

New laryngoscope for quantitative high-speed imaging of human vocal folds vibration in the horizontal and vertical direction

Nibu A. George
Frits F. M. de Mul
Qingjun Qiu
Gerhard Rakhorst
Harm K. Schutte

University Medical Centre Groningen and University of Groningen
Groningen Voice Research Lab
Department of Biomedical Engineering
9700 AD Groningen
The Netherlands
E-mail: ffm@demul.net

Abstract. We report the design of a novel laser line-triangulation laryngoscope for the quantitative visualization of the three-dimensional movements of human vocal folds during phonation. This is the first successful *in vivo* recording of the three-dimensional movements of human vocal folds in absolute values. Triangulation images of the vocal folds are recorded at the rate of 4000 fps with a resolution of 256×256 pixels. A special image-processing algorithm is developed to precisely follow the subpixel movements of the laser line image. Vibration profiles in both horizontal and vertical directions are calibrated and measured in absolute SI units with a resolution of $\pm 50 \mu\text{m}$. We also present a movie showing the vocal folds dynamics in vertical cross section. © 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3041164]

Keywords: vocal folds; laryngeal imaging; laser triangulation; 3D endoscope.

Paper 08066R received Feb. 26, 2008; revised manuscript received Oct. 6, 2008; accepted for publication Oct. 8, 2008; published online Dec. 11, 2008.

Introduction

Optical triangulation is one of the oldest and widely used optical shape measurement methods, particularly in industrial applications, but its use in medical applications is not yet fully utilized. Laser triangulation can be performed in different ways, such as single-point measurements, pointed laser position scanning, or simultaneous multipoint measurements using laser line projection method.¹⁻⁴ One major advantage of laser triangulation is that its working range, resolution, adaptiveness to different applications, etc., can be easily manipulated by controlling the triangulation angle, dimensions, magnification of the imaging optics, etc. However, for clinical *in vivo* applications, the design of a triangulation laryngoscope is much more complex due to the relatively small working space available inside the mouth and, hence, the device has to be compact and easy to use by the clinician.

Vocal folds, the major source of sound production in human beings, are two similar flexible muscular structures covered with soft mucous membrane stretched horizontally inside the larynx. Vocal folds vibration is a complex three-dimensional movement, and its fundamental frequency of closing and opening spread over several hundred Hertz. These complex three-dimensional movements are not yet completely investigated due to the lack of experimental data. One of the most commonly used methods to track the vibration dynamics is stroboscopic imaging, in which a flash stroboscope triggered with the simultaneously acquired audio signal to visualize the vocal folds vibrations in slow motion.⁵ However, during some clinical investigations on subjects having irregu-

lar vocal folds vibrations or for very short phonation, the stroboscopic visualization may fail due to false audio triggering, because the microphone may pick up the fundamental frequency of vibrations as well as its harmonics. Švec and Schutte introduced high-speed videokymography (VKG) in 1996 to visualize the horizontal movements of certain selected positions of vocal folds in very high speed and resolution.^{6,7} Two-dimensional imaging of complete vocal folds using a high-speed camera is also introduced in the same period and in recent years it gained a high momentum due to the technological advances in digital imaging and efficient data storing.^{8,9} However, all these laryngeal imaging methods are only capable of the 2-D imaging of the horizontal movements of the vocal folds. Vertical movements of the vocal folds have never been successfully imaged in a quantitative way *in vivo*, except for some work by Larsson and Hertegård.¹⁰ However, their measurement was a single-point laser spot projection method, with which only one single spot on one of the vocal folds was imaged at very low resolution and speed and, hence, the results were very limited for any further analysis.

