IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **53** (2008) 2667–2675 doi:10.1088/0031-9155/53/10/015

# Depth-kymography: high-speed calibrated 3D imaging of human vocal fold vibration dynamics

**Nibu A George, Frits F M de Mul, Qingjun Qiu, Gerhard Rakhorst and Harm K Schutte**

Groningen Voice Research Lab, Department of Biomedical Engineering, University Medical Center Groningen and University of Groningen, 9700 AD Groningen, The Netherlands

Received 3 December 2007, in final form 17 March 2008
Published 28 April 2008
Online at stacks.iop.org/PMB/53/2667

**Abstract**

We designed and developed a laser line-triangulation endoscope compatible with any standard high-speed camera for a complete three-dimensional profiling of human vocal fold vibration dynamics. With this novel device we are able to measure absolute values of vertical and horizontal vibration amplitudes, length and width of vocal folds as well as the opening and closing velocities from a single *in vivo* measurement. We have studied, for the first time, the generation and propagation of mucosal waves by locating the position of its maximum vertical position and the propagation velocity. Precise knowledge about the absolute dimensions of human vocal folds and their vibration parameters has significant importance in clinical diagnosis and treatment as well as in fundamental research in voice. The new device can be used to investigate different kinds of pathological conditions including periodic or aperiodic vibrations. Consequently, the new device has significant importance in investigating vocal fold paralysis and in phonosurgical applications.

(Some figures in this article are in colour only in the electronic version)

## Introduction

Laryngeal imaging has been a prominent subject of interest to clinicians and voice researchers for a long time. *In vivo* imaging of human vocal fold (vocal cord) vibration dynamics can shed light onto pathological conditions and its influence on the fundamental mechanism of voice production. Human vocal fold vibration is a complex three-dimensional movement and its fundamental frequency of closing and opening spreads over a wide range. To give an example, in male chest voice vibration mode the frequency range can be 100–250 Hz. One of the most established methods to track this dynamics is stroboscopic imaging, in which a normal video camera is optically or electronically manipulated to visualize the high-speed vibrations in slow motion (Bless *et al* 1987, Qiu *et al* 2005). However, this method has the major drawback that it will not be able to track the vibrations if one or both of the folds vibrate

  

aperiodically. Introduction of high-speed videokymography by Švec and Schutte in 1996 was a breakthrough in this field (Švec and Schutte 1996, Schutte *et al* 1998). With this method people were able to analyze the horizontal movements of certain selected positions of vocal folds in very high speed and resolution. Two-dimensional imaging using a high-speed camera was also introduced in the same period and in recent years it gained a high momentum due to the technological advances in digital imaging (Wittenberg *et al* 1995, Deliyski 2005). These methods have a common major limitation that they can provide us only the two-dimensional image of vocal fold vibration in the horizontal plane. One of the major obstacles in three-dimensional visualization of vocal fold dynamics is the location of vocal folds. An endoscope inserted through the mouth or via the nose is normally used to visualize the vocal folds from the top and hence the vertical vibrations in the same plane as that of the viewing plane cannot be imaged.

Vertical vibration of the vocal folds along the direction of airflow, which is also significant for investigations, is oblivious to the conventional imaging techniques. Knowledge of the absolute values of vibration amplitudes in both directions and the mucosal wave propagation properties can be extremely helpful in phonosurgical applications such as thyroplasty, injection medialization laryngoplasty, etc. Also these data can provide excellent feedback for the theoretical modeling of vocal fold vibration dynamics and hence assist clinicians in getting a deeper insight into the vocal fold pathology.

