

# Visualizing the movement of the contact between vocal folds during vibration by using array-based transmission ultrasonic glottography

Bowen Jing, Pengju Chigan, Zhengtong Ge, Liang Wu, Supin Wang, and Mingxi Wan<sup>a)</sup>

*The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an, Shaanxi, 710049, People's Republic of China*

(Received 14 December 2016; revised 26 April 2017; accepted 28 April 2017; published online 17 May 2017)

For the purpose of noninvasively visualizing the dynamics of the contact between vibrating vocal fold medial surfaces, an ultrasonic imaging method which is referred to as array-based transmission ultrasonic glottography is proposed. An array of ultrasound transducers is used to detect the ultrasound wave transmitted from one side of the vocal folds to the other side through the small-sized contact between the vocal folds. A passive acoustic mapping method is employed to visualize and locate the contact. The results of the investigation using tissue-mimicking phantoms indicate that it is feasible to use the proposed method to visualize and locate the contact between soft tissues. Furthermore, the proposed method was used for investigating the movement of the contact between the vibrating vocal folds of excised canine larynges. The results indicate that the vertical movement of the contact can be visualized as a vertical movement of a high-intensity stripe in a series of images obtained by using the proposed method. Moreover, a visualization and analysis method, which is referred to as array-based ultrasonic kymography, is presented. The velocity of the vertical movement of the contact, which is estimated from the array-based ultrasonic kymogram, could reach 0.8 m/s during the vocal fold vibration. © 2017 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4983472>]

[CCC]

Pages: 3312–3322

## I. INTRODUCTION

For the purpose of noninvasively visualizing the dynamics of the contact between vibrating vocal fold medial surfaces, an ultrasonic imaging method which is referred to as array-based transmission ultrasonic glottography (ATUG) is proposed in this article. Investigation of the medial surface dynamics of the paired vocal folds is indispensable for comprehensively understanding the patterns of vocal fold vibration in the process of voice production.<sup>1–3</sup> However, one of the major challenges is the inaccessibility of the vocal fold medial surface within intact larynges. The laryngeal endoscopy, which has been widely applied in observation of vocal fold vibration, provides only a superior view of the supraglottal surface of the vocal folds. In order to expose the medial surface and observe it directly, Baer<sup>4</sup> built a glass window at one end of the excised subglottal trachea and recorded the vibration of the medial surface from the inferior view using a stroboscopic dissecting microscope. Berry *et al.*,<sup>1</sup> Doellinger *et al.*,<sup>2,5</sup> Kobler *et al.*,<sup>6</sup> and Chang *et al.*<sup>7</sup> recorded the vibration of the vocal fold medial surface by using a hemilarynx setup. These optical imaging technique-based methods permit the mucosal wave traveling on the medial surface to be easily visualized and recorded. The trajectory of the flesh points on the medial surface is obtained and the tissue displacement is measured. However, these optical imaging technique-based methods cannot be used in

the situation where the intactness of the larynx is required. The application of these methods to investigating vocal fold vibration in human subjects is limited by their invasiveness as well. Besides, the application of stroboscopic x ray laminagraphy<sup>8,9</sup> to the observation of the medial surface is limited by the unnecessary radiation exposure.

Ultrasonography provides an alternative approach for the investigation of the vocal fold vibration.<sup>10–16</sup> Compared with the optical imaging techniques, the major advantage of ultrasonography is its capability to noninvasively visualize and record the vibration of laryngeal tissues beneath the skin surface without interfering with voice production. In our previous work, the frame rate of laryngeal ultrasonography has been increased from only several ten frames/s,<sup>10–13</sup> which is lower than the fundamental frequency [72 to 1000 Hz (Ref. 17)] of vocal fold vibration, to several thousand frames/s.<sup>16</sup> The high frame rate could provide a sufficient sampling of the vibration phases. The vibration displacement of laryngeal tissues including the body layer of the vocal folds has been quantified by using motion estimation algorithms.<sup>13–16</sup> However, as demonstrated by our previous studies,<sup>14–16</sup> the quantitative investigation of the medial surface dynamics using these ultrasonic imaging techniques is limited by the fact that the visualization of the medial surface is impeded by the poor image quality induced by the strong reverberation artifact and the anisotropic reflection of the ultrasound at the medial air–mucosa interface of the vocal folds.

The fact that the deformation of the paired medial surfaces during vocal fold vibration usually results in the contact

<sup>a)</sup>Electronic mail: mxwan@mail.xjtu.edu.cn

between the medial surfaces suggests that observation of the contact between the medial surfaces could be another approach for investigating the medial surface dynamics. As claimed in the previous literature, the amplitude of electroglottographic,<sup>18</sup> and ultrasonic glottographic<sup>19–21</sup> signals is highly correlated with the contact area between the medial surfaces of the paired vocal folds. More importantly, electroglottography, which has been widely used for voice assessment, and ultrasonic glottography are both noninvasive. However, compared with the direct measurements obtained by using the optical imaging techniques mentioned above, the information of medial surface dynamics obtained from electroglottographic and ultrasonic glottographic signals is quite limited. Basically, both the electroglottography and the ultrasonic glottography provide only indicators of the extent of the overall contact between the medial surfaces. It is unachievable to know where the contact happens on the medial surface and how the location of the contact changes along with the deformation of the medial surface during the whole vibration cycle.

In this report, the authors propose an ultrasonic imaging method, ATUG, aiming at visualizing the movement of the contact between the vibrating vocal fold medial surfaces. The ultrasonic imaging method used in previous studies is based on the basic scheme that the emission of the ultrasound and the reception of the reflected ultrasound wave are implemented by using a single ultrasound imaging probe. These imaging techniques belong to the category of the ultrasonic echography. ATUG is different from the ultrasonic echography. The strategy employed by ATUG is using an array of ultrasound transducers to detect the ultrasound wave transmitted, rather than reflected, from one side of the paired vocal folds to the other side through the small-sized contact between the vocal folds. Moreover, it should be noted that the method (ATUG) proposed in the present report is different from the previous ultrasonic glottography<sup>19–21</sup> by which a pair of single-element ultrasonic transducers rather than ultrasonic transducer arrays is employed to register the variation of the amplitude of the ultrasound transmitted through the vibrating glottis.

In this preliminary study, a phantom-based imaging experiment was carried out to validate the proposed method (ATUG) for visualizing and locating the small-sized contact between tissues and compare the proposed method (ATUG) with the conventional B mode ultrasonic echography. Then, the proposed method (ATUG) was used for investigating the movement of the contact between the vibrating vocal folds of excised canine larynges. With the different settings of the subglottal air pressure level, two different types of vocal fold vibration which is the steady-state quasi-periodic vibration and the irregular aperiodic vibration were achieved by using the excised canine larynx setup. Therefore, we were able to investigate how the method performs under different types of vibration. The ultrasonic images were recorded in the coronal plane for the purpose to observe the vertical movement of the contact along the vocal fold medial surface. The high speed glottograph was also recorded as a reference during the experiments.

## II. MATERIALS AND METHODS

### A. Imaging scheme of the ATUG

The basic scheme of ATUG is that an array of ultrasound transducers is used to detect the ultrasound wave propagating from one side of the paired vocal folds to the other side through the small-sized contact between them (Fig. 1). In the present study, the imaging plane of the ultrasound transducer array is the coronal plane that is perpendicular to anterior–posterior direction of the vocal folds [Figs. 1(a-1) and 1(c)]. As the paired vocal folds lie within the vocal tract that is full of air and the ultrasound utilized for medical imaging cannot propagate in air, the only passage for the ultrasound wave to propagate across the glottis is the contact between the medial surfaces of the paired vocal folds when the glottis is closed during phonation.