In this paper, we report a compact and high-resolution novel laser-sheet triangulation laryngoscope capable of visualizing vertical movements of certain selected positions of both vocal folds, similar to the function of VKG for horizontal movements. Because we used high-speed imaging of the full vocal folds, 3-D visualization of the movement of the upper surface of the vocal folds during vibration can be done by generating a VKG pattern at the same position where we measured the vertical movements. More significantly, with our new device we measured the vertical and horizontal movements of the vocal folds vibration in absolute SI units. The 3-D measurements not only give an insight into the complex

Address all correspondence to: Frits de mul, Faculty of Medicine, Biomedical Engineering, University of Groningen, UMCG, University Medical Center Groningen UMCG-Anthonius Deusinglaan 1, Groningen—9713 AV Netherlands; Tel: xx.31.50.3633138; E-mail: ffm@demul.net

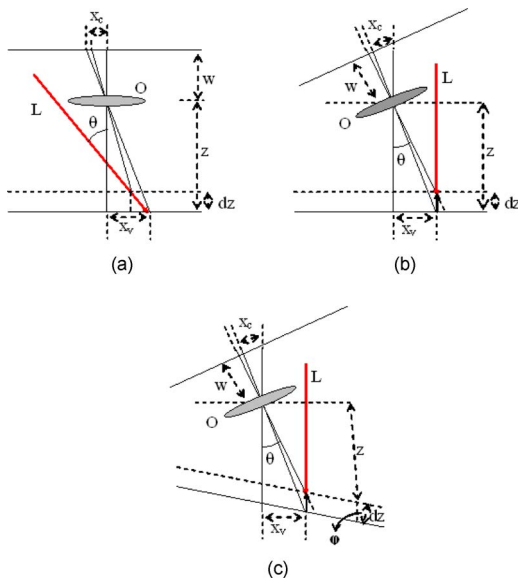


Fig. 1 Possible variations of small-angle laser triangulation. (a) The laser beam is tilted and imaging optics is perpendicular to the moving object surface. (b) Imaging optics is tilted and laser beam is perpendicular to the moving object surface. (c) Both imaging optics and laser beam are tilted with respect to the moving object surface, making an angle φ between the incident laser beam and the object surface. In the above three cases, the triangulation angle θ is the same. Here, L represents the incident laser beam, O represents the imaging optics, z is the object distance, and w is the image distance.

movements of human vocal folds, but they also provide sufficient feedback for numerical modeling and its analysis, and can also help the medical doctors for a quantitative understanding of the pathological conditions of the vocal folds.

Theory

Laser triangulation is projecting a laser beam onto the object surface and collecting its image at an angle with respect to the incident laser beam. The angle between the incident beam and the imaging optics is the triangulation angle. It is well known that the resolution of a triangulation device increases with the increase in the triangulation angle,¹¹ and hence, in most of the triangulation devices a large triangulation angle in the range of 45–60 deg is used. However, in this particular triangulation laryngoscope, the triangulation angle has to be kept very small due to the limited working space available inside the human mouth. Figure 1 shows three different possible configurations of small-angle laser triangulation. In Fig. 1(a), the laser beam incident at an angle to the object surface and the image collection optics is perpendicular to the object surface. In Fig. 1(b), the laser beam is incident perpendicular to the object surface but the image collection optics is placed at an angle to the object surface. In both cases, the angle between the incident laser beam and the imaging optics, θ , is the same.

In the case of Figs. 1(a) and 1(b), the change in image position dx_c on the detector corresponding to an object movement of dz is

