic Imaging and High-Speed Video

Xulei Qin, Liang Wu, Hujie Jiang, Shanshan Tang, Supin Wang*, and Mingxi Wan*

***Abstract*—Vibration of vocal folds is a body-cover layered vibration pattern due to the two-layer tissue structures of vocal folds. A method based on a synchronal imaging system is proposed in order to image and measure the body-cover vibration pattern of vocal folds. This imaging system contains two parts: high-frame-rate ultrasonic imaging part and high-speed video part, which can synchronously image the vibration of the body and cover layers at high speed. Then, image analysis methods are applied to measure the body-cover vibration of vocal folds from both recorded image sequences. We analyze characteristics of body-layer vibration based on the measurements from designed experiments. Moreover, these results meet simulations of a body-cover model.**

***Index Terms*—Body-cover layers of vocal folds, high-frame-rate ultrasonic (HFRU), high-speed video (HSV), vibration measurement.**

I. Introduction

THE vibration of vocal folds directly decides the properties of voice during phonations, such as pitch, loudness, and voice quality. It is always an important issue for voice production research and clinical laryngeal examination. However, this vibration is a fast, fine, and complicated process, and layered as a body-cover vibration pattern, which is due to the two-layer tissue structures of vocal folds [1]. Therefore, determining how to dynamically image and precisely measure the complex layered vibration is a basic and tough problem for the research of vocal fold vibration.

General instruments with low imaging speed, such as MRI and current transformer, cannot acquire enough details for the fast vibration of vocal folds. Therefore, high-speed video (HSV) was introduced into this area to image and measure the vibration of vocal folds fast enough [2]–[5]. It could record the images of vocal fold vibration above 1000 frames/s, which was several times more than the frequency of vocal fold vibration. Moreover, physiological parameters of vocal folds could be inversed according to biomechanical models [14]–[19] based on the experimental parameters extracted from HSV images [6]–[13]. However, HSV can only image upper surface of the cover layer, and not directly provide any information of the body layer, which was enveloped by the cover layer.

On the other hand, ultrasonic imaging was applied in larynx examination in some studies [20]–[22] because ultrasound could propagate in laryngeal tissues. Similar to stroboscope, dynamic ultrasonic imaging could record the vibration information of both vocal fold layers during stationary phonations [23], [24], with which Tsai *et al*. analyzed the vibration of the body layer [25]. However, this method could not supply sufficient dynamic details for each vibration cycle, because their image speed was still much lower than that of vocal fold vibration, which led to disability to investigate irregular phonation or changing instants of phonations, such as onset and offset. Moreover, although the vibration of the cover layer could be imaged by ultrasonic imaging, it was ambiguous in these images due to its larger movements and severe artifact reverberated by the air–tissue interface. Therefore, it is not suitable to investigate the body-cover layered vibration pattern of vocal folds based on ultrasonic images only.

A synchronal imaging system is proposed in this paper to measure the body-cover vibration of vocal folds. This system contains two imaging parts: one is high-frame-rate ultrasonic (HFRU) imaging and the other is HSV. HFRU imaging is introduced to record the vibration of the body layer at high speed, which can even be up to 1000 frames/s when imaging region is smaller. Meanwhile, HSV is applied to record the vibration of the cover layer also at high speed. Then, image analysis methods are applied to measure and analyze the body-cover vibration of vocal folds from both recorded image sequences. At last, we discuss the vibration characteristics of the body layer and the relationship between the vibrations of the body and cover layers based on these results.

II. Methods

A. Vibration Imaging of Vocal Fold Layers

We set a synchronous imaging system including HFRU and HSV parts in order to image the vibration of the body and cover layers during phonations. The HFRU part was an ultrasonic imaging machine (Sonix RP, Ultrasonix Medical Corporation, Richmond, Canada) with a linear-array transducer (L14–5/38,

Manuscript received December 20, 2010; revised March 13, 2011 and April 23, 2011; accepted April 29, 2011. Date of publication May 19, 2011; date of current version July 20, 2011. This work was supported by the National Natural Science Foundations of China under Grant 30770544 and Grant 10874137, and by the National Basic Research Program 973 under Grant 2010CB732603 from the Ministry of Science and Technology of China. *Asterisks indicate corresponding author.*

X. Qin, L. Wu, H. Jiang, and S. Tang are with the Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China.

*S. Wang and *M. Wan are with the Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: spwang@mail.xjtu.edu.cn; mxwan@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/TBME.2011.2157156

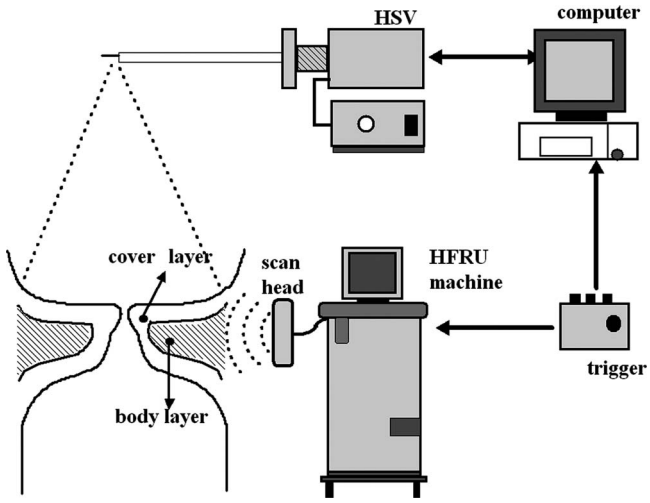


Fig. 1. Illustration of the experiment using a synchronal imaging system.