In recent years, some reports have come out in which the vibration dynamics of excised larynges are visualized and analyzed (Döllinger and Berry 2006, Tao *et al* 2007). Although some attempts to measure the vibration amplitudes of a single point on the vocal fold surface have been reported recently, results are very limited (Larsson and Hertegård 2004). Calibration of 2D images using a laser spot projection and the estimation of vocal folds dimensions in the horizontal direction have been reported previously (Schuster *et al* 2005, Schuberth *et al* 2002, Schade *et al* 2005). We report a sophisticated approach to measure the absolute values of human vocal fold vibration in both the horizontal and vertical directions. A novel laser line-triangulation endoscope was used in conjunction with a high-speed area-scan digital video camera. With our hybrid endoscopic system we could measure the absolute value of the vertical amplitudes along a line, perpendicular to the glottal midline, on both vocal folds. In addition to that, the 2D image of the vocal folds in the horizontal plane is also calibrated with this approach.

## Method

Laser line triangulation is a well-known method for distance measurement and its working principle can be found elsewhere (Ji and Leu 1989). We designed and developed a unique endoscopic laser projection system which can stretch a normal laser beam in one direction and project it as a laser line at an angle onto the vocal fold surface. Figure 1 shows a schematic view of the laser line-triangulation measurement configuration. The laser projection channel consists of a number of optical components which effectively form a cylindrical lens configuration to stretch the laser line in one direction while keeping it focused in the other direction. This laser projection channel is attached to a 90° rigid endoscope which acts as the receiving channel. 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 7.5°. The laser line is aligned in such a way that it appears perfectly parallel to the horizontal pixels in the 2D image. Within a normal working distance of 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 delivers an effective laser power density of 1.8 mW $mm^{-2}$, keeping below the exposure

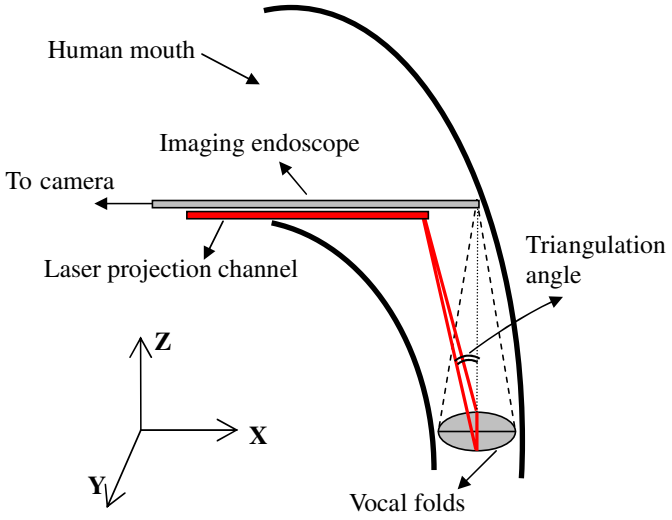


**Figure 1.** Schematic of the laser line-triangulation endoscopic configuration for the *in vivo* imaging the human vocal fold vibrations. The laser beam is optically stretched in one direction to form a line to cover both vocal fold surfaces. The high-speed camera records the image of the surface of vocal folds in the *XY*-plane. Laser line triangulation enables us to extract the vocal fold vibration in the *YZ*-plane, in addition to the horizontal movements in the *XY*-plane.

limit of 2 mW mm$^{-2}$ for human skin recommended by the International Commission on Non-ionizing Radiation Protection (ICNIRP) (Matthes 2000). The red laser is used because it gives maximum reflectance and minimum absorption by the tissue in the visible spectral region. A compact high-speed digital color camera (HResEndocam 5562, Richard Wolf GmbH, Germany) is used for recording the images. The system can record images continuously for a maximum of 2 s at a rate of 4000 fps with an image resolution of 256 $\times$ 256 pixels.