The ATUG imaging system is composed of two modules which are the ultrasound emitting module and the ultrasound receiving module. In the present study, a linear array probe (L14-5/38, Ultrasonix, Richmond, BC, Canada) equipped on a medical ultrasonic imaging device (SonixTouch, Ultrasonix, Richmond, BC, Canada) is used as the ultrasound emitting module. Another linear array probe (L14-5/38, Ultrasonix, Richmond, BC, Canada) installed on another medical ultrasonic imaging device (SonixRP, Ultrasonix, Richmond, BC, Canada), which is equipped with a 128-channel parallel data receiver (SonixDAQ, Ultrasonix, Richmond, BC, Canada), is used as the ultrasound passive receiving module. The emitting probe and the passive receiving probe are placed against the outer surface of the larynx and face each other during imaging [Fig. 1(a-1)]. It should be noted that the ultrasonic probe used in the present study is a single line of 128 ultrasonic transducer elements. The pitch size is 0.3048 mm. These two modules are synchronized by external triggering pulses. An ultrasound plane wave is emitted by simultaneously triggering each element of the emitting probe. In order to achieve a stronger acoustic penetration through the thyroid cartilage and, thus, a better signal-to-noise ratio, an ultrasound pulse at 5 MHz, which is lower than the center frequency (7.2 MHz) of the probe, is employed for the emission. The pulse duration of the ultrasound plane wave is approximately 1.2  $\mu$ s (6 cycles, 5 MHz). Although using a longer pulse could potentially increase the signal-to-noise ratio, longer pulses were not used in the present study because the probe could be damaged by emitting long pulses at a high pulse repetition frequency (5000 Hz). The ultrasound wave which propagates to the other side of the vocal folds is received, digitized, and stored by the receiving module. The received data are pre-beamformed radio-frequency ultrasonic signals.

It should be noted that the contact between the medial surfaces of the vocal folds actually plays a role as an acoustic slit as well as the only passage for the ultrasound wave to propagate from one side of the vocal folds to the other side [Fig. 1(a-1)]. The size of the contact between the vocal folds, which constantly varies during the vibration, could be comparable to or much smaller than the wavelength (0.3 mm at frequency of 5 MHz and ultrasound speed of 1500 m/s) of the ultrasound. It is hardly achievable to directly locate the slit from the ultrasound wave pattern received by the linear

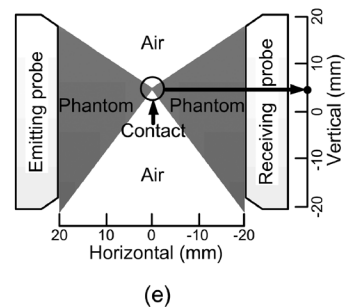
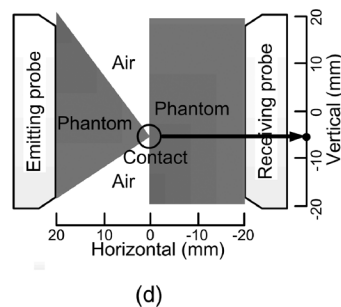
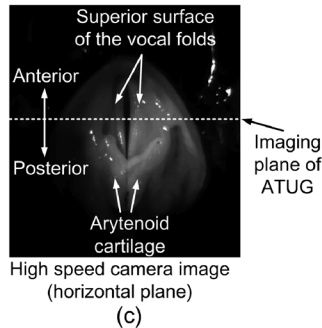
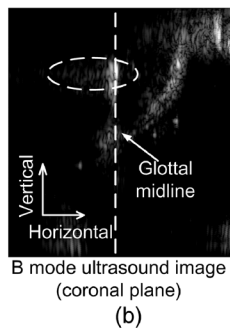
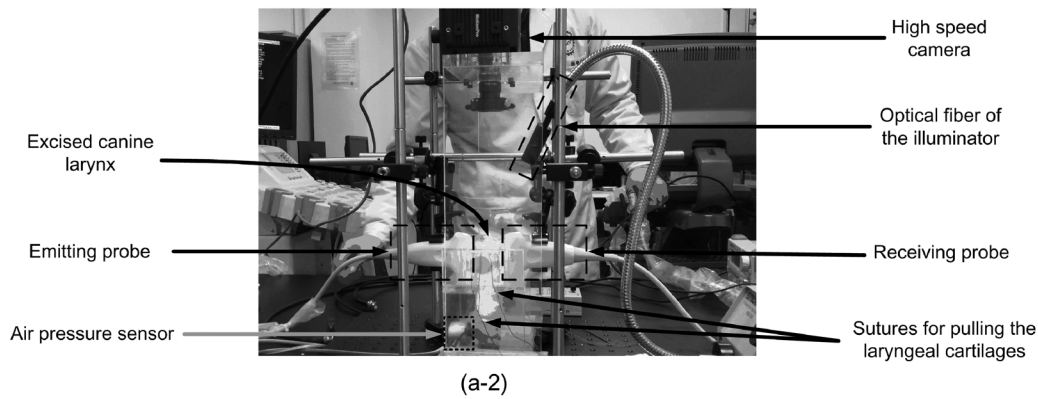
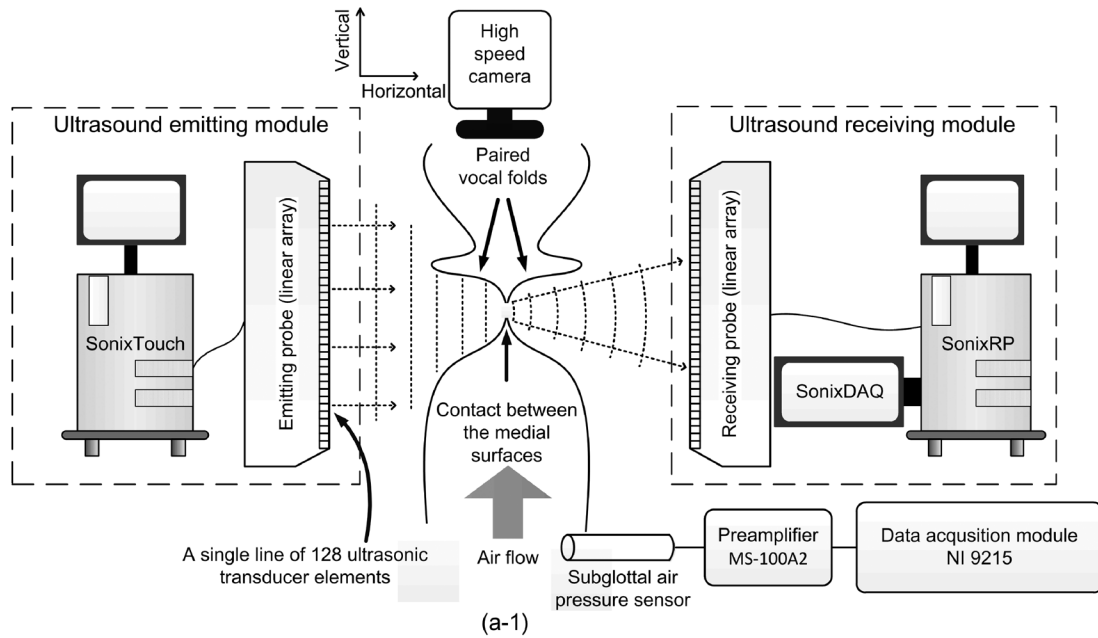


FIG. 1. Diagram illustration of the experimental setup. (a-1) Illustration of the imaging scheme of the ATUG and (a-2) the photo of the experimental setup of the recording system corresponding to (a-1). A sketch of the paired vocal folds in the coronal plane is shown in (a-1). The vertical and horizontal directions are illustrated as well. (b) A frame of B-mode ultrasonic images of the excised canine larynx in the coronal plane. The glottal midline (the vertical dashed line) is located by inserting a thin metal stick into the glottis as a reference. The stick is identified as a hyperechoic object with comet tail artifacts (the dashed ellipse). (c) A frame of photos of the excised larynx taken from the superior view by using the high speed camera. (d) Diagram illustration of the cuboid-and-triangle phantom. (e) Diagram illustration of the dual-triangle phantom.

array transducer [Fig. 2(a)], since the ultrasound wave could diffract strongly rather than keep a straight path when it encounters the slit. Therefore, in order to locate the small-sized contact between the vocal folds, a passive acoustic mapping method based on a passive time-domain delay-and-sum beamforming strategy was employed in offline processing of the received pre-beamformed radio-frequency ultrasonic signal. It should be noted that delay and sum beamformers are,

used to deal with a non-planar ultrasound wave received by the array probe. A similar approach has been successfully used in mapping the source of wide band acoustic wave generated by cavitation events in high-intensity focused ultrasound therapy.<sup>22,23</sup> In the passive acoustic mapping method, the acoustic slit, the small-sized contact, from where the ultrasound wave diffracts during propagation is modeled as the acoustic radiation source in the