38 mm, Sonix Series). It could record high-resolution B-model ultrasonic images at high speed, which was up to 1000 frames/s or even more when imaging region was smaller. The HSV part, containing a high-speed digital camera (model 9700, Kay Electronic) and an endoscope (model 9106), could record optical images at the speed of 2000 frames/s. Both parts could be started to record data synchronously by an external trigger.

First, the scan head of HFRU part was placed on subjects' neck surface of right side along the coronal plane [25], and the endoscope of HSV part was inserted into the open mouth. These different placements ensured both parts imaging synchronously during phonation. The parameters of HFRU part were adjusted to image right vocal fold at high speed. Then, the subjects were introduced to pronounce the vowel /u:/ steadily in chest register and loudly to produce larger vibration. Meanwhile, the trigger was pressed to start HFRU and HSV parts to record images simultaneously. Whole recording time once lasted about 8 s, which included about several hundred vibration cycles of vocal folds. The illustration of this experiment is shown in Fig. 1.

B. Vibration Measurements of Body Layer From HFRU Images

An optical flow field algorithm is introduced to measure the vibration of the body layer from the recorded HFRU image series. It can calculate the motion velocities of brightness patterns in an image, which is called optical flow, according to the time-varying brightness of pixels in an image series [26].

Here, $I(X,t)$ is supposed to be the brightness of point X in the image at moment t, when the time distance and motion of objects between two neighbored frames are both smaller. If V is the optical flow at this point, it will move to the place of $X = X + dX$ at moment $t + dt$; here, $dX = V \cdot dt$. Because of the smaller motion, the brightness at this point $I(X + dX,\ t + dt)$ should need the following equation:

$$I(X,t) = I(X + dX,\ t + dt). \tag{1}$$

It is written in the second-order formal Taylor expansion as

$$\nabla I(X,t) \cdot V + \frac{dI(X,t)}{dt} = 0. \tag{2}$$

This is a gradient restriction equation and was calculated by Lucas and Kanade [27] through a weighted least-squares method. They calculated the optical flow field algorithm by minimizing the following equation in each smaller domain Ω, which could yield motion vector field at subpixel level [28]:

$$\sum_{X \in \Omega} [W(X)]^2 . \left[\nabla I(X,t) \cdot V + \frac{dI(X,t)}{dt}\right]^2 \tag{3}$$

where $W(X)$ is a window function.

The time distance between two neighbored HFRU images is very small, which is about 1 ms. The motion displacements of vocal folds in these image series, especially the motion of the body layer, is smaller than several pixels. Therefore, the motion displacements of the body layer can be measured by an optical flow field algorithm. Unfortunately, although the motion of the cover layer is more intense, it is hardly measured from ultrasonic images accurately because the air–tissue surface of the cover layer scatters ultrasound to cause severe reverberation artifact.

C. Vibration Measurements of Cover Layer From HSV Images

1) Image Processing Approach for HSV Images: An image processing approach is applied to analyze the recorded HSV image series [13] for the purpose of measuring the vibration of the cover layer. This approach contains two steps. The first one is to segment glottal regions at pixel level from HSV image series by an improved level set algorithm, which is based on the dynamics properties of this series. The second one is a subpixel edge detection algorithm based on *Zernike* operator, which can extract more accurate vibration parameters because the HSV image resolution was lower. Moreover, the lateral vibration displacements of vocal fold are smaller, about only several pixels, which will cause a larger error by even 1-pixel segment mistake. Then, the distances from the glottal midline lateral to the right glottal edge at the medial position (glottal amplitudes) are extracted as the displacements of the cover layer of the right vocal fold. These displacements represent the horizontal vibration of the cover layer along the lateral direction of vocal folds.

2) Estimation of Vocal Fold Length: Those pixels in HSV images need to be directly calculated to metric unit for investigating the relationship between vibrations of the body and cover layers.