For the small amplitude vertical movement of vocal folds, the position of the image of the laser line shifts only a few pixels. So we have to use a proper image processing technique to achieve a sub-pixel resolution to track the vocal fold 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. 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 $y = y_0 + 2Aw/[4\pi(x - x_c)^2 + \pi w^2]$, 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 laser line at every point on the vocal folds. High-frequency noise in the fitted data is filtered out using a two-level Daubechies wavelet (db10) filter. Since the 'region of interest' for this curve-fitting procedure is only an area belonging to the image of the laser line, only a selected area around the laser intensity profile is used for the processing. The algorithm works automatically and we set several constraints in the fit variables to ensure that the fit is reliable. The algorithm takes about 20 min to process 50 ms duration data of the high-speed images in a 4.2 GHz computer.

The resulting data provide us information about the vertical position of the upper surface of the vocal folds during phonation. The values obtained from the above-described fitting method are in pixels. In order to quantify these 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 the horizontal and vertical directions. For this purpose, a black and white line pattern separated

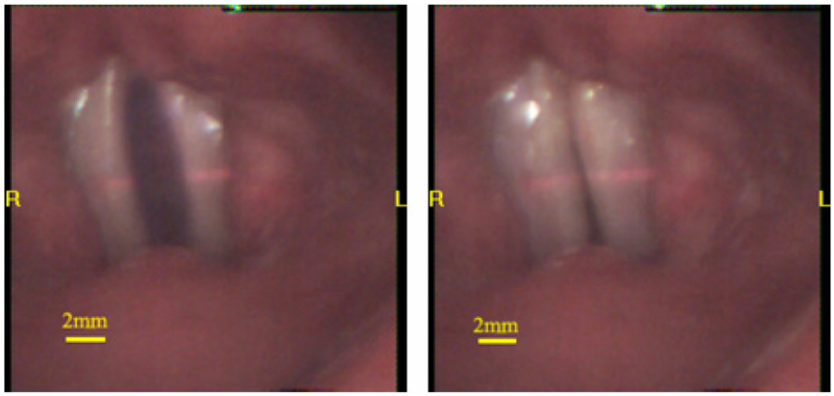


**Figure 2.** Two calibrated frames of the high-speed video showing the maximum open and maximum closed phases of the vocal folds. Inside the picture, R and L represent right and left vocal folds. Projected laser line is visible on the vocal fold surface.

by 5 mm in the horizontal and vertical directions is placed at a series of distance ranging from 50 mm to 80 mm from the tip of the endoscope system and imaged. For each distance, with a step of 2 mm, 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. The focus of the endoscope is adjusted for each distance between the endoscope and the calibration object, consequently the position of the image of the laser line as well as the size of the calibration pattern varied in a nonlinear fashion with respect to the object distance. We have determined that the experimental data points followed a second-order polynomial curve fit. These data are used for the calibration of *in vivo* measurements. The measurements were repeated several times to confirm the accuracy of the calibration procedure and we got consistent results each time. A reverse calibration method, placing the object at an unknown distance and estimating it from the image, is also carried out to validate the accuracy of our device. We have achieved an accuracy of $\pm 50$ $\mu$m. 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 kymogram from the high-speed video. The algorithm picks one specific line in the horizontal direction of the 2D image and displays one over the other continuously for a number of frames of the high-speed video.

One of the measurements carried out on a healthy male subject is demonstrated in this paper. The subject is a non-smoker and he does not have any history of voice problems. The measurements were carried out by a trained medical doctor. The subject was in the sitting position and the rigid endoscope is inserted through the mouth without using any anesthesia.

## Results and discussion

Two frames, showing the maximum open and maximum closed phases, of the high-speed video of the vocal folds obtained from the *in vivo* imaging of the subject during a sustained phonation are shown in figure 2. The laser line projected onto the vocal fold surface is clearly visible in these images (particularly in the color images). During the measurement the endoscope is positioned in such a way that the projected laser line is situated around the middle in the anterior–posterior direction of the vocal folds. This position is selected for projecting the laser line because for a normal healthy person the maximum displacements are expected to occur here. However, from figure 2 it is clear that the glottal axis is tilted slightly