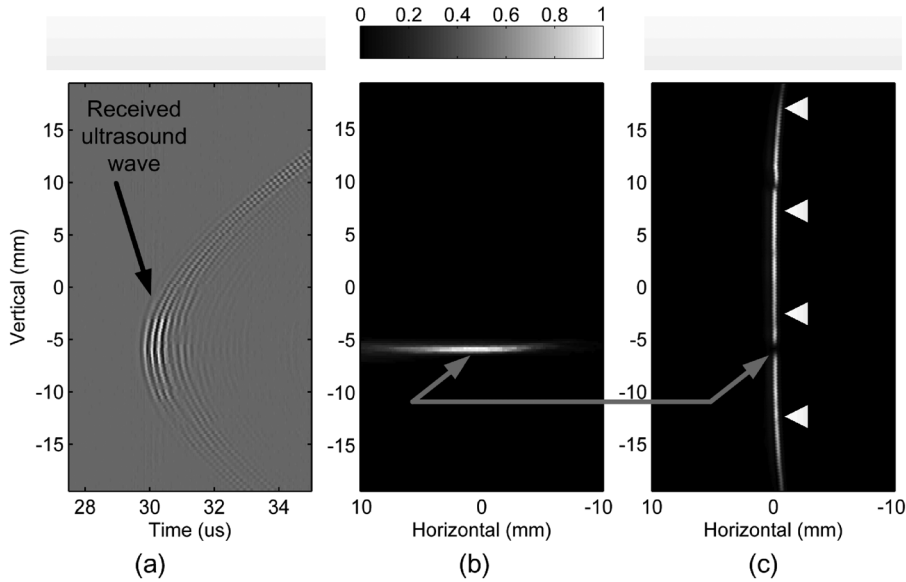


FIG. 2. The results of the investigation using the cuboid-and-triangle phantom. (a) The ultrasound wave received by the receiving module. (b) The image obtained by using ATUG. The image is a frame of normalized acoustic intensity maps. The colorbar in (b) indicates the normalized acoustic intensity. (c) The B mode ultrasonic image obtained by using the receiving module. The four arrow heads point out the flat edge of the cuboid phantom. The location of the contact between the tissue-mimicking phantoms is pointed out by the arrows in (b) and (c), respectively.

acoustic field. The acoustic intensity map of the region of the acoustic radiation source was estimated by

$$I(\mathbf{r}) = \frac{1}{T} \int_{t_0}^{t_0+T} \left( CF(\mathbf{r}, t) \sum_{i=1}^M H_i(\mathbf{r}, t) \right)^2 dt, \quad (1)$$

where  $I(\mathbf{r})$  is the estimated acoustic intensity at the spatial location  $\mathbf{r}$ ,  $M$  is the total number of the transducer elements in the beamforming aperture,  $t_0$  is the time when the ultrasound wave arrives at the ultrasonic probe,  $T$  is the time interval over which the integral is implemented,  $H_i(\mathbf{r}, t)$  is the back-propagated ultrasound signal of the  $i$ th ultrasound receiving transducer element at the spatial location  $\mathbf{r}$ , and  $\mathbf{r}_i$  is the spatial location of the  $i$ th ultrasound receiving transducer element.

$$H_i(\mathbf{r}, t) = \sqrt{d(\mathbf{r}, \mathbf{r}_i)} p_i(t + d(\mathbf{r}, \mathbf{r}_i)/c). \quad (2)$$

$d(\mathbf{r}, \mathbf{r}_i)$  is the distance between the spatial locations  $\mathbf{r}$  and  $\mathbf{r}_i$ ,  $c$ , which is the speed of sound, was set to 1500 m/s for soft tissue,  $p_i(t)$  is the ultrasound pressure sensed by the  $i$ th element of the ultrasound receiving transducer array, and  $CF(\mathbf{r}, t)$  is the coherence factor given by

$$CF(\mathbf{r}, t) = \left| \sum_{i=1}^M H_i(\mathbf{r}, t) \right|^2 / \left( M \sum_{i=1}^M |H_i(\mathbf{r}, t)|^2 \right). \quad (3)$$

The aperture size used in beamforming is approximately 10 mm [ $M=32$  in Eq. (1)] when the acoustic intensity is estimated at each spatial location within the region of interest (ROI). The coherence factor, which has been shown to be effective in improving the lateral resolution in previous studies,<sup>24,25</sup> was used in beamforming. The time-domain integral in Eq. (1) for estimating the acoustic intensity at each spatial location was implemented over the duration of the emitted ultrasound pulse, which is a 6-cycle 5 MHz pulse. The acoustic intensity is estimated every 0.2 mm along both the horizontal and vertical directions in the imaging plane, which means the spatial distance between two neighboring pixels of the image obtained by using ATUG is 0.2 mm.

## B. Experimental setup

### 1. Phantom study

The phantom study is designed to validate ATUG. It is also designed to compare performances of the conventional B mode echography and ATUG. Gelatin-based phantoms mimicking a pair of vocal folds were constructed during the investigation. The phantom is made of Gelatin, Formaldehyde, graphite powder, and distilled water. The concentration of Gelatin is 12% m/V. The concentration of Formaldehyde used for enhancing cross-linking of the phantom and improving its long-term stability is 0.1% m/V. The concentration of the graphite powder that is used to mimic the ultrasound scattering in tissues is 0.3% m/V. This type of phantoms has been shown feasible of mimicking soft tissues for research in the field of ultrasonic medical imaging.<sup>26</sup>

It should be noted that both probes can be used for conventional B mode echography to acquire ultrasonic images from both sides since the probes are both connected to ultrasonic imaging devices. Before the experiment, in order to make the imaging planes of both the ultrasound emitting and receiving probes as coincident as possible, the position of the probes was adjusted using B mode echography. Before ATUG was deployed, the conventional B mode ultrasonic images of the phantom had been acquired on the receiving module of the system. Then, the emitting module of the system was switched to plane wave emission mode and the receiving module of the system was switched to passive listening mode, so that the data of ATUG were recorded.

First, a cuboid-and-triangle phantom was constructed. The small-sized contact between tissues was mimicked by placing the tip of the triangle part against the cuboid part of the phantom [Fig. 1(d)]. The emitting probe was placed against the triangle part of the phantom and the passive receiving probe was placed against the cuboid part of the phantom. With a frame of B mode ultrasonic image of the cuboid part obtained as a reference, we were able to validate ATUG for visualizing and locating the contact.



Second, a dual-triangle phantom was constructed to mimic a pair of vocal folds with a small-sized contact at the tip between two sloping edges [Fig. 1(e)]. This phantom is designed to compare the performances of the conventional B mode echography and ATUG in visualizing and locating the small-sized contact between tissues. More importantly, it is also designed to illustrate the difference between these two imaging methods.

## 2. Imaging of the vibration of excised canine larynges

As opposed to the hemilarynx setup employed in previous studies,<sup>1,2,5-7</sup> two excised canine larynges with both sides of the vocal folds intact were used in the present investigation. The larynx was frozen and stored in a  $-30\text{ }^{\circ}\text{C}$  freezer after being obtained from the research animal-breeding center of the medical school of Xi'an Jiaotong University, and slowly thawed at  $3\text{ }^{\circ}\text{C}$  the day before the experiment. During the experiment, the adduction of the paired vocal folds was achieved by anteriorly pulling the muscular processes of the arytenoid cartilage with sutures. By pulling the anterior tip of the thyroid cartilage, the gap between the paired vocal folds was further closed and the vocal folds are elongated as well. The posterior section of the glottis, which is the gap between the arytenoid cartilages, was further closed by clamping the paired arytenoid cartilages together using a small light-weight alumina clamp. In order to expose the superior surface of the vocal folds to the high speed camera, the supraglottal structures including the epiglottis and the false vocal folds were removed. A polyacrylamide hydrogel-based phantom mimicking soft tissue was constructed around the excised larynx to gain acoustical coupling between the flat surface of the linear array probe and the curved outer surface of the excised larynx. The air which came out of an air compressor (VT2-8C, SIRC, Shanghai, China) was humidified (relative humidity  $>95\%$ ) and heated (temperature  $>32\text{ }^{\circ}\text{C}$ ) by a humidifier (MR850, Fisher & Paykel Healthcare, Panmure, Auckland, New Zealand) and then went through a 1.5 liter air chamber before entering the trachea of the excised larynx and propelling the vocal folds. The length of the remaining trachea of the excised larynx was approximately 8 cm.