$$\frac{dx_c}{dz} = \frac{w \tan \theta}{z} = m \tan \theta \quad (1a)$$

$$\frac{dx_c}{dz} = \frac{w \sin \theta}{z} = m \sin \theta \quad (1b)$$

where m is the magnification factor ($m=w/z$) of the imaging system. For small θ values, Eqs. (1a) and (1b) lead to almost identical results. As an example, with typical values for a triangulation laryngoscope, such as $z \approx 70$ mm, $w \approx 25$ mm, and $\theta \approx 70$, $dz \approx 1$ mm [Eqs. (1a) and (1b)], we arrive at $dx_c = 43.85$ and 43.52 mm, respectively. The basic difference between Figs. 1(a) and 1(b) is that in the case of Fig. 1(a), when z is changing, the beam will illuminate different positions and, hence, the measured vertical positions correspond to different closely spaced positions on the object (vocal folds in the present case) surface. However, if the triangulation angle is small, the change in laser illumination position is very small to take into account. For the same values of z , w , and θ , given above, a vertical movement of 1 mm by the vocal folds will lead to a horizontal positional displacement of the laser spot on the vocal folds surface by 122 mm. But, in the case of Fig. 1(b), the same spot on the vocal folds surface keeps being illuminated as the laser beam is incident perpendicular to the object surface. Therefore, the option in Fig. 1(b) is more favorable than the option in Fig. 1(a). But in real *in vivo* measurements, it is practically impossible to introduce the endoscope into the mouth at a predefined angle with respect to the vocal folds, and hence, an *in vivo* measurement may satisfy either Fig. 1(a) or 1(b) or it can be a composite of both options, which is shown in Fig. 1(c). In Fig. 1(c), both the incident laser beam and the imaging optics are at an angle with respect to the vocal folds surface. If φ is the angle between the vocal folds surface and the incident laser beam, then the measured movement of the vocal folds along the direction of the laser beam will be the component of dz along the direction of the laser beam given by, $dz \cos(\varphi)$. For small tilt angles, this effect will be very small. For a reasonable assumption of maximum tilt angle of 15 deg of the endoscope with respect to the vocal folds, the measured vertical amplitude will only be $\sim 3.5\%$ less than the actual value. Please note that this is the theoretically calculated error due to the endoscope tilt and not the total error in the final measured data.

Design of the Triangulation Laryngoscope

We designed and developed a unique endoscopic laser projection system, which stretches a laser line beam in one direction and projects it as a thin laser sheet at an angle onto the vocal folds surface. The laser projection channel consists of a semiconductor diode laser, an aspheric convex lens to focus the laser beam on to the vocal folds surface, a pair of convex and concave cylindrical lenses to stretch the laser beam in one direction, and a mirror at the end to achieve the desired triangulation angle. In our first prototype, the laser projection channel is firmly attached to a 90-deg rigid endoscope, which acts as the receiving channel. Figure 2 shows the schematic view of the laser triangulation laryngoscope. The optical axes of the two channels are separated by a distance of 9 mm at the tip of the system, and the triangulation angle is ~ 8 deg. From trial experiments, we have concluded that this is the maximum triangulation angle that can be used for a successful *in vivo* measurement. Within a normal working distance of

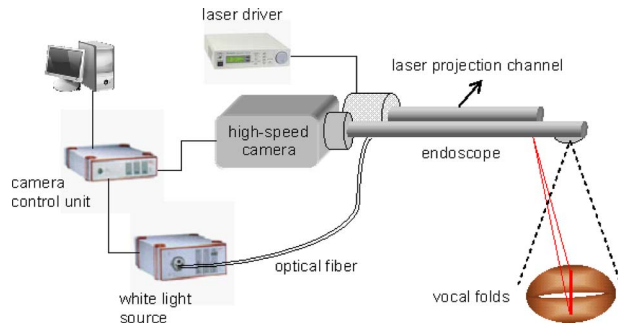


Fig. 2 Schematic view of the laser triangulation laryngoscope. The laser-sheet projection channel is attached to the image-receiving endoscope in its one side. Both the laser projection channel and the receiving endoscope are 10 mm diam and made of stainless steel.

60–70 mm from the tip of the endoscope, the laser line is 18–20 mm long and ~ 0.4 mm wide. We used a semiconductor laser emitting at 653 nm, which deliver an effective laser power density of 1.8 mW/mm^2 , keeping below the exposure limit of 2 mW/mm^2 for human skin recommended by the International Commission on Nonionizing Radiation Protection.¹² The red laser is used because it gives minimum absorption and satisfactory reflectance by the tissue in the visible spectral region. Only a small area around the center of the original laser beam is used for the projection to keep a uniform intensity distribution ($<10\%$ variation) along the length of the stretched laser beam.

