The Laryngoscope
Lippincott Williams & Wilkins, Inc., Philadelphia

High-Precision Measurement of the Vocal Fold Length and Vibratory Amplitudes

Stefan Schuberth, Dipl Phys; Ulrich Hoppe, PD Dr Dr; Michael Döllinger, Dipl Math;
Jörg Lohscheller, Dipl Ing; Ulrich Eysholdt, Prof Dr Dr

Objective/Hypothesis: **Standard laryngoscopy suffers from a lack of information about the actual size of the observed objects (i.e., vocal fold length and oscillating amplitudes). However, there is much interest in absolute measures for both clinical and research purposes. Therefore, a laser projection device has been developed that enables the precise determination of absolute units in endoscopic investigation during respiration and phonation.** ***Study Design:*** **An experimental study in which 9 adults underwent high-speed endoscopy combined with a laser projection device.** ***Methods:*** **The projection system consists of two parallel laser beams with a distance of 3.8 mm. It is mounted on the tip of a rigid endoscope, which is attached to a digital high-speed camera. During development and design, examination situations were taken into account. Two laser spots are projected onto the vocal folds and enable the definition of a metric scale within the endoscopic image. Knowledge-based image processing algorithms were used for evaluation.** ***Results:*** **First measurements of the vocal fold length during phonation were performed in a group of nine healthy male students. The determination of glottal length during phonation agrees with former results. Quantifying vocal fold velocities in absolute units makes it possible to estimate the initial collision forces.** ***Conclusions:*** **The presented laser projection system allows the determination of absolute measures in the larynx. Because of the simple functional principle, the system is open for use without digital high-speed recording as well. Absolute units may also be helpful in voice diagnosis and in monitoring during voice therapy.** ***Key Words:*** **laryngoscopy, phonation, vocal fold length, high-speed camera, laser projection system.**

***Laryngoscope*, 112:1043–1049, 2002**

INTRODUCTION

The recording of vocal fold oscillations during phonation by endoscopic methods is important for the diagnosis of voice disorders. Standard laryngoscopy as the most commonly used imaging method for vocal fold examination allows rough estimation of the vibratory patterns. Using digital high-speed glottography, examinations were extended to irregular vocal fold oscillations.[1] Both methods fail to procure an absolute scale and allow only a determination of the relative size of anatomical structures. Therefore, laryngoscopic recordings performed in separate sessions are not comparable. Consequently, long-term effects (e.g., during voice therapy) usually cannot be observed.

Introducing an absolute scale may avoid these disadvantages. If both the optical properties of the endoscopic optic and the distance between endoscope and the glottal plane were known, an absolute scale could be derived. However, these parameters are not known in the examination situation. Therefore, many attempts were made to yield an absolute scale by using sophisticated techniques. For example, stereo endoscopy, where simultaneously two endoscopes are brought into the throat, provides absolute scaling information.[2] Another technique to receive metric information is the projection of objects with known size into the larynx.

The purpose of the current report is to present a new laser projection system (LPS) for determining metric information. It consists of two parallel laser beams in a fixed distance, which are projected on the glottal plane. The system allows derivation of anatomical structures, vocal fold elongations, and vocal fold velocities. The design of the LPS is described, and the accuracy of this method is estimated. First measurements in a group of nine male subjects are presented.

MATERIALS AND METHODS

Laser Projection System

The LPS was designed by our group to enable determination of absolute measures. It consists of a portable laser unit and a projection device. The laser unit is a battery-driven semiconductor laser diode with a wavelength of 633 nm (GaAs). A light pipe channels the laser beam to the projection device (Figs. 1 and 2).

This device consists of a metal casing with a height of 10 mm, length of 38 mm, and width of 6 mm. The bottom is a quartz glass plate, which is passed by the laser light. A mirror of 50% reflectivity splits the laser beam from the light pipe. One part of

From the Department of Phoniatrics and Pediatric Audiology, University of Erlangen-Nürnberg, Erlangen, Germany.

Supported by the Deutsche Forschungs Gemeinschaft (project DFG EY 15/8-3).

Editor's Note: This Manuscript was accepted for publication February 1, 2002.

Send Correspondence to Stefan Schuberth, Dipl-Phys, Department of Phoniatrics and Pediatric Audiology, Bohlenplatz 21, D-91054 Erlangen, Germany. E-mail: Stefan.Schuberth@gmx.de

the beam propagates through the glass plate, and the second part of the beam is reflected at a mirror with 100% reflectivity parallel to the first one. The beams are adjusted to be parallel with an accuracy better than the beam divergence of 0.17°. They propagate from the device under a compensation angle of 6.8° to the normal of the glass surface (Fig. 2). This angle was introduced because of examination conditions: During examination the glottal plane is tilted a few degrees toward the endoscope. Therefore, this angle guarantees that the laser beams hit the glottal plane perpendicularly.

The laser projection device was designed to be easily mounted on a rigid standard endoscope with 90° optics. The projected laser spots can be visually identified in the endoscopic image and adjusted to the center of the recording by rotating the laser projection device around the endoscope axis (Fig. 1).

The distance between the two laser beams was determined by a calibration measurement. Therefore, grids with increments of 1.0, 0.5, and 0.2 mm were measured with the LPS and the high-speed camera at a distance of 10 cm. A precision of ± 0.02 mm could be determined by repetition of the measurement. The average value was 3.82 mm.