The conventional B mode echography is used for adjusting the position of the linear array probes. The imaging plane of both probes is the coronal plane [Fig. 1(a-1)]. A thin metal stick, which was inserted into the closed glottis between the vocal folds from above and identified as a hyperechoic object with comet tail artifacts in the B mode ultrasonic image [Fig. 1(b)], served as a reference during the adjustment of the position of the linear array probes. The imaging plane was approximately at the middle part between the anterior and the posterior commissures of the vocal folds [Fig. 1(c)]. The probes were placed approximately 40 mm away from each other. Therefore, the glottal midline where the vocal folds contact each other during vibration was approximately 20 mm away from both the ultrasound probes.

A high speed camera (Motion Pro Y3-S1, Integrated Design Tools, Tallahassee, FL) was used for recording the vibration from the superior view of the glottis. The frame

rate of the camera is 5000 frames/s. The subglottal pressure level was monitored by using an air pressure sensor (PTL, Glottal Enterprises, Syracuse, NY) and digitized at a sampling rate of 50 KHz by using a data acquisition module (NI 9215, National Instrument, Austin, Texas). The frame rate of ATUG is 5000 frames/s. ATUG and the high speed camera was started simultaneously by using the same trigger source (Trigger box, Photron, Japan). For each of the two excised larynges, multiple recordings of the vocal fold vibration in different settings of the subglottal pressure level were conducted. The duration of each recording was less than 200 ms, so that there was no significant variation of the pressure level of the air compressor during each recording session.

## III. RESULTS

### A. Phantom study

The results of the investigation using the cuboid-and-triangle phantom [Fig. 1(d)] are illustrated in Fig. 2. It should be noted that the only passage for the ultrasound wave to be transmitted from the emitting module to the receiving module is the contact between the two parts of the tissue-mimicking phantoms according to the experiment setup illustrated in Figs. 1(d) and 1(e). The signal received by the receiving module is illustrated in Fig. 2(a). It is shown that the ultrasound wave was transmitted through the contact and received by the receiving module of the ATUG system. The received wave is a diffracted arc-shaped wave rather than the plane wave emitted by the emitting module of the imaging system.

The image [Fig. 2(b)] obtained by using ATUG is a frame of normalized acoustic intensity maps which was generated by employing the passive acoustic mapping algorithm described in Sec. II and normalized by the maximum intensity of all locations within the ROI. The contact resulting in the reception of the diffracted ultrasound wave also resulted in a high-intensity stripe pattern in the ATUG image. Actually, on the other hand, when there is no contact, there is neither any diffracted ultrasound wave received nor any high-intensity stripe pattern observed in ATUG images. The result indicates that the small-sized contact between tissue-mimicking phantoms can be visualized as a diffracted ultrasound wave pattern as well as a high-intensity stripe by using the proposed imaging method (ATUG).

The B mode ultrasonic image illustrating the location of the small-sized contact between the tip of the triangle phantom and the edge of the cuboid phantom is shown in Fig. 2(c). The result that the edge of the cuboid phantom appears to be a hyperechoic line in the B mode image indicates that the ultrasound wave is strongly reflected at the edge. Moreover, the contact is visualized as a small hypoechoic crack on the flat hyperechoic edge of the cuboid phantom in the B mode image [Fig. 2(c)].

As illustrated by Figs. 2(b) and 2(c), along the vertical axis, the location of the high-intensity stripe in the ATUG image coincides with the location of the contact which is visualized in the B mode ultrasonic image [Fig. 2(c)]. The results indicate that it is feasible to visualize the contact and, more importantly, locate the contact on the vertical direction by using ATUG. However, as illustrated in Fig. 2(b), it is

hardly achievable to locate the contact on the horizontal direction in the ATUG image since the contact between the vocal folds is visualized as a high-intensity stripe elongated on the horizontal direction.

The results of the investigation using the dual-triangle phantom [Fig. 1(e)] are illustrated in Fig. 3. The received ultrasound wave pattern, the image obtained by using ATUG and the B mode ultrasonic image of the dual-triangle phantom are shown in Figs. 3(a), 3(b), and 3(c), respectively. It is the same as the previous results [Figs. 2(a) and 2(b)] that the contact resulting in the reception of the diffracted ultrasound wave [Fig. 3(a)] also resulted in a horizontally elongated high-intensity stripe pattern in the ATUG image [Fig. 3(b)]. The small-sized contact between the tissue-mimicking phantoms can be visualized and located on the vertical axis. However, it is quite different from the cuboid-and-triangle phantom that it is unachievable to locate the contact between the triangle-shaped phantoms in the B mode image [Fig. 3(c)]. There are too many hypoechoic cracks on the edge of the triangle phantom.

Moreover, only one of the sloping edges of the triangle phantom is visible in the B mode ultrasonic image [Fig. 3(c)]. The other edge appears to be hypoechoic and is hardly discernable in the image even when different settings of acoustic power, gain, and dynamic range are applied. It should be noted that, as illustrated in Fig. 1(e), the angle between the ultrasound beam and one of the edges, which is the hypoechoic indiscernible one, is much larger than the angle between the beam and the other edge which is hyperechoic and visible in Fig. 3(c).

## B. Investigation of the vibration of excised canine larynges

### 1. Steady-state quasi-periodic vibration of the excised larynx

A recording of the vibration of larynx No. 1 is illustrated in Fig. 4. The high speed video-kymogram, the corresponding high speed photographs, and the ATUG images are

illustrated in Figs. 4(a), 4(b), and 4(c), respectively. The interval between the successive frames illustrated in Figs. 4(b) and 4(c) is 0.8 ms rather than 0.2 ms which is the actual frame-to-frame interval of the recording data. The duration of the whole series of frames in Figs. 4(b) and 4(c) is 10.4 ms which is less than 2 vibration cycles as illustrated in the videokymogram [Fig. 4(a)]. The acoustic intensity maps which are the ATUG images in Fig. 4(c) are normalized by the maximum intensity in the whole series of the ATUG frames rather than a single frame. The recording of the vibration illustrated in Fig. 4 is also uploaded as a video clip (Mm. 1).

**Mm. 1.** The video clip of the recording of the vibration of larynx No. 1. This is a file of type “avi” (3.6 Mb).

The periodically-varying pattern in the videokymogram [Fig. 4(a)] indicates that a steady-state quasi-periodic vibration was recorded. There is a high-intensity stripe in the ATUG image when the glottis is closed as is observed in the high speed photographs (0 to 2.4 ms, 6.4 to 8.8 ms). On the other hand, there is no discernable stripe in the ATUG image when the middle part of the glottis is opened and the paired vocal folds are out of contact with each other in the middle part (3.2 to 5.6 ms, 9.6 to 10.4 ms). The results indicate that the contact between a pair of vibrating vocal folds is visualized as a horizontally elongated high-intensity stripe pattern periodically emerging and disappearing in a series of images obtained by using ATUG (Mm. 1).