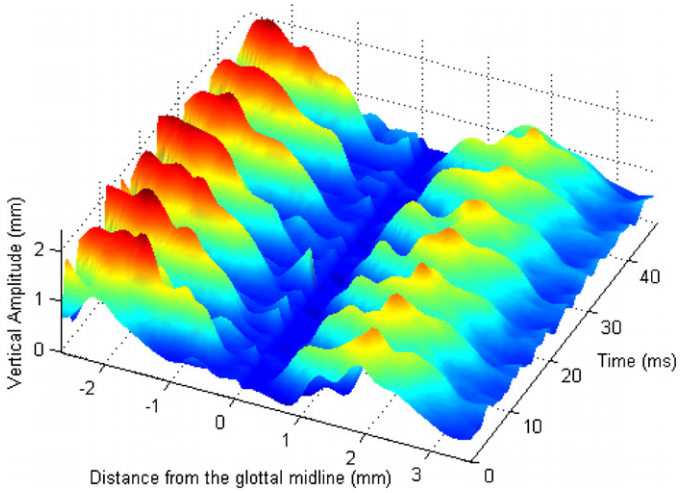


**Figure 3.** 3D view of vocal fold vibration. In the axis 'distance from the glottal midline', zero corresponds to glottal midline, negative and positive values correspond to right and left vocal folds, respectively. Along the vertical direction, zero corresponds to lowest vertical position of the upper surface of the vocal folds.

with respect to the laser line. We have measured the tilt to be about 13°. In an extensive study conducted by Wittenberg *et al*, a glottal axis rotation of less than 15° has negligible influence in the videokymogram generated from a high-speed video (Wittenberg *et al* 2000). We have also adopted the same benchmark in our device measurement, since the vertical vibration is analogous to the horizontal vibrations.

Original movie frames are saved in avi (audio video interleave) format without any compression for further processing. Figure 3 shows the extracted 3D profile of both vocal folds, derived from the position of the image of the laser line. We call this combined display of horizontal and vertical vibrations 'depth-kymogram'. Distances along the lateral direction (width) of the vocal folds are measured from the glottal midline as the origin, negative and positive directions correspond to right and left vocal folds, respectively. Details of the image processing algorithm and the calibration method are described in the method section. Taking into account the fact that the horizontal and vertical amplitudes are different at different positions along the anterior–posterior direction, a videokymogram is generated from the high-speed video at the same position as that of the laser line. Figure 4 shows a color-coded (greyscale) 2D view of the vertical amplitudes together with the kymogram for a side-by-side comparison. In figure 5, the vertical vibration amplitude profile of the left vocal fold alone is shown for the clarity of visualization, especially to explain the mucosal wave propagation. In table 1 absolute values of horizontal and vertical vibration amplitudes, opening and closing velocities of both folds and their length and width are given. The observed values of the opening velocities in the horizontal direction and the vocal folds dimensions are in good agreement with the earlier reports of calibrated 2D imaging of human vocal folds (Larsson and Hertegård 2004, Schuster *et al* 2005, Schuberth *et al* 2002, Schade *et al* 2005). The position at which maximum vertical amplitude occurs and the average value of lateral propagation velocity of the mucosal waves are also measured and presented in the table.

Vocal fold vibration amplitudes and velocities are highly dependent on different phonation attributes such as the fundamental frequency of phonation, type of phonation as well as