The distance between anterior commissure and arytenoids, pointed by arrows A and B in Fig. 2, will keep stable during same phonation patterns. After the experiment mentioned earlier, keeping endoscope in the mouth, the scan head is adjusted along the horizontal plane, and imaging region is enlarged to image the anterior commissure and arytenoid, shown in Fig. 2(a) and (b). Then, the subject keeps the same way to pronounce vowel /u:/, and ultrasonic and HSV images are synchronal recorded. The same lengths between anterior commissure and arytenoid are measured from ultrasonic images in metric and

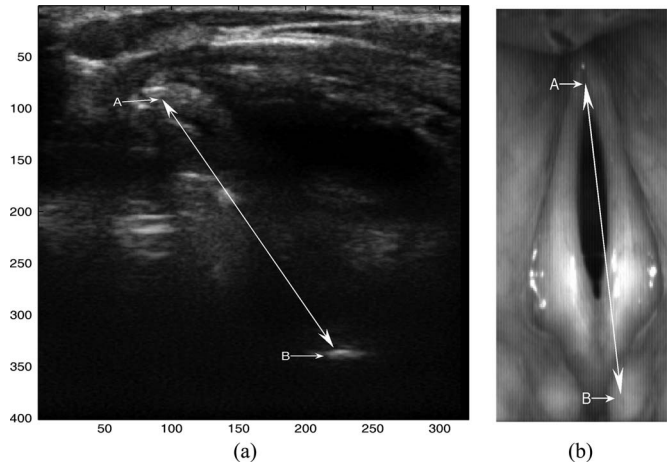


Fig. 2. Estimation of the vocal fold length. (a) HFRU image of vocal fold. (b) HSV image of vocal folds. Arrows A and B point to the anterior commissure and arytenoids, respectively, and the double arrow line means the distance between them.

from HSV images in pixel. At last, the pixels in HSV images are calculated to metric unit based on both measured lengths.

III. Results and Discussions

A. Identifying Structures of Vocal Folds From Ultrasonic Images

The imaging speed of HSV part was 2000 frames/s in this experiment, and the imaging region of HFRU scanner was set as 6-mm width and 30-mm depth. The ultrasonic frequency of transducer was 10 MHz, and the imaging speed of HFRU part was 1000 frames/s based on the size of imaging region, which was several times faster than the phonation pitch of subject. Then, the optical (120×256) and ultrasonic (80×400) images were recorded synchronously during phonation.

Before measuring the vibration of the body layer from ultrasonic images, we first need to identify the vocal fold structures from ultrasonic images. However, there was no clear boundary among those tissue layers in ultrasonic images [25]. Therefore, we first identified the whole vocal fold part in ultrasonic images with an excised canine larynx experiment and then classified both layers according to dynamic imaged ultrasonic series because of their different vibration patterns.

We used an excised canine larynx, whose structures were similar to those of human, to confirm the ability of ultrasound to image vocal fold structures. The relationship between vocal fold structures and B-mode ultrasonic images are shown in Fig. 3. Fig. 3(a) shows an optical image of an excised canine larynx, and Fig. 3(b) shows the corresponding ultrasonic image to Fig. 3(a) imaged in transverse plane. Comparing Fig. 3(a) and (b), we can see that the structures of this excised larynx are imaged by ultrasound. Moreover, the thyroid cartilage (A), lateral part of the thyroarytenoid muscle (B), and water–mucosa interface (C) can be clearly identified from the ultrasonic image. Then, we imaged vocal folds of human *in vivo* by optical and ultrasonic methods, as shown in Fig. 3(c) and (d). The vocal fold structures of human in both images are similar to those of canine in Fig. 3(a)

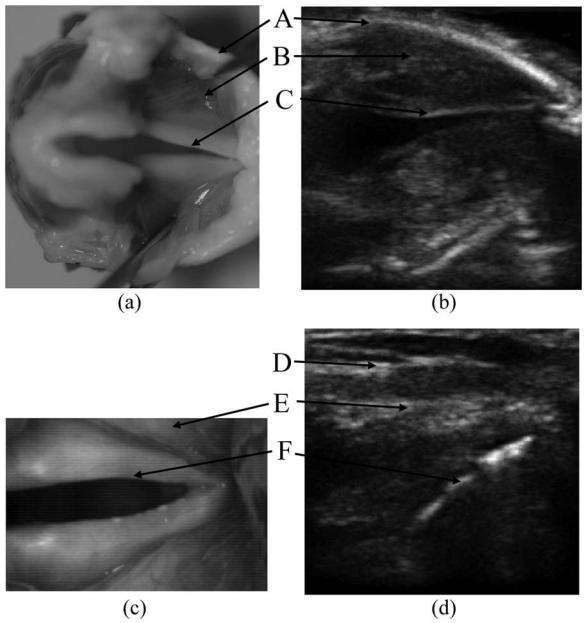


Fig. 3. Identifying vocal fold structures from B-mode ultrasonic images based on optical images. (a) Optical image of an excised canine larynx and (b) the corresponding ultrasonic image to (a) imaged in transverse plane. Arrows A, B, and C point to the structures of canine vocal folds: thyroid cartilage, the lateral part of the thyroarytenoid muscle, and water–mucosa interface, respectively. (c) Optical image of a human vocal fold *in vivo* and (d) the corresponding ultrasonic image to (c). Arrows D, E, and F point to the same structures of human vocal folds, respectively.