Moreover, as illustrated by Fig. 4(c), the brightness of the stripe, which denotes the passively received acoustic intensity, varies from one frame to another during the vibration of the vocal folds. Actually, this is not the first time that the variation of the intensity of the ultrasound transmitted through the vibrating glottis is found. According to the result of previous investigations by using the ultrasonic glottography, the magnitude of the transmitted ultrasound signal varies periodically with the periodic opening and closing of the glottis during vibration.<sup>19–21</sup>

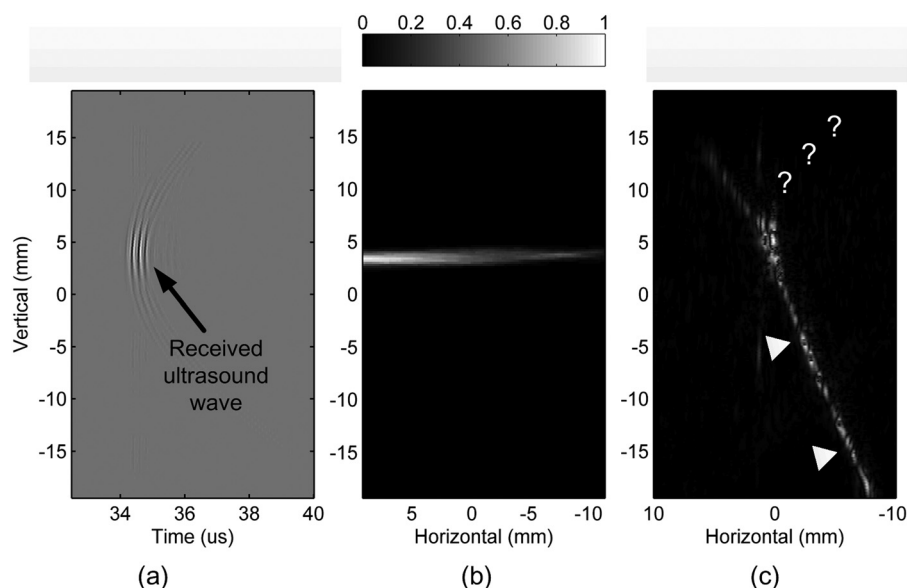


FIG. 3. The results of the investigation using the dual-triangle phantom. (a) The ultrasound wave received by the receiving module. (b) The image obtained by using ATUG. The image is a frame of normalized acoustic intensity maps. The colorbar in (b) indicates the normalized acoustic intensity. (c) The B mode ultrasonic image obtained by using the receiving module. The two arrow heads point out one sloping edge of the triangle phantom. The three question marks indicate the potential location of the other sloping edge.

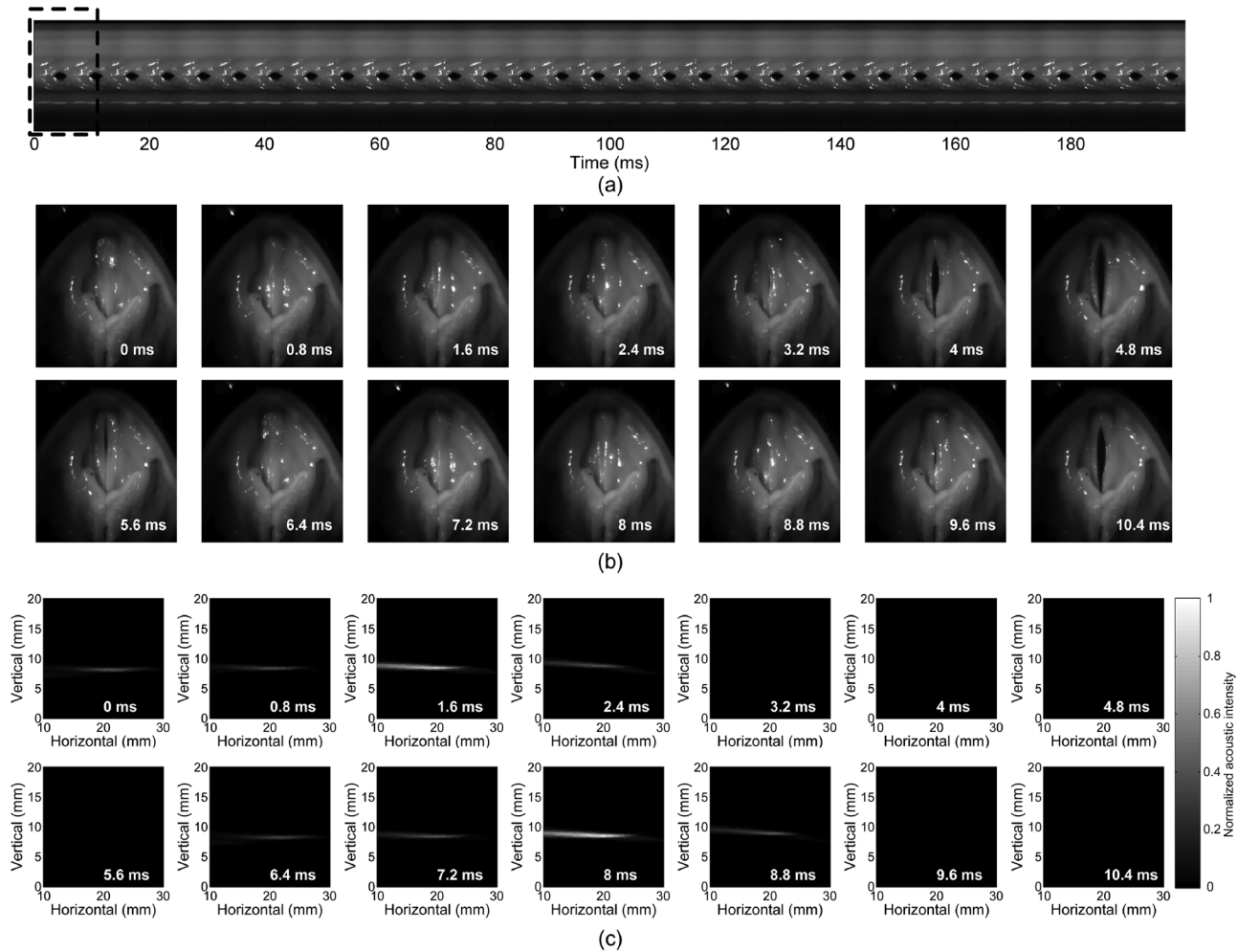


FIG. 4. A recording of the vibration of larynx No. 1. (a) The high speed video-kymogram generated from high speed photos. The dashed rectangle in (a) indicates the time duration of the recorded frames illustrated in (b) and (c). (b) A series of photos of the excised larynx taken from the superior view by using the high speed camera. (c) A series of ATUG images captured in the coronal plane. The recording of the vibration illustrated in this figure is also uploaded as a video clip ([Mm. 1](#)).

More importantly, in Fig. 4(c) and the uploaded video clip ([Mm. 1](#)), it can be seen that the contact-resulted high-intensity stripe in the ATUG images moves upward on the vertical direction from the beginning to the end of the glottal closing phase (6 to 9 ms) within each vibration cycle. The result indicates that there is an upward movement of the contact between the paired vocal fold medial surfaces during the closing of the glottis within each vibration cycle.

## 2. Visualization of the vertical movement of the contact along the glottal midline

In order to give a clear view of the vertical movement of the contact-resulted high-intensity stripe, hundreds of frames of the ATUG image are mapping into a two-dimensional diagram, as illustrated in Fig. 5. According to the previous results obtained from the investigation using the tissue-mimicking phantoms, it is infeasible to locate the small-sized contact on the horizontal direction due to the limited horizontal resolution. Nevertheless, according to the experiment setup, the glottal midline, where the paired vocal folds contact each other during vibration, is located approximately 20 mm away from the receiving probe on the horizontal

direction [the glottal midline illustrated in Fig. 1(b)]. Therefore, a visualization and analysis strategy which is similar to the method used for generating the high speed video-kymogram<sup>27,28</sup> is employed to generate a two-dimensional diagram from the ATUG images. The two-dimensional diagram is referred to as array-based ultrasonic kymogram (ATU kymogram). It can be seen that there is a periodically-varying pattern in the ATU kymogram [Fig. 5(a)] which is obtained from the recording of the periodic vibration of the vocal folds. More importantly, the vertical movement of the contact, which is shown in [Mm. 1](#), is now clearly visible in the two-dimensional ATU kymogram [Fig. 5(c)].