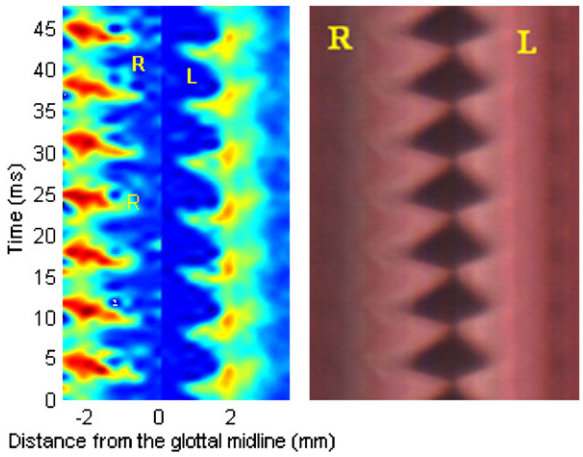


**Figure 4.** Color-coded (greyscale in print) 2D view of the vertical amplitude of both vocal folds (left) shown together with the kymogram (right) from the high-speed video at the same position of the laser line. Inside the pictures, R and L represent right and left vocal folds, respectively. Time duration in both plots is 47.5 ms. Fundamental frequency of the vocal fold vibration is 149 Hz.

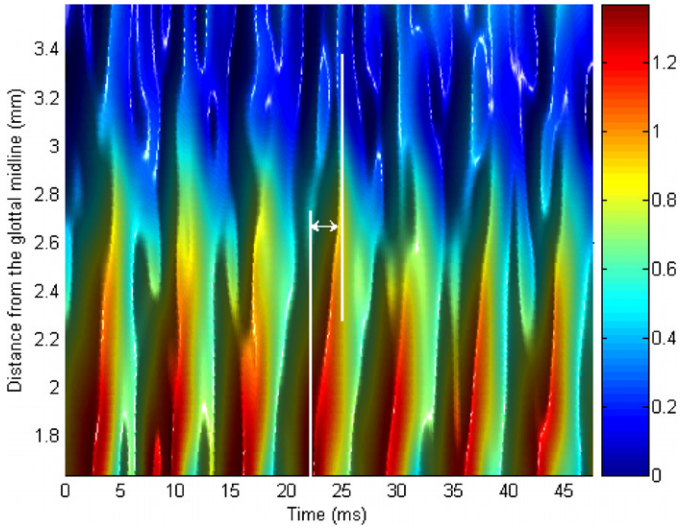


**Figure 5.** Color-coded (greyscale in print) 2D view of the vertical amplitude of left vocal fold. Mucosal wave propagation along the lateral direction is marked between two vertical white lines.

sound pressure level (Sulter *et al* 1996, Titze 1989). However, knowing the values of different vibration parameters within a particular measurement where all the above-mentioned parameters remain constant can help us to understand the relationship between the various vibration parameters and their mutual influence. The male subject under present investigation is a trained professional able to produce a sustained phonation over a long period of time. It is obvious from the table that all the vibration parameters, except the fundamental frequency, show a left–right asymmetry, and the values are higher for the right vocal fold. But, we have not noticed any significant irregularities in the audio signal which is recorded together with the high-speed video. It is remarkable that we could quantitatively measure the left–right vocal

**Table 1.** Different vibration parameters of human vocal folds measured using the laser line-triangulation method.

| | Left vocal fold | Right vocal fold |
|---|---|---|
| Vertical amplitude (mm) | 0.7 | 1.2 |
| Horizontal amplitude (mm) | 1.2 | 1.6 |
| Distance from the glottal midline at which mucosal wave reaches its highest position (mm) | 1.8 | 2.0 |
| Opening velocity (m $s^{-1}$) | 0.63 | 0.92 |
| Closing velocity (m $s^{-1}$) | 0.35 | 0.51 |
| Average value of lateral propagation velocity of mucosal waves (m $s^{-1}$) | 0.51 | 0.81 |
| Length (mm)[a] | 8.2 | 9.5 |
| Width (mm)[a] | 3.2 | 3.9 |
| Fundamental frequency (Hz) | 149 | 149 |

[a] The values reported here are only the width and length of the visible portions of the vocal folds during the closed phase.

fold asymmetry in the vertical and horizontal vibration amplitudes as well as in the opening and closing velocities even though the audio signal does not carry any such information.