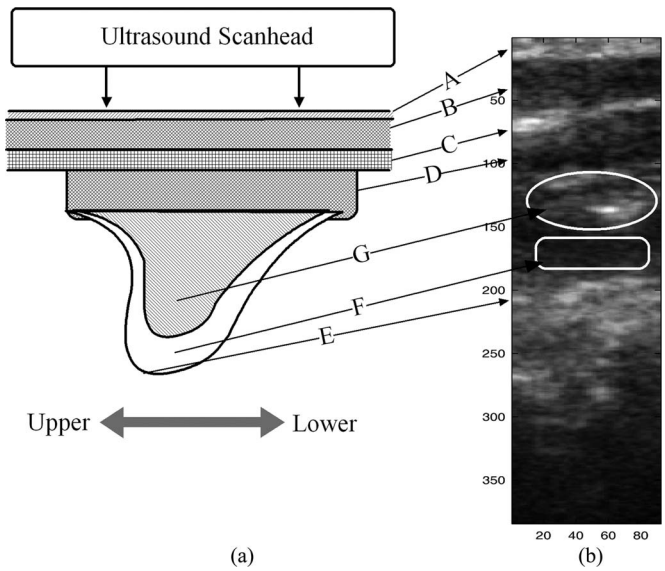


Fig. 4. Relationship between vocal fold structures and B-mode ultrasonic images. (a) Illustration of vocal fold structures. (b) Ultrasonic image of human vocal folds imaged in coronal plane *in vivo*. Arrows point to different structures of vocal folds in both images, where A is skin layer, B is extrinsic muscle of larynx, C is thyroid cartilage, D is the lateral part of the thyroarytenoid muscle, and E is air-mucosa interface. Region pointed by F is Reinke's space, and region pointed by G is the body layer of vocal folds.

and (b), except the skins and muscles in Fig. 3(d), which is outside the thyroid cartilage (d). We can generally draw the region of vocal folds from both images, which is in between the lateral part of the thyroarytenoid muscle and water–mucosa interface. However, it is hard to exactly classify the body layer

and the cover layers from the ultrasonic image [see Fig. 3(d)]. Therefore, we need more information for classification.

Fig. 4(b) shows one ultrasonic image from a dynamic series of human vocal fold vibration in coronal plane selected, and Fig. 4(a) shows the corresponding illustration of the structures of vocal folds in the coronal plane. Moreover, we can easily identify the main structures of larynx base on the anatomy of human larynx and preresults in Fig. 3, such as skin layer (A), extrinsic muscle of larynx (B), thyroid cartilage (C), lateral part of the thyroarytenoid muscle (D), and air–mucosa interface (E). The vocal fold is the part between (D) and (E), which vibrates periodically in dynamic series. The boundary between vocal fold and lateral part of the thyroarytenoid muscle (D) can be classified from the dynamic series because vocal fold is periodic moving and the lateral part of the thyroarytenoid muscle is almost still during vocal fold vibration. The vocal fold region can be divided into two different periodical motion parts according to the different vibration patterns in the dynamic series, which are both the body (G) and cover layers (E) and (F). Moreover, unlike the mucosa surface (E) of the cover layer was brighter in ultrasonic images, the *Reinke'*s space in the cover layer under mucosa will be darker because its tissues contain less scattering elements to scatter ultrasound. Therefore, the Reinke's space is identified from Fig. 4(b) as darker region F, which is in between the air–mucosa interface (E) and the body layer (G).

B. Body-Cover Vibration Measurements

1) Motion Estimation From HFRU Image Series: Both body and cover layers are identified from two neighbored ultrasonic images shown in Fig. 5(a) and (b) based on the brightness and movements in dynamic series. Fig. 5(c) was part of the optical flow field between them, and we can see from it that the optical flows in the body layer are different from other parts. In this image series, the left side corresponds to the upper side of vocal fold, and the right side corresponds to the lower side of vocal folds.

The average relative motions of a chosen region in the body layer between neighbored images during one vibration cycle are listed in Table I. The sign + of U means the vertical direction of motion is from upper side of vocal fold to the lower and − means the contrary. The sign + of V means the direction of motion is along the lateral direction to close glottis and − means the opening direction.

We can draw base on the data from Table I that the motion directions in the body layer region are mainly along the vertical direction of vocal folds. We can also draw that those along the lateral direction is minor, which was also supported by the observations of Tsai *et al.* [23], [25]. HFRU imaging unlike their methods could supply more information of about six time instants in each vibration cycle.

2) Error Analysis for HFRU Imaging Measurements: Errors of the body vibration measurements are caused by both imaging and measuring methods.

The first main factor causing error will be the HFRU imaging methods. There are three main aspects that will affect measurement precision: imaging spatial resolution, temporal resolution, and the placed position of ultrasonic scan head.