Furthermore, the velocity of the vertical movement is also estimated from the ATU kymogram. First, the displacement of the contact from the moment of the emergence of the high-intensity stripe, which is the moment when the vocal folds begin to contact, to the moment before the disappearance of the stripe, which is the end of the contact, is estimated, as illustrated in Fig. 5(c). It should be noted that the location of the contact is estimated by using a simple peak-picking procedure. The pixel with the maximum absolute value in the vertical line of the ATU kymogram is identified

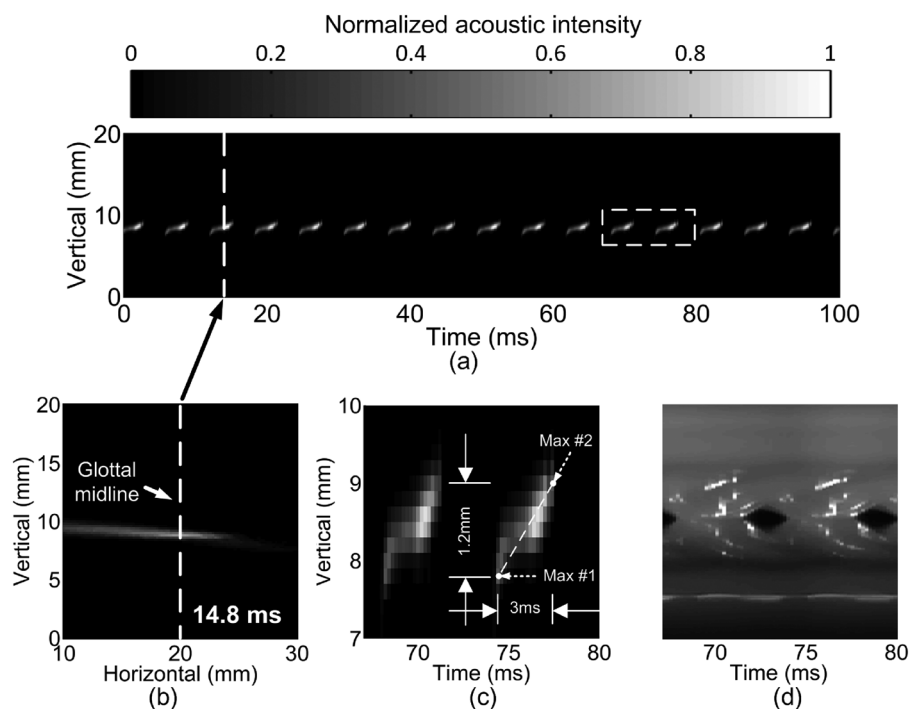


FIG. 5. ATU kymogram which is generated from a series of ATUG images recorded. (a) The ATU kymogram of the vibration of larynx No. 1. (b) A frame of ATUG images captured at 14.8 ms after the start of the recording. (a) combined with (b) illustrates the procedure of mapping the midline of each frame of the ATUG images to the ATU kymogram. (c) An enlarged ATU kymogram of the section in the dashed rectangle in (a). The peak-picking procedure for estimating the vertical displacement of the contact is illustrated in (c). Max #1 and #2 indicate the pixels with the maximum absolute value in the vertical line of the ATU kymogram. The recording of the vibration illustrated in this figure is also uploaded as a video clip (Mm. 1). (d) The recorded high speed video-kymogram corresponding to (c).

and regarded as the location of the contact [Fig. 5(c)]. For each vibration cycle, the moving velocity is obtained by dividing the estimated displacement by the duration of the movement. For each recording, the velocity is estimated from one vibration cycle to another and averaged over a number of cycles recorded during each session. The frequency of contact, which indicates the frequency of the occurrence of the contact between the medial surfaces in the imaging plane, is also obtained from the ATU kymogram. The velocity and the frequency of contact obtained from each recording are illustrated in Fig. 6.

The estimated velocity of the contact between the vocal folds of larynx No. 1 is from 0.05 to 0.51 m/s while the subglottal pressure ranges from 0.77 to 2.4 kPa and the frequency of contact ranges from 156 to 307 Hz [Fig. 6(a)]. The results indicate that there is an increase of the velocity while the subglottal pressure rises. Besides, the estimated velocity of the contact between the vocal folds of larynx No. 2 is from 0.01 to 0.82 m/s while the subglottal pressure ranges from 0.96 to 2.49 kPa and the frequency of contact ranges from 92 to 175 Hz [Fig. 6(b)]. The moving velocity is close to zero and there is hardly any upward movement of the contact when the subglottal pressure is around 1 kPa.

### 3. Irregular vibration of the excised larynx

The vibration of larynx No. 2 at a subglottal pressure level of 1.4 kPa is illustrated by Fig. 7 and the video clip (Mm. 2). The significant variation of the period of vibration cycles, which is illustrated by the video-kymogram [Fig. 7(a)], indicates the significant irregularity of the vibration recorded. Besides, it is feasible as well to directly visualize and quantify the perturbation of the vibration period from the ATU kymogram [Fig. 7(b)]. The difference in vibration period between two successive cycles reaches 6 ms. More importantly, the vertical upward movement of the contact

is also visualized at the beginning of the closing phase of each cycle, which is similar to the results obtained from the investigation of the steady-state quasi-periodic vibration (Figs. 4 and 5). However, the irregular vibration illustrated in Fig. 7 differs from the steady-state quasi-periodic vibration (Fig. 5) in that the vertical movement of the contact appears to significantly slow down and nearly remain still after the upward movement at the beginning of each vibration cycle [Fig. 7(b)].

Mm. 2. The video clip of the recording of the irregular vibration of larynx No. 2. This is a file of type “avi” (3.4 Mb).

## IV. DISCUSSION

The results of the investigation using the dual-triangle phantom indicate that it is feasible to visualize and locate the small-sized contact between tissues by using ATUG while it is difficult to do the same job by using B mode echography (Fig. 3). There are a large number of hypoechoic cracks on the sloping edge as shown in the B mode image [Fig. 3(c)]. According to the previous result (Fig. 2), one of these cracks is the result of the small-sized contact between phantoms. Apart from the contact, another potential reason that could cause the hypoechoic cracks on the sloping edge is the anisotropy<sup>29,30</sup> of the ultrasound reflection at the sloping edge. The problem is that it is unachievable to differentiate the tissue contact-resulted crack from the anisotropic reflection-resulted crack in the B mode ultrasonic image. Therefore, it is unachievable to locate the contact in the B mode image. Actually, the anisotropy of the ultrasound reflection is also demonstrated by the result that one of the edges of the tissue-mimicking phantom is hypoechoic and hardly discernable when the angle between the edge and the



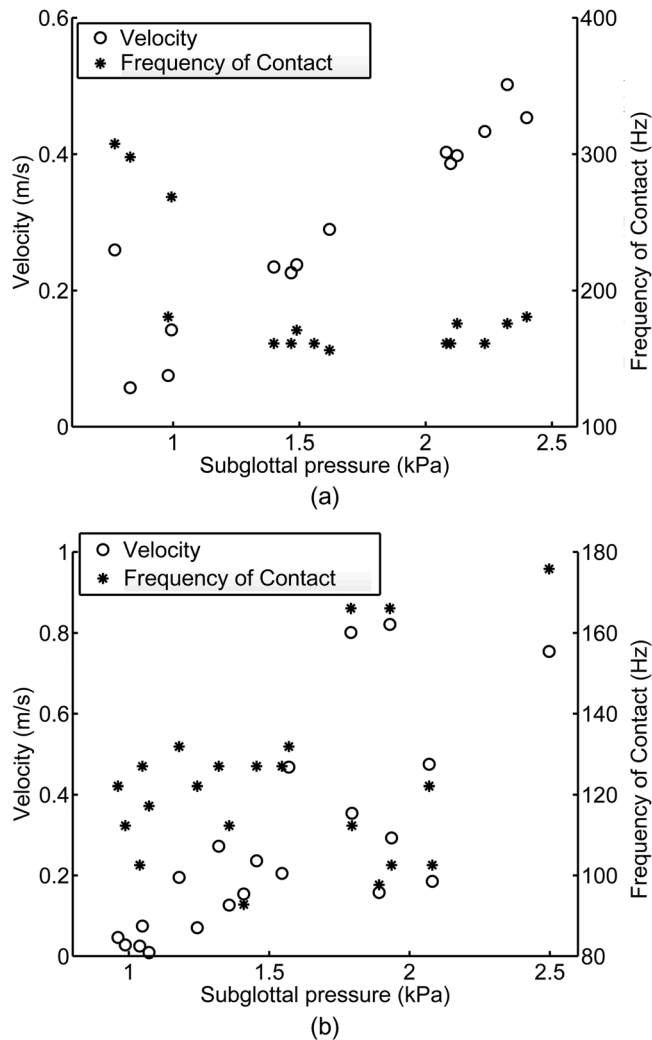


FIG. 6. The vertical velocity of the movement of the contact estimated from the ATU kymogram. (a) The results obtained from larynx No. 1. (b) The results obtained from larynx No. 2. The frequency of contact during vocal fold vibration is also illustrated in (a) and (b).