From the analysis of the vertical vibration amplitudes, we have observed that the mucosal waves reached a maximum vertical position away from the vocal fold edges (table 1). This finding is in good agreement with the previous *in vitro* measurements on an excised larynx (Döllinger and Berry 2006). In our *in vivo* measurements we achieved a high horizontal position resolution of 77 $\mu$m which in turn produced a large number of measurement points across the entire width of the vocal folds. We suggest that this peak wave position in the mucosal wave propagation occurs at a line, anatomically known as 'linea arcuata superior', which firmly attaches the mucous membrane to the covering of the vocal muscle on the upper side of the vocal folds (Reidenbach 1998). The sudden burst opening of the vocal folds triggers a lateral mucosal wave and the linea arcuata superior offers a resistance to this wave in the lateral direction and the energy is transferred into the vertical direction causing an increased vertical amplitude. Then the mucosal wave amplitude slowly decays as it propagates laterally. The vertical amplitude profile shown in figure 5 is plotted by skipping the data very near to the edges and hence this shows the vertical motion of the upper surface of the vocal fold. This implies that the vertical amplitude profile depicted in this figure represents the true mucosal wave propagation. There is a finite time delay between the generation of mucosal wave at the vocal fold edge and its complete decay. The propagation of the mucosal wave is shown in-between two vertical lines in figure 5. Average propagation velocity of the mucosal waves is calculated from this time delay and the distance covered by it (table 1). For both vocal folds, we found that the mucosal wave velocity is lower than the opening velocity of the respective vocal fold. The amplitude decay of the mucosal wave as well as the decrease in mucosal wave velocity with respect to the vocal fold opening velocity is accounted for by the elastic properties of the vocal folds.

From the calibrated 2D image shown in figure 2 and the kymogram in figure 4, it is clear that there is no complete closure of the glottis. The gap between the two vocal folds in the most closed phase is 0.4 mm. But from the vertical amplitude profiles shown in figures 3 and 4, the gap between the two vocal folds in the most closed phase seems to be 1 mm. This discrepancy is due to the fact that during opening and closing phases the vocal fold edges do not appear very sharp in the images and this caused some ambiguity in the fitting procedure used to detect the position of the laser line. The vocal fold opening and closing velocities

might be too high for the high-speed camera to record sharp images, leading to a slightly blurred image of vocal fold edges. All such data points were skipped and this caused a wider gap to appear in the vertical amplitude profile compared to a 2D videokymogram. These missing data points do not have any influence in any of the calculations and analysis. Also, in figure 3, some peaks are visible near the glottal midline and these peaks are due to erroneous fitting and have no other significance. In the present case, the projected laser line was not covering the full width of the right vocal fold (see figure 2) and hence we were not able to extract the vertical amplitude profile of the right vocal fold completely, until the complete decay of mucosal wave.

A side-by-side comparison of the vertical and horizontal amplitudes (figure 4) shows that for the left vocal fold the maximum vertical amplitude corresponds to closed position and the minimum vertical amplitude corresponds to maximum open position. However, for the right vocal fold the maximum vertical amplitude is delayed by 1 ms with respect to the closed phase and the minimum vertical amplitude is delayed by 2 ms with respect to the open phase. Regarding the length and width measurements, in the present case the vocal folds are not completely visible because the sides are masked in the images by the ventricular folds, whereas the full length is not visible at both ends mainly due to the epiglottis and arytenoid cartilages and hence the values reported here are only the width and length of the visible portions of the true vocal folds. However, in normal clinical examinations a skillful endoscopic imaging can provide a full view of the vibrating vocal folds.

The accuracy of the absolute values of different vibration parameters (amplitude and velocity) measured with the present method depends on the *in vivo* measurement configuration. During the external calibration procedure the calibration object surface is viewed from the top without any titling between the endoscope and the object surface. That means the endoscope axis is parallel to the object surface. However, in real *in vivo* measurements this may not be the case all the time, from measurement to measurement the endoscope may be tilted at different angles with respect to the vocal fold surface. This tilting (in the *XZ*-plane in figure 1) can cause an error in the measured absolute values. A reasonable assumption of a maximum tilt angle of 15° between the endoscope and vocal fold surface leads to an error of 3.5% in the vertical and horizontal amplitudes. Consequently, the opening and closing velocity values can also have the same error in the measurement.