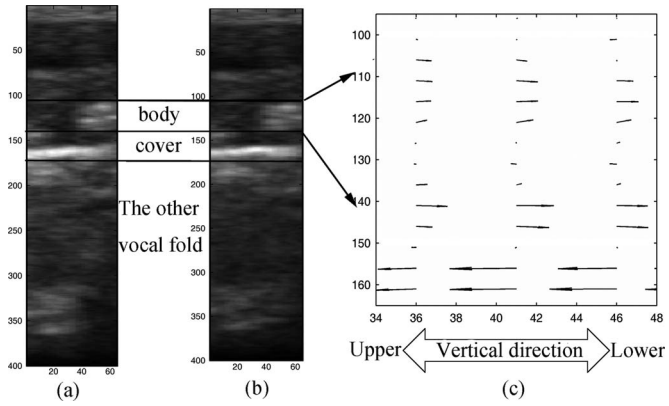


Fig. 5. Two neighbored HFRU images and their corresponding optical flow field. (a) and (b) Two neighbored HFRU images, where the cover-body structure of vocal folds is identified. (c) Part of the optical flow field between (a) and (b), where the left side corresponds to the upper side of vocal folds and the right side corresponds to the lower side of vocal folds.

TABLE I
AVERAGE RELATIVE MOTIONS OF A CHOSEN REGION IN BODY LAYER BETWEEN NEIGHBORED IMAGES IN ONE VIBRATION CYCLE

Images of one cycle	Average optical flow(pixel)	
	U	V
1	+0.94	0.15
2	+0.79	0.02
3	+2.02	0.13
4	+0.43	0.31
5	-2.31	-0.35
6	-1.63	-0.47

The sign + of U means the motion along the vertical direction from upper side of vocal folds to lower and – means the contrary. The sign + of V means the motion is along the lateral direction to close glottis and – means the opening direction.

The spatial resolution related to the amplitude precision is mainly positively related with the frequency of ultrasonic transducer. However, the frequency should be restricted under about 20 MHz to suit for the imaging larynx depth, which will also restrict the motion estimation precision. The optical flow field algorithm may partly improve this precision because of its ability to estimating motion at subpixel level.

The temporal resolution related to precision of vibration period and phase is mainly decided by the imaging speed of ultrasonic machine. This speed is mainly decided by spread speed of ultrasound and the hardware capability. The main temporal resolution restriction in this experiment is the capacity of RAM. Larger capacity of RAM will lead faster imaging speed and larger imaging regions.

Generally, the placed position of scan head related to the imaging direction would affect the measurement results, especially displacements. However, the scan head in this method was placed along the vertical direction, and the vibration of the body layer is also mainly along vertical direction. It would cause less measurement error because of both parallel directions. However, the lateral results would heavily affected by this placed position or angle.

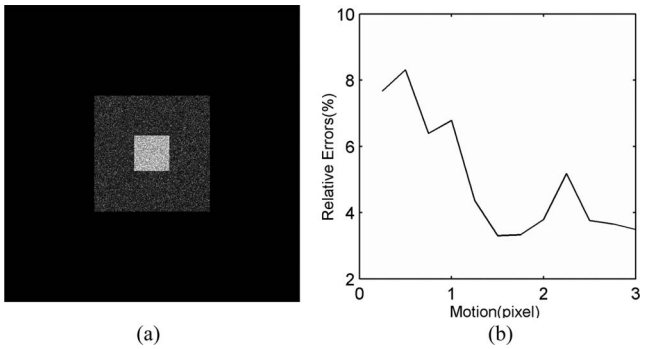


Fig. 6. Relative error analysis for motioned couples of artificial images (resolution 512 × 512) estimated by an optical flow field algorithm at a low resolution level (128 × 128). (a) Simulated image of ultrasound, where gray and white regions are different parts of an object and black region is the background. (b) Relative errors of the optical flow field method according to the different motion displacements from 0.25 to 3 pixels at low-resolution level.

The second main factor causing error will be the optical flow field algorithm, which estimates motions by calculating the spatial and temporal variety of ultrasonic speckles. It is decided by the intrinsic errors of algorithm and the errors related to the constant brightness assumption that is violated on ultrasound data due to speckle noise and acquisition artifacts [29]. These errors have been analyzed in [30] and [31]. However, their analysis was mainly about myocardial motion, which contains large displacements, rotation, torsion, and even escaping from the imaging plane. Unlike myocardial motion, the body motion is translation motion patterns with smaller motion displacements and less transmutation, which will cause less motion estimation errors because the ultrasonic speckle pattern and their brightness keeps stable meeting the constant brightness assumption.

The other estimation error source is the optical flow field algorithm. We can also see from Table I that the displacements are measured at subpixel level. We created a higher resolution (512 × 512) image in order to testify the accuracy of these measurements. The gray and white regions in the image were created following Rayleigh random distributions standing for different parts of an object. Then, the object shifted with different displacements from 1 to 12 pixels and sampled to a low resolution (128 × 128) as ultrasonic images, shown in Fig. 6(a), where the different displacements were from 0.25 to 3 pixels. Then, the displacements were estimated from low-resolution images and relative errors were calculated, shown in Fig. 6(b). We can see from these results that all relative errors are below 9%, which can be accepted for this method.