ultrasound beam is large [Fig. 3(c)]. A similar issue was encountered when the ultrasonic echography was employed to visualize the vibration of the vocal fold medial surface which is an air-mucosa interface. As shown in our previous studies,<sup>14–16</sup> it is unachievable to visualize and track the movement of the contact between a pair of vibrating vocal folds due to the anisotropic reflection of the ultrasound beam at the air-mucosa interface of which the normal direction is drastically changing during vibration.

Although the result of the investigation using tissue-mimicking phantoms indicates that it is feasible to locate the contact on the vertical direction by using ATUG, it is hardly achievable to locate the contact on the horizontal direction due to the poor horizontal resolution of ATUG. The contact between the vocal folds is visualized as a high-intensity stripe elongated on the horizontal direction. It should be noted that the horizontal direction in the ATUG image [Figs. 2(b) and 3(b)] is the axial direction of the linear array probe. According to the results obtain in the previous studies of the passive acoustic mapping method,<sup>23</sup> the axial resolution of the acoustic intensity map is much lower than the lateral

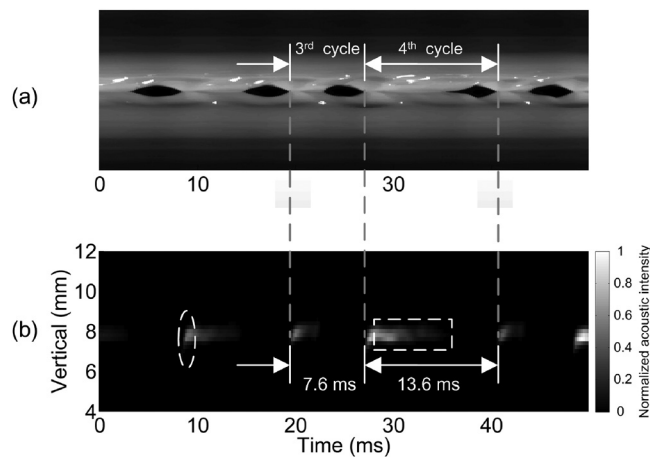


FIG. 7. A recording of the irregular vibration of larynx No. 2. (a) The high speed video-kymogram of the vocal fold vibration. (b) The ATUG kymogram of the vocal fold vibration. The dashed ellipse indicates the beginning of the glottal closing. The dashed rectangle indicates the slowing down of the contact after the beginning of the glottal closing. The recording of the vibration illustrated in this figure is also uploaded as a video clip (Mm. 2).

resolution when a linear array probe is employed for passive acoustic mapping. Therefore, the poor horizontal resolution of the images obtained using ATUG in the present study is not unexpected.

On the other hand, the vertical resolution which is much finer than the horizontal resolution permits the localization of the small-sized contact between the paired vocal folds as well as the tissue-mimicking phantoms. As a consequence, the vertical movement of the contact between the vocal folds can be visualized by using ATUG (Mm. 1 and Mm. 2). However, as is similar to the situation where the high speed video-endoscope is employed for imaging of vocal fold vibration, thousands of frames of the ATUG images, which are usually recorded in a few seconds, could cause the visualization and analysis of the dynamics of the vocal fold contact to be troublesome. The proposed processing strategy, which is used to map a large number of ATUG images to a two-dimensional ATU kymogram, makes the visualization and analysis of the vertical movement of the vocal fold contact much more convenient, as illustrated in Fig. 5.

Actually, the vertical upward movement of the contact between the vocal folds, which is not a new finding in the field of voice production research, has been demonstrated in numerous previous studies. However, there are only several quantitative investigations in which the velocity of this upward movement is quantified. It is shown in Doellinger and Berry's investigation using a human hemilarynx<sup>5</sup> that the velocity of the upward movement of the contact between the medial surface and the plate at the glottal midline reaches approximately 1.2 m/s when the subglottal pressure reaches about 3.2 kPa. In the study conducted by Boessenecker *et al.*, who used the same technique as Doellinger *et al.* did, the velocity of the upward movement of the contact reaches approximately 1.5 m/s.<sup>31</sup> The velocity of the upward movement estimated by using ATUG is slightly lower than that obtained from human hemilarynges. The study,<sup>7</sup> in which the optical coherence tomography is used for capturing the rapid periodic motion of the medial surface of a calf

hemilarynx, shows that the velocity of the upward movement of the contact is from 0.2 to 0.6 m/s which are close to the results obtained in this present study. It should be noted that the velocity of the contact movement obtained in this present study is the mean velocity estimated from the beginning to the end of the contact. The mean velocity over a whole vibration cycle could be lower than 0.1 m/s (Fig. 6) while the transient velocity at certain moments could be much higher than the mean velocity since the slowing down of the movement happens during the vibration (Fig. 7).

In addition, the result (Fig. 7) indicates that the visualization and quantification of the perturbation of irregular vibrations can be achieved by using ATUG as well as high speed video-kymography although extracting the temporal features of the vocal fold contact is not the main purpose of the proposed method. The recording shown in Fig. 7 is a short period (50 ms) of irregular vibration after a period of regular and quasiperiodic vibration. It seems that the vibration pattern during this recording was not stable and alternating between irregular and regular vibrations. A possible cause of the irregularity could be the insufficient medial compression between the paired arytenoids of larynx No. 2. It should be noted that the glottis of both larynges had been fully closed before the recording began. However, when the air pressure was built up to excite larynx No. 2, the posterior part of the glottis opened up and remained unclosed during vibration while the glottis of larynx No. 1 could fully close during vibration. This main difference between larynx No. 1 and No. 2 could be a potential cause of the unstable and irregular vibration pattern of larynx No. 2 shown in Fig. 7 and Mm. 2.

The dimension of the contact could vary considerably during the closing phase of each vibration cycle. However, one of the current limitations of ATUG is that the quantification of the dimension of the contact between the vocal folds is not achievable. Although previous reports of the single-element transducer-based transmission ultrasonic glottography claim that the amplitude of the ultrasonic glottographic signal is correlated with the contact area between the paired vocal folds,<sup>20,21</sup> it is still needed in future work to quantitatively investigate the relationship between the amplitude of the ultrasonic glottographic signal and the vocal fold contact area.

Compared with previous optical imaging techniques, ATUG provides a huge advantage which is its capability to visualize the movement of the contact between the medial surfaces without compromising the intactness of the larynx by removing one of the vocal folds. Therefore, ATUG could potentially be used for noninvasively investigating the dynamics of the vocal fold medial surface in human subjects during phonation. However, a tricky issue, which should be dealt with before human subjects are involved, is the alignment of the imaging planes of the ultrasonic probes. Apparently, the method that a metal stick is inserted into the glottis for adjusting the position of the probes is inconvenient and impractical for the application involving human subjects. Therefore, in the future work, a proper method for adjusting the probe position should be developed.