A compact high-speed digital color camera (HResEndocam 5562, Richard Wolf GmbH, Germany) is used for recording the images. This camera is developed exclusively for the laryngeal imaging. This system can record images continuously for a maximum of 2 s at the rate of 4000 fps with an image resolution of 256×256 pixels, or for 4 s at the rate of 2000 fps. We used 4000 fps in our measurements to achieve maximum temporal resolution. Temporarily stored data in the camera memory is downloaded to the computer for further analysis. A 300-W white light source coupled to the endoscope through an optical fiber is used to illuminate the larynx to record the high-speed 2-D image of the vocal folds. The white light is used to visualize the vocal folds in its natural color, because the 2-D image is used to investigate any pathological conditions, in addition to the 3-D profiling.

Image Processing

For typical vibration amplitudes of the vocal folds of 1–2 mm, the maximum intensity position of the image envelope of the laser line shifts only one or two pixels. A typical width of this envelope is 20 pixels. Therefore, we have to employ a proper image-processing technique to achieve sub-pixel resolution to track the vocal folds movements with sufficient resolution. The center of the image of the laser line is tracked for this purpose, and a curve fitting method is used to locate the exact center with subpixel resolution. The algorithm written in Matlab reads each frame of the movie, and the intensity profile of the image of the laser along its width is fitted with a Lorentzian curve of the form

$$y = y_0 + \frac{2A}{\pi} \frac{w}{4(x - x_c)^2 + w^2} \quad (2)$$

where A is the area, w is the width, and x_c is the maximum intensity position of the profile. The procedure is repeated for all points along the length of the laser line to locate the peak intensity position of the whole laser. Only the significant area of the image with sufficient laser intensity is used for the fitting procedure. High-frequency noise in the fitted data is filtered out using a two-level Daubechies wavelet (db10) filter. The resulting data provide us the information about the vertical position of vocal folds during phonation, with respect to a baseline position. In fact, the vertical position of the vocal folds is measured as the distance between the endoscope tip and the vocal folds surface. Later, this distance is converted as the vertical height of the vocal folds surface from the baseline. The baseline is the largest distance between the endoscope tip and the vocal folds surface, or in other words, it is the lowest position of the vocal folds during vibration.

The values obtained from the above fitting method are in pixels. In order to quantify the data, we have to calibrate the image to convert the pixel values into SI units. A single-step external calibration method is used for both horizontal and vertical directions. For this purpose, a black-square pattern of known length and width in white background placed at a series of known distances from the tip of the endoscope system, and the images are recorded for each distance. For any particular distance between the endoscope and the object, the projected laser line assumes a unique position in the image and the image size of the calibration pattern is also different for different distances. Because of the complex optical focusing mechanism of the endoscope, the position of the image of the laser line as well as the size of the calibration pattern varies in a nonlinear fashion with respect to the distance between the endoscope and the calibration object surface. A second-order polynomial fit is used to correlate them, and the resulting data are used for the calibration of *in vivo* measurements. In order to compare the horizontal movements at the same position as that of the laser line, we used a Matlab algorithm to generate a 2-D kymogram from the high-speed video.

In Vivo Measurements

The rigid endoscope is inserted through the mouth, and the high-speed video from the camera is displayed in a TV monitor in real time for guiding and proper positioning of the endoscope. The endoscope is positioned in such a way that the laser line is projected near the middle (in the anterior-posterior direction) of the vocal folds, where the vibration amplitudes are expected to be maximum. The recording procedure takes only few seconds, and the recorded high-speed video is processed later. Depending on the quality of the image, the entire image-processing procedure takes 15–20 min using a 4.2-GHz computer. Such a long processing time may restrict the widespread use of this method in real-time clinical investigations, but it can be used for the in-depth analysis of the clinical data.