3) Synchronal Measurements of the Body-Cover Vibration: Subjects in this experiment were three males with healthy voice, aged from twenty four to thirty years old. The lengths between anterior commissure and arytenoids of the subjects were first measured following the estimation approach arrange from 19 to 26 mm. Then, the glottal displacements extracted from HSV images were calculated to metric unit. Fig. 7 shows the measured displacement time series of the body and cover layers during a vibration period of on subject. Fig. 7(a) shows the lateral displacements between glottal midline and vocal fold edge extracted from HSV image series. They are related to the lateral

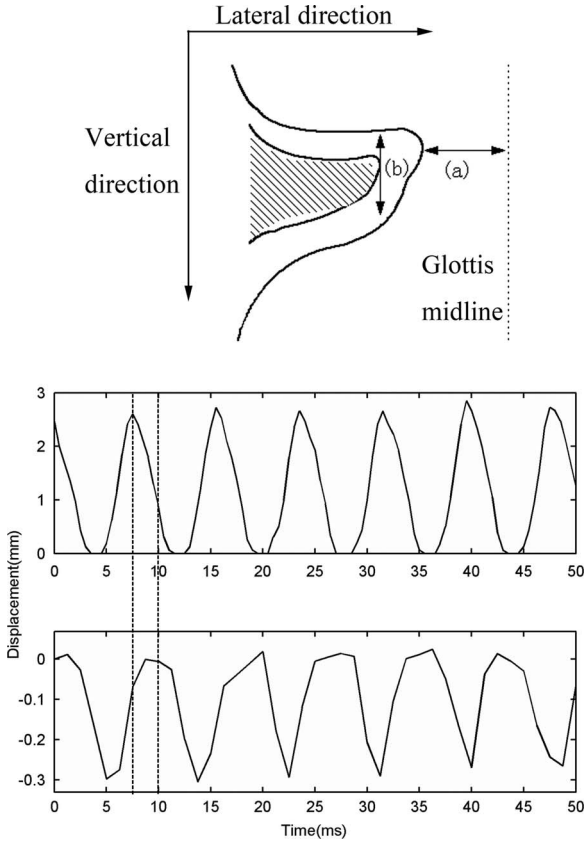


Fig. 7. Measured displacement time series of body and cover layers of a chosen time distance. (a) Maximum glottal displacements extracted from HSV image series, which is related to the cover layer. Their directions are along the lateral direction of vocal folds. (b) Displacements of body layer extracted from HFRU image series, which are calculated based on the motions estimated by the optical flow field algorithm as setting the displacement of first image as zero. Their displacements are along the vertical direction of vocal folds.

motion of the cover layer, and their direction is along the lateral of vocal folds. Fig. 7(b) shows the displacements of body layer extracted from HFRU image series. These displacements were calculated with the motions estimated by the optical flow field algorithm as setting the displacement of first image as zero, and the directions of these displacements are along the vertical direction. The positive slopes of this curve mean that the motion of the body layer is from lower side to the upper side, and the negative slopes mean the contrary. Then, the vibration relations between both layers can be drawn by comparing both time series from Fig. 7(b) as followed.

Furthermore, four groups of body-cover vibration characteristics of the three subjects are analyzed in Table II based on those measurements. Therefore, we can summarize the body-cover vibration characteristics from Table II as follows.

First, the most distinct difference between both vibrations and their directions are different, seeing Table I. The vibration direction of the cover layer is complicated, which is not only along the lateral direction of vocal folds but also along the vertical direction. Unlike the cover layer, the vibration of the body layer is mainly along the vertical direction and little along the lateral direction.

TABLE II
VIBRATION CHARACTERISTICS OF BOTH BODY AND COVER LAYERS MEASURED FROM FOUR DIFFERENT RECORDED DATA

Subject Number	Maximum Motion (mm)		Vibration Period (ms)		Phase delay between boy and cover layers (ms)
	Body layer	Cover layer	Body layer	Cover layer	
1	0.30	2.67	7.7	7.9	2.5
1	0.34	2.41	7.5	7.6	2.7
2	0.26	1.51	6.7	7.1	1.6
3	0.47	3.14	8.1	8.3	2.1

Second, it can be drawn from Table II that vibration periods of both layers are similar but with different vibration phases. The instants from glottis opening to closing are earlier than that of body layer moving from upper side to lower side, which means the opening phases changing to closing.

Third, amplitudes of the cover layer are much larger than that of the body layer, shown in Table II, which can be also qualitatively observed from the dynamic ultrasonic image series. The amplitudes along the lateral direction of the cover layer are several times larger than the maximum ones along the vertical direction of body layer.

These body-cover vibration patterns are not only measured from synchronal imaging data but also can be simulated by a two-layer continuous model [32]. Its simulation showed that increasing body-cover stiffness ratio would gradually reduce the vibration amplitude of body layer and restricts vocal fold motion to the medial surface. On the other hand, the real physiological characteristic of vocal folds is that the body layer containing much muscle and collagen fibers are much stiffer than the cover layer during phonations. Therefore, these conclusions of a physiological model are accordant with our measurements.