## V. CONCLUSION

An imaging method referred to as ATUG is proposed in this report. The feasibility of using ATUG to visualize and locate the contact between soft tissues has been validated by conducting an imaging experiment using tissue-mimicking phantoms. Moreover, the results indicate that ATUG offers an advantage over conventional B mode ultrasonic echography in the visualization and localization of the contact between tissues. The proposed method (ATUG) was also employed and tested for visualizing the movement of the contact between the vibrating vocal folds of excised canine larynges. The results indicate that the vertical movement of the contact between the vocal folds can be visualized as the vertical movement of the high-intensity stripe in a series of ATUG images recorded. Moreover, a visualization and analysis method, which is referred to as array-based ultrasonic kymography, is presented. The velocity of the vertical movement, which is estimated from the ATU kymogram, could reach 0.8 m/s during the vibration of the vocal folds.

This present study based on the excised larynx setup is a preliminary step toward the *in vivo* application of ATUG in the investigation of the dynamics of vocal fold medial surfaces during phonation. Since ATUG provides a noninvasive approach for visualizing the vocal fold vibration, the improvement and evaluation of ATUG for clinical applications should also be included in future work.

## ACKNOWLEDGMENTS

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

<sup>1</sup>D. A. Berry, D. W. Montequin, and N. Tayama, "High-speed digital imaging of the medial surface of the vocal folds," *J. Acoust. Soc. Am.* **110**, 2539–2547 (2001).

<sup>2</sup>M. Doellinger, D. A. Berry, and G. S. Berke, "Medial surface dynamics of an *in vivo* canine vocal fold during phonation," *J. Acoust. Soc. Am.* **117**, 3174–3183 (2005).

<sup>3</sup>M. Doellinger, J. Kobler, D. A. Berry, D. D. Mehta, G. Luegmair, and C. Bohr, "Experiments on analysing voice production: Excised (human, animal) and *in vivo* (animal) approaches," *Curr. Bioinform.* **6**, 286–304 (2011).

<sup>4</sup>T. Baer, "Investigation of phonation using excised larynges," Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA, 1975.

<sup>5</sup>M. Doellinger and D. A. Berry, "Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation," *J. Voice* **20**, 401–413 (2006).

<sup>6</sup>J. B. Kobler, E. W. Chang, S. M. Zeitels, and S. H. Yun, "Dynamic imaging of vocal fold oscillation with four-dimensional optical coherence tomography," *Laryngoscope* **120**, 1354–1362 (2010).

<sup>7</sup>E. W. Chang, J. B. Kobler, and S. H. Yun, "Triggered optical coherence tomography for capturing rapid periodic motion," *Sci. Rep.* **1**, 48 (2011).

<sup>8</sup>H. Hollini, R. Coleman, and P. Moore, "Stroboscopic laminagraphy of the larynx during phonation," *Acta Otolaryngol.* **65**, 209–215 (1968).

<sup>9</sup>T. Kusuyama, H. Fukuda, A. Shiotani, H. Nakagawa, and J. Kanzaki, "Analysis of vocal fold vibration by x-ray stroboscopy with multiple markers," *Otolaryng. Head Neck* **124**, 317–322 (2001).

<sup>10</sup>T. Y. Hsiao, C. L. Wang, C. N. Chen, F. J. Hsieh, and Y. W. Shau, "Noninvasive assessment of laryngeal phonation function using Color Doppler ultrasound imaging," *Ultrasound Med. Biol.* **27**, 1035–1040 (2001).

- <sup>11</sup>Y. W. Shau, C. L. Wang, F. J. Hsieh, and T. Y. Hsiao, "Noninvasive assessment of vocal fold mucosal wave velocity using color Doppler imaging," *Ultrasound Med. Biol.* **27**, 1451–1460 (2001).
- <sup>12</sup>T. Y. Hsiao, C. L. Wang, C. N. Chen, F. J. Hsieh, and Y. W. Shau, "Elasticity of human vocal folds measured in vivo using color Doppler imaging," *Ultrasound Med. Biol.* **28**, 1145–1152 (2002).
- <sup>13</sup>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.* **35**, 1812–1818 (2009).
- <sup>14</sup>X. L. Qin, L. Wu, H. J. Jiang, S. S. Tang, S. P. Wang, and M. X. Wan, "Measuring body-cover vibration of vocal folds based on high-frame-rate ultrasonic imaging and high-speed video," *IEEE Trans. Biomed. Eng.* **58**, 2384–2390 (2011).
- <sup>15</sup>S. S. Tang, Y. Y. Zhang, X. L. Qin, S. P. Wang, and M. X. Wan, "Measuring body layer vibration of vocal folds by high-frame-rate ultrasound synchronized with a modified electroglottograph," *J. Acoust. Soc. Am.* **134**, 528–538 (2013).
- <sup>16</sup>B. W. Jing, S. S. Tang, L. Wu, S. P. Wang, and M. X. Wan, "Visualizing the vibration of laryngeal tissue during phonation using ultrafast plane wave ultrasonography," *Ultrasound Med. Biol.* **42**, 2812–2825 (2016).
- <sup>17</sup>D. D. Deliyiski, M. E. G. Powell, S. R. C. Zacharias, T. T. Gerlach, and A. de Alarcon, "Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment," *Biomed. Signal Process.* **17**, 21–28 (2015).
- <sup>18</sup>V. Hampala, M. Garcia, J. G. Svec, R. C. Scherer, and C. T. Herbst, "Relationship between the electroglottographic signal and vocal fold contact area," *J. Voice* **30**, 161–171 (2016).
- <sup>19</sup>S. L. Hamlet and J. M. Reid, "Transmission of ultrasound through larynx as a means of determining vocal-fold activity," *IEEE Trans. Biomed. Eng.* **BM19**, 34–37 (1972).
- <sup>20</sup>N. G. Holmer and H. E. Rundqvist, "Ultrasonic recording of the fundamental frequency of a voice during normal speech," *Ultrasound Med. Biol.* **2**, 123–127 (1976).
- <sup>21</sup>S. L. Hamlet, "Ultrasonic measurement of larynx height and vocal fold vibratory pattern," *J. Acoust. Soc. Am.* **68**, 121–124 (1980).
- <sup>22</sup>M. Gyongy and C. C. Coussios, "Passive spatial mapping of inertial cavitation during HIFU exposure," *IEEE Trans. Biomed. Eng.* **57**, 48–56 (2010).
- <sup>23</sup>K. J. Haworth, T. D. Mast, K. Radhakrishnan, M. T. Burgess, J. A. Kopechek, S. L. Huang, D. D. McPherson, and C. K. Holland, "Passive imaging with pulsed ultrasound insonations," *J. Acoust. Soc. Am.* **132**, 544–553 (2012).
- <sup>24</sup>B. M. Asl and A. Mahloojifar, "Minimum variance beamforming combined with adaptive coherence weighting applied to medical ultrasound imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56**, 1923–1931 (2009).
- <sup>25</sup>C. I. Nilsen and S. Holm, "Wiener beamforming and the coherence factor in ultrasound imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **57**, 1329–1346 (2010).
- <sup>26</sup>J. R. Cook, R. R. Bouchard, and S. Y. Emelianov, "Tissue-mimicking phantoms for photoacoustic and ultrasonic imaging," *Biomed. Opt. Exp.* **2**, 3193–3206 (2011).
- <sup>27</sup>J. G. Svec and H. K. Schutte, "Videokymography: High-speed line scanning of vocal fold vibration," *J. Voice* **10**, 201–205 (1996).
- <sup>28</sup>H. K. Schutte, J. G. Svec, and F. Sram, "First results of clinical application of videokymography," *Laryngoscope* **108**, 1206–1210 (1998).
- <sup>29</sup>J. M. Rubin, P. L. Carson, and C. R. Meyer, "Anisotropic ultrasonic backscatter from the renal cortex," *Ultrasound Med. Biol.* **14**, 507–511 (1988).
- <sup>30</sup>J. J. Wan, F. L. He, Y. F. Zhao, H. M. Zhang, X. D. Zhou, and M. X. Wan, "Non-invasive vascular radial/circumferential strain imaging and wall shear rate estimation using video images of diagnostic ultrasound," *Ultrasound Med. Biol.* **40**, 622–636 (2014).
- <sup>31</sup>A. Boessenecker, D. A. Berry, J. Lohscheller, U. Eysholdt, and M. Doellinger, "Mucosal wave properties of a human vocal fold," *Acta Acust. Acust.* **93**, 815–823 (2007).