We have performed a series of *in vivo* measurements on several volunteers. In order to demonstrate the working of our new device, one of the measurements carried out on a vocally

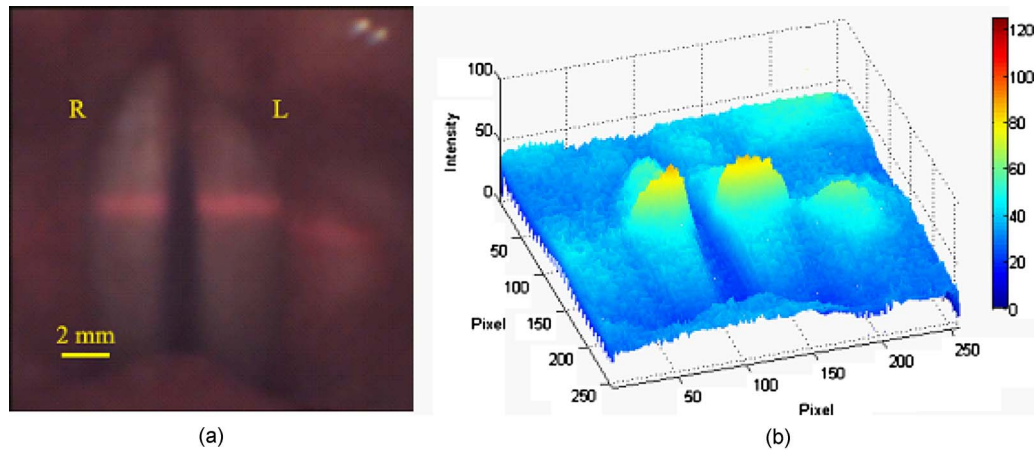


Fig. 3 (a) One frame of the high-speed video showing the vocal folds. The projected laser line is visible on the surface of both vocal folds. Here, R and L represent the right and left vocal folds. (b) Intensity profile of the red component of the same image extracted using Matlab. (Color online only.)

healthy male subject is shown here. The subject produced a sustained phonation at 240 Hz. Figure 3(a) shows one frame of the high-speed movie. This frame of the movie corresponds to partially opened vocal folds. The projected laser beam on the both vocal folds surface is visible in the (color) image. Because we used a red laser beam, the red component of each frame of the video is used for further processing. Figure 3(b) shows the red component of the image shown in Fig. 3(a). Here you can see the intensity profile of the laser beam on top of the relatively flat background intensity profile that is contributed by the red component of the white light. Several consecutive frames of the movie is converted into such intensity patterns and then processed to extract the triangulation data over a period of several opening and closing cycles. Figure 4 shows the laser intensity profile on the right vocal fold at a distance of 1.2 mm from the glottal midline. For visual clarity, only four alternate frames are shown in Fig. 4. The position corresponding to the peak laser intensity is calculated with an accuracy of $\pm 50 \mu\text{m}$. Wavelet filtering of the data further improved the accuracy of the measurement. In Fig. 4, different positions of the peak intensity in the theoretical fit correspond to different vertical positions of the upper surface

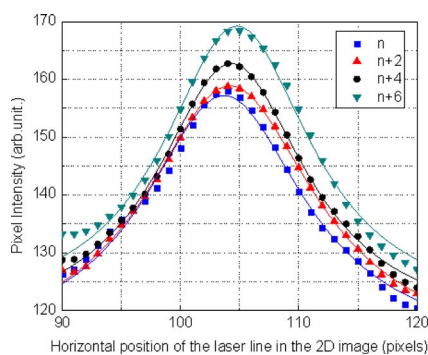


Fig. 4 Laser intensity profile on the right vocal fold at a distance of 1.2 mm from the glottal midline. Here, n represents the number of movie frame. For visual clarity, only alternate movie frames are shown here. The continuous line is the theoretical fit to the experimental data.

of the vocal folds. The absolute values of the vertical positions of the vocal folds are determined from the external calibration procedure described in the image-processing section. Figure 5(a) depicts the colour coded 3-D vibration pattern of the left and right vocal folds. Figure 5(b) shows the horizontal component of the vibration pattern at the same location where the vertical component is measured. The “glottal midline” is an imaginary line of contact of the two folds in their closed position, and normally, the position of the two vocal folds are marked in either direction from this glottal midline. Because glottal midline is used as the reference line for marking the horizontal position, the calculated vertical positions for each lateral distance from this reference line is, in fact, the instantaneous vertical positions of the vocal folds upper surface at a certain time. Hence, the mucosal wave propagation in the lateral direction is also included in the calculation and is clearly visible in the 3-D profile. Video 1 shows the vocal fold’s 3-D vibration dynamics, where, t (in milliseconds) is the time with respect to the first frame. From Fig. 5(a), it is clear that the vertical component of the vibration is slowly decaying as it