Unlike in stroboscopy, high-speed video recording does not require any audio triggering and hence our triangulation device which makes use of a high-speed camera can be used to investigate the horizontal as well as the vertical vibrations during any complex vibration dynamics such as hoarse voice, paralyzed folds, etc. Also the new device can be very useful in investigating vibration parameters before and after phonosurgical treatments. A trained medical doctor can easily position the laser line on the vocal folds at any desired location and hence this device can be easily used to study vocal fold nodules, polyps or cysts and their local influence on vibration parameters. In the future, a high-resolution camera with higher acquisition rate can give more insight into vocal fold vibration dynamics. The same method can also be further extended for the 3D profiling of other anatomical structures and their dynamics.

## Acknowledgments

This work was supported by the Technology Foundation STW (Stichting Technische Wetenschappen) Project 06633, Applied Science Division of NWO (Natuurwetenschappelijk Onderzoek) and the technology program of the Ministry of Economic Affairs, the Netherlands.

The authors are grateful to Richard Wolf GmbH, Germany, for providing the high-speed camera.

## References

Bless D M, Hirano M and Feder R J 1987 Videostroboscopic evaluation of the larynx *Ear Nose Throat J.* **66** 48–58

Deliyski D D 2005 Endoscope motion compensation for laryngeal high-speed videoendoscopy *J. Voice* **19** 485–96

Döllinger M and Berry D A 2006 Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation *J. Voice* **20** 401–13

Ji Z and Leu M C 1989 Design of optical triangulation devices *Opt. Laser Technol.* **21** 335–8

Larsson H and Hertegård S 2004 Calibration of high-speed imaging by laser triangulation *Logoped. Phoniatr. Vocol.* **29** 154–61

Matthes R *et al* 2000 Revision of guidelines on limits of exposure to laser radiation of wavelengths between 400 nm and 1.4 $\mu$m *Health Phys.* **79** 431–40

Qiu Q J, Schutte H K and van Geest L K 2005 An optical triggering system for stroboscopy *European Patent* Appl. No. 06076120.2, PCT 07/050249, The Netherlands

Reidenbach M M 1998 Subglottic region: normal topography and possible clinical implications *Clin. Anat.* **11** 9–21

Schade G, Kirchhoff T and Hess M 2005 Speed measurements of vocal fold movement *Folia Phoniatr. Logop.* **57** 202–15

Schuberth S, Hoppe U, Döllinger M, Lohscheller J and Eysholdt U 2002 High-precision measurement of the vocal fold length and vibratory amplitudes *Laryngoscope* **112** 1043–9

Schuster M, Lohscheller J, Kummer P, Eysholdt U and Hoppe U 2005 Laser projection in high-speed glottography for high-precision measurements of laryngeal dimensions and dynamics *Eur. Arch. Otorhinolaryngol.* **262** 477–81

Schutte H K, Švec J G and Šram F 1998 First results of clinical application of videokymography *Laryngoscope* **108** 1206–10

Sulter A M, Schutte H K and Miller D G 1996 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** 175–89

Švec J G and Schutte H K 1996 Videokymography: high-speed line scanning of vocal fold vibration *J. Voice* **10** 201–5

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

Titze I R 1989 On the relation between subglottal pressure and fundamental-frequency in phonation *J. Acoust. Soc. Am.* **85** 901–6

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

Wittenberg T, Tigges M, Mergell P and Eysholdt U 2000 Functional imaging of vocal fold vibration: digital multislice high-speed kymography *J. Voice* **14** 422–42