IV. CONCLUSION

The fast and complex vibration of vocal folds contains two different vibration patterns based on the body-cover layer structures. Previous research mainly focused on the imaging and measurement of vibration of the cover layer rather than that of the body layer because the cover layer was easier to be imaged. There were few high-speed imaging methods for the vibration of the body layer, let alone the vibration of both layers. Therefore, a method based on a synchronal image system containing HFRU and HSV, and image analysis methods were proposed in order to image and measure the body-cover vibration. The body layer could be identified from HFRU images, and its vibration characteristics were drawn by the estimation of an optical flow field algorithm. The body-cover vibration patterns were discussed from the aspects of vibration directions, amplitudes, and time phases based on the synchronal imaging and measurements during sustained phonation experiments. The factors causing measurement errors were also analyzed. Our next work will focus on the statistical body-cover vibration patterns of regular vibration and will also include investigation of the body-cover vibration during irregular phonation or bifurcations.

REFERENCES

[1] M. Hirano, "Morphological structure of the vocal cord as a vibrator and its variations," *Folia Phoniatr.*, vol. 26, pp. 89–94, 1974.

[2] J. R. Booth and D. G. Childers, "Automated analysis of ultra high-speed laryngeal films," *IEEE Trans. Biomed. Eng.*, vol. 26, no. 4, pp. 185–192, Apr. 1979.

[3] D. G. Childers, D. M. Hicks, G. P. Moore, L. Eskenazi, and A. L. Lalwani, "Electroglottography and vocal fold physiology," *J. Speech Hear. Res.*, vol. 33, pp. 245–254, 1990.

[4] O. Köster, B. Marx, P. Gemmar, M. M. Hess, and H. J. Kunzel, "Qualitative and quantitative analysis of voice onset by means of a multidimensional voice analysis system using high-speed imaging," *J. Voice*, vol. 13, pp. 355–374, 1999.

[5] L. Li, N. P. Galatsanos, and D. Bless, "Eigenfolds: A new approach for analysis of vibrating vocal folds," in *Proc. 2002 IEEE Int. Sym. Biomed. Imag.*, pp. 589–592.

[6] Y. Yan, K. Ahmad, M. Kunduk, and D. Bless, "Analysis of vocal-fold vibrations from high-speed laryngeal images using a hilbert transform-based methodology," *J. Voice*, vol. 19, pp. 161–175, 2005.

[7] T. Wittenberg, M. Moser, M. Tigges, and U. Eysholdt, "Recording, processing, and analysis of digital high-speed sequences in glottography," *Mach. Vis. Appl.*, vol. 8, pp. 399–404, 1995.

[8] A. K. Saadah, N. P. Galtsanos, D. Bless, and C. A. Ramos, "Deformation analysis of the vocal folds from videostroboscopic image sequences of the larynx," *J. Acoust. Soc. Am.*, vol. 103, pp. 3627–3641, 1998.

[9] B. Marendic, N. Galatsanos, and D. Bless, "A new active contour algorithm for tracking vibrating vocal folds," in *Proc. 2001 IEEE. Int. Conf. Proc. Image Process.*, vol. 1, pp. 397–400.

[10] S. Allin, J. Galeotti, G. Stetten, and S. H. Dailey, "Enhanced snake based segmentation of vocal folds," presented at the IEEE Int. Symp. Biomed. Imag., Washington, DC, 2004.

[11] Y. Yan, X. Chen, and D. Bless, "Automatic tracing of vocal-fold motion from high-speed Digital Images," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1394–1400, Jul. 2006.

[12] J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, and M. Döllinger, "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos," *Med. Image Anal.*, vol. 11, pp. 400–413, 2007.

[13] X. L. Qin, S. P. Wang, and M. X. Wan, "Improving reliability and accuracy of vibration parameters of vocal folds based on high-speed video and electroglottography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 6, pp. 1744–1754, Jun. 2009.

[14] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *J. Acoust. Soc. Am.*, vol. 97, pp. 1874–1884, 1995.

[15] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt, "Vibration parameter extraction from endoscopic image series of the vocal folds," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 8, pp. 773–781, Aug. 2002.

[16] R. Schwarz, U. Hoppe, M. S. T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1099–1108, Jun. 2006.

[17] T. Wurzbacher, R. Schwarz, M. Döllinger, U. Hoppe, U. Eysholdt, and J. Lohscheller, "Model-based classification of nonstationary vocal fold vibrations," *J. Acoust. Soc. Am.*, vol. 120, pp. 1012–1027, 2006.

[18] C. Tao, Y. Zhang, and J. J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 794–801, May 2007.

[19] R. Schwarz, M. Döllinger, T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Spatio-temporal quantification of vocal fold vibrations using high-speed video endoscopy and a biomechanical model," *J. Acoust. Soc. Am.*, vol. 123, pp. 2717–2732, 2008.

[20] E. M. Friedman, "Role of ultrasound in the assessment of vocal cord function in infants and children," *Ann. Otol. Rhinol. Laryngol.*, vol. 106, pp. 199–209, 1997.

[21] J. S. Rubin, S. Lee, J. M. Guinness, I. Hore, D. Hill, and L. Berger, "The potential role of ultrasound in differentiating solid and cystic swellings of the true vocal fold," *J. Voice*, vol. 18, pp. 231–235, 2004.