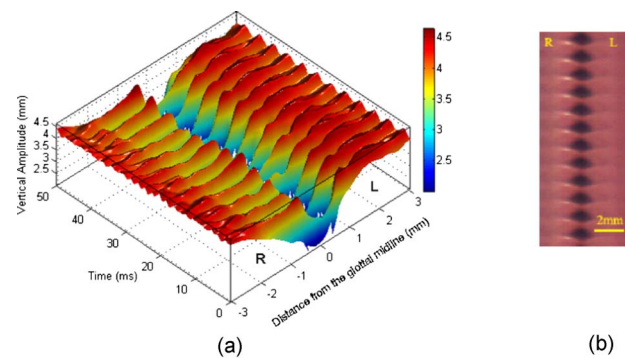
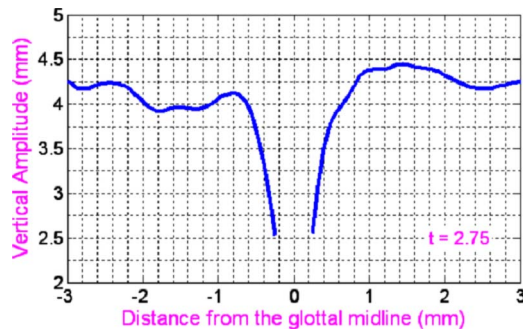


Fig. 5 (a) Colour coded 3D vibration pattern of the vocal folds. The vertical amplitude is normalized with respect to the position of vocal folds surface at its lowest position. The closed phase of the glottis is visible as a continuous line between the left and right vocal folds. (b) Horizontal component of the vibration pattern of both vocal folds. Here R and L represent the right and left vocal folds. (Color online only.)



Video 1 Three-dimensional vibration dynamics of vocal fold (QuickTime, 1.5 MB). [URL: <http://dx.doi.org/10.1117/1.3041164.1>].

goes away from the glottal midline, which might correspond to a mucosal surface wave. The fact that no irregularities in the vibration amplitudes is visible in either the horizontal or vertical component of the vibration dynamics of the left and right vocal folds indicates that our test subject is vocally healthy. Further clinical measurements on subjects having vocal problems are in progress, and the results and detail analysis will be published in a clinically oriented journal.

One of the major problems with the acquired image is the specular reflection from the vocal folds surface. The vocal folds surface is highly moisturized and, depending on the positioning of the endoscope, certain frames or even the entire movie may contain the specular reflection. In most cases, the specular reflection is limited to small-area localized patterns, but in some case, the reflection pattern may spread over the entire vocal folds surface. For small-area localized patterns, the specular reflection of the laser beam in one or two movie frames can be filtered out using intensity thresholding with respect to the surrounding pixels along the length of the laser line. Another difficulty faced while performing the measurement was the reduction of laser intensity during the measurement due to the condensation on the laser channel window. The tip of the endoscope is warmed up to $\sim 40^\circ\text{C}$ using a stream of hot air to avoid the condensation. Considering the fact that the examiner holds the endoscope in his hand, during the measurement a relative movement of the endoscope with respect to the vocal folds is possible. However, this movement is at very low frequency and, because the recording is at very high speed, it is very easy to single out regions with no endoscope movement from the movie. The absolute values of the vibration amplitudes measured with the present method depend on the *in vivo* measurement configuration. During the external calibration procedure, the calibration object surface is moved in a direction perpendicular to the plane of the projected laser beam. However, during an *in vivo* examination, the endoscope may be slightly tilted with respect to the vocal folds surface and this tilt may be different during different measurements. This tilting can cause a small error in the measured absolute values as we discussed in the theory section. In a few cases, patients cannot tolerate a rigid endoscope for laryngeal examination. In these cases, the aforescribed triangulation method may not be fully successful.