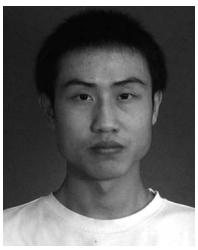
[22] C. C. Huang, L. Sun, S. H. Dailey, S. H. Wang, and K. K. Shung, "High frequency ultrasonic characterization of human vocal fold tissue," *J. Acoust. Soc. Am.*, vol. 122, pp. 1827–1832, 2007.

[23] C. G. Tsai, T. Y. Hsiao, Y. W. Shau, and J. H. Chen, "Towards an intermediate water wave model of vocal fold vibration: Evidence from vocal-fold

dynamic sonography," presented at the Int. Conf. Voice Physiol. Biomech., Tokyo, Japan, Jul. 2006.
[24] C. G. Tsai, T. Y. Hsiao, Y. W. Shau, C. G. Chang, and J. H. Chen, "The in-vivo investigation of vocal fold vibration by dynamic sonography," presented at the 19th Int. Congr. Acoustics, Madrid, Spain, Sep. 2007.
[25] C. G. Tsai, J. H. Chen, Y. W. Shau, and T. Y. Hsiao, "Dynamic B-mode ultrasound imaging of vocal fold vibration during phonation," *Ultrasound Med. Biol.*, vol. 35, pp. 1812–1818, 2009.
[26] B. K. P. Horn and B. G. Schunch, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, 1981.
[27] B. D. Lucas and T. Kanade, "An Iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–779.
[28] C. Q. Davis, Z. Z. Karu, and D. M. Freeman, "Equivalence of subpixel motion estimators based on optical flow and block matching," in *Proc. Int. Symp. Comput. Vis.*, Nov. 1995, pp. 7–12.
[29] E. D. Angelini and O. Gerard, "Review of myocardial motion estimation methods from optical flow tracking on ultrasound data," in *Proc. 28th IEEE Eng. Med. Biol. Soc. Annu. Int.Conf.*, 2006, pp. 1537–1540.
[30] P. Baraldi, A. Sarti, C. Lamberti, A. Prandini, and F. Sgallari, "Evaluation of differential optical flow techniques on synthesizedecho images," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 3, pp. 259–272, Mar. 1996.
[31] M. Suhling, M. Arigovindan, C. Jansen, P. Hunziker, and M. Unser, "Myocardial motion analysis from B-mode echocardiograms," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 525–553, Apr. 2005.
[32] Z. Y. Zhang, "Characteristics of phonation onset in a two-layer vocal fold model," *J. Acoust. Soc. Amer.*, vol. 125, pp. 1091–1102, 2009.

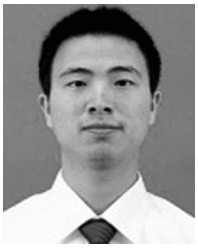
Xulei Qin was born in Shandong, China, in 1982. He received the B.S. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2005, where he is currently working toward the Ph.D. degree in the Department of Biomedical Engineering.

His current research interests include biomedical image analysis and biomechanics measurements of phonation.



Liang Wu was born in Shaanxi, China, in 1985. He received the B.S. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2007, where he is currently working toward the Ph.D. degree in the Department of Biomedical Engineering.

His current research interests include voice rehabilitation and speech processing.



Hujie Jiang was born in Zhejiang, China, in 1984. He received the B.S. degree in biomedical engineering, from Xi'an Jiaotong University, Xi'an, China, in 2007, where he is currently working toward the M.S. degree in biomedical engineering.

His research interests include medical ultrasonic imaging.

Shanshan Tang was born in Shaanxi, China, in 1986. She received the B.S. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2008, where she is currently working toward the Ph.D. degree in the Department of Biomedical Engineering.

Her current research interests include high frame rate and super resolution ultrasonic imaging.

Supin Wang was born in Hebei, China, in 1950. She received the Diploma degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1976.

She is currently a full Professor in the Department of Biomedical Engineering, Xi'an Jiaotong University. She is the author and coauthor of more than 30 publications and three books about electrical engineering. Her current research interests include voice science and ultrasonic imaging.

Mingxi Wan was born in Hubei, China, in 1962. He received the B.S. degree in geophysical prospecting from Jianghan Petroleum Institute, Jingzhou, China, in 1982, and the M.S. and Ph.D. degrees in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 1985 and 1989, respectively.

He is currently a full Professor in the Department of Biomedical Engineering, Xi'an Jiaotong University. From 1995 to 1996, he was a Visiting Scholar and Adjunct Professor at Drexel University, Philadelphia, PA, and Pennsylvania State University, University Park. From 2000 to 2001, he was a Visiting Scholar in the Department of Biomedical Engineering, University of California, Davis. From 2000 to 2010, he was the dean of the School of Life Science and Technology, Xi'an Jiaotong University. He is the author and coauthor of more than 80 publications and three books about medical ultrasound. His current research interests include voice science, ultrasonic imaging, especially in tissue elasticity imaging, contrast and tissue perfusion evaluation, and high-intensity focused ultrasound.

Dr. Wan has received several important awards from the Chinese government and university.