Conclusions

With the help of laser triangulation, we have not only extracted the 3-D vibration profile of the vocal folds but also

calibrated and quantified the 2-D image. Vocal fold's vibration amplitudes and velocities are dependent on different parameters, such as the fundamental frequency of phonation, type of phonation, and the sound pressure level.^{13,14} However, the fluctuations in these variables are minimum or negligible within one measurement for a very short time. Hence, a comparison of complete 3-D vibration profile of left and right vocal folds can be very useful in the investigation of vocal folds pathology. The present high-speed triangulation method does not require any audio triggering and, hence can be used to investigate any type of complex vibration dynamics, such as hoarse voice, paralyzed folds, etc., where stroboscopes cannot precisely track such motions. The position of the laser line on the vocal folds can be easily controlled by tilting the endoscope, and hence, this device can be easily used to study the physical size of vocal folds nodules, polyps, or cysts in 3-D, and this can help a phonosurgeon for vocal folds surgeries.

Acknowledgements

This work was supported by the Technology Foundation STW (Stichting Technische Wetenschappen) Project No. 06633, Applied Science Division of NWO (Natuurwetenschappelijk Onderzoek), and the technology program of the Ministry of Economic Affairs, The Netherlands. The authors are thankful to Richard Wolf GmbH, Germany, for providing the high-speed camera.

References

1. F. Chen, G. M. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," *Opt. Eng.* **39**, 10–22 (2000).
2. P. Commer, C. Bourauel, K. Maier, and A. Jager, "Construction and testing of a computer-based intra-oral laser scanner for determining tooth positions," *Med. Eng. Phys.* **22**, 625–635 (2000).
3. L. S. Wang, D. L. Lee, M. Y. Nie, and Z. W. Zheng, "A study of the precision factors of large-scale object surface profile laser scanning measurement," *J. Mater. Process. Technol.* **129**, 584–587 (2002).
4. W. J. Hsueh and E. K. Antonsson, "Automatic high-resolution optoelectronic photogrammetric 3-D surface geometry acquisition system," *Mach. Vision Appl.* **10**, 98–113 (1997).
5. D. M. Bless, M. Hirano, and R. J. Feder, "Videostroboscopic evaluation of the larynx," *Ear Nose Throat J.* **66**, 48–58 (1987).
6. J. G. Švec and H. K. Schutte, "Videokymography: High-speed line scanning of vocal fold vibration," *J. Voice* **10**, 201–205 (1996).
7. Q. Qiu and H. K. Schutte, "A new generation videokymography for routine clinical vocal fold examination," *Laryngoscope* **116**, 1824–1828 (2006).
8. T. Wittenberg, R. Moser, M. Tigges, and U. Eysholdt, "Recording, processing, and analysis of digital high-speed sequences in glottography," *Mach. Vision Appl.* **8**, 399–404 (1995).
9. D. D. Deliyiski, "Endoscope motion compensation for laryngeal high-speed videoendoscopy," *J. Voice* **19**, 485–496 (2005).
10. H. Larsson and S. Hertegård, "Calibration of high-speed imaging by laser triangulation," *Logoped. Phoniatr. Vocol.* **29**, 154–161 (2004).
11. Z. Ji and M. C. Leu, "Design of optical triangulation devices," *Opt. Laser Technol.* **21**, 335–338 (1989).
12. ICNIRP (International Commission on Non-Ionizing Radiation Protection), "Revision of guidelines on limits of exposure to laser radiation of wavelengths between 400 nm and 1.4 μm ," *Health Phys.* **79**, 431–440 (2000).
13. A. M. Sulter, H. K. Schutte, and D. G. Miller, "Standardized laryngeal videostroboscopic rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age," *J. Voice* **10**, 175–189 (1996).
14. I. R. Titze, "On the relation between subglottal pressure and fundamental-frequency in phonation," *J. Acoust. Soc. Am.* **85**, 901–906 (1989).