Endoscopic Investigation

For measuring the vocal folds the laser projecting system and a digital high-speed camera were combined.[3,4] The camera has a maximal spatial resolution of 128×128 pixels. To receive reasonable temporal resolution of 3700 images per second, a reduced spatial resolution of 128×64 pixels was used. Hence, the interval between each video frame and the next represents a time interval of 0.27 milliseconds. The high-speed camera allows continuous recording of up to 8 seconds.

The laser projection device was mounted on the tip of a rigid 90° endoscope as shown in Figure 1. The two projected laser points were positioned in the center of the recorded image. For this purpose, the angle between the normal of the glass plate and the optical axis of the endoscope was adjusted by rotating the laser projection device around the endoscope axis.

Nine men between 24 and 35 years of age participated in the study. They had to sustain the vowel [*i:*] after the endoscope was inserted into the oropharynx. They were asked to phonate within the modal register in a comfortable way. Two digital recordings were made of the vocal folds, one in respiratory position and the other during phonation.

Image Processing

Figure 3 shows nine succeeding images (frames) of a high-speed recording. The aim of image processing is to extract the glottal length and the fundamental frequency as a function of time. The fundamental frequency of vocal fold oscillations was determined in the time domain by calculating the inverse of the averaged time of a glottal cycle.

The most important definition of vocal fold length in the literature was taken from Hirano et al.,[5] who measured the distance between the vocal process of the arytenoid cartilage and the thyroid cartilage. This definition is useful for visual examination by eye. For a machine-based evaluation, the glottal length as shown in Figure 3 is more accurate because the glottis can usually be extracted as the darkest region within each image. In addition, the glottal length corresponds more closely to the length of the vibrating part of the vocal folds, which is decisive for voice production. The glottal length was extracted by a knowledge-based algorithm that enhances image quality and allows automatic recognition of the vibratory pattern of the vocal folds.[6] This algorithm was shown to be quite stable and to represent a useful tool for automatic image processing. For the purpose of the current study, the previous algorithm was modified to focus on extraction of the glottal area for a predefined high-speed sequence.[7]

Within the digital video images the laser spots were bright compared with the reflected light of the surrounding tissue and could be identified easily even during voice onset. Compared with other reflections of the light source, the laser spots were brighter and more stable concerning the position within the images.

Once the laser spot position was located in the recording, a two-dimensional gaussian function was fit to the spots. Therefore, a rectangular area of 20×10 pixels around the spot was extracted from the image data. The corresponding matrix of gray values was fitted by a two-dimensional gaussian function using the Levenberg-Marquardt algorithm. The extension of the spots given by the standard derivation usually was between 1 pixel and 3×3 pixels depending on the reflectivity of the tissue. The distance of the maxima of the gaussian functions in pixels corresponds to the previously mentioned distance of 3.8 mm. If the laser spots are projected onto the vocal folds, this method allows the direct measurement of all anatomical structures within the glottal plane.

RESULTS

Glottal Length During Phonation

The glottal length as a function of fundamental frequency for nine male subjects is shown in Figure 4. For each subject, the length of the glottis was determined

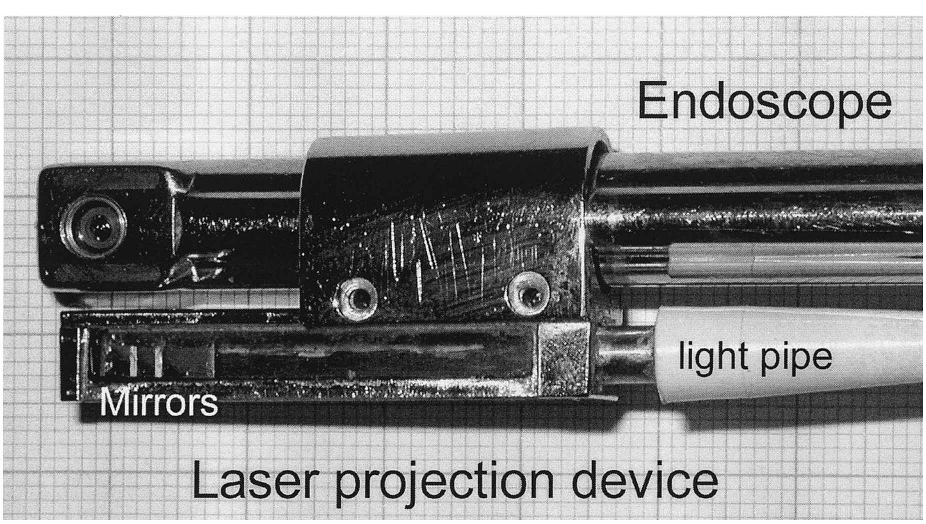

Fig. 1. The laser projection device consists of a light pipe channelling the laser light and two mirrors for beam splitting. A metal casing and a glass plate protect the device against disinfection agents. The device can be easily mounted onto a standard endoscope. A grid of 1×1 mm in the background shows the size of the device. The endoscope is a rigid standard type with 90° optics and a diameter of 9 mm (Wolf Corp., Knittlingen, Germany). A 250-W Xe-light lamp served as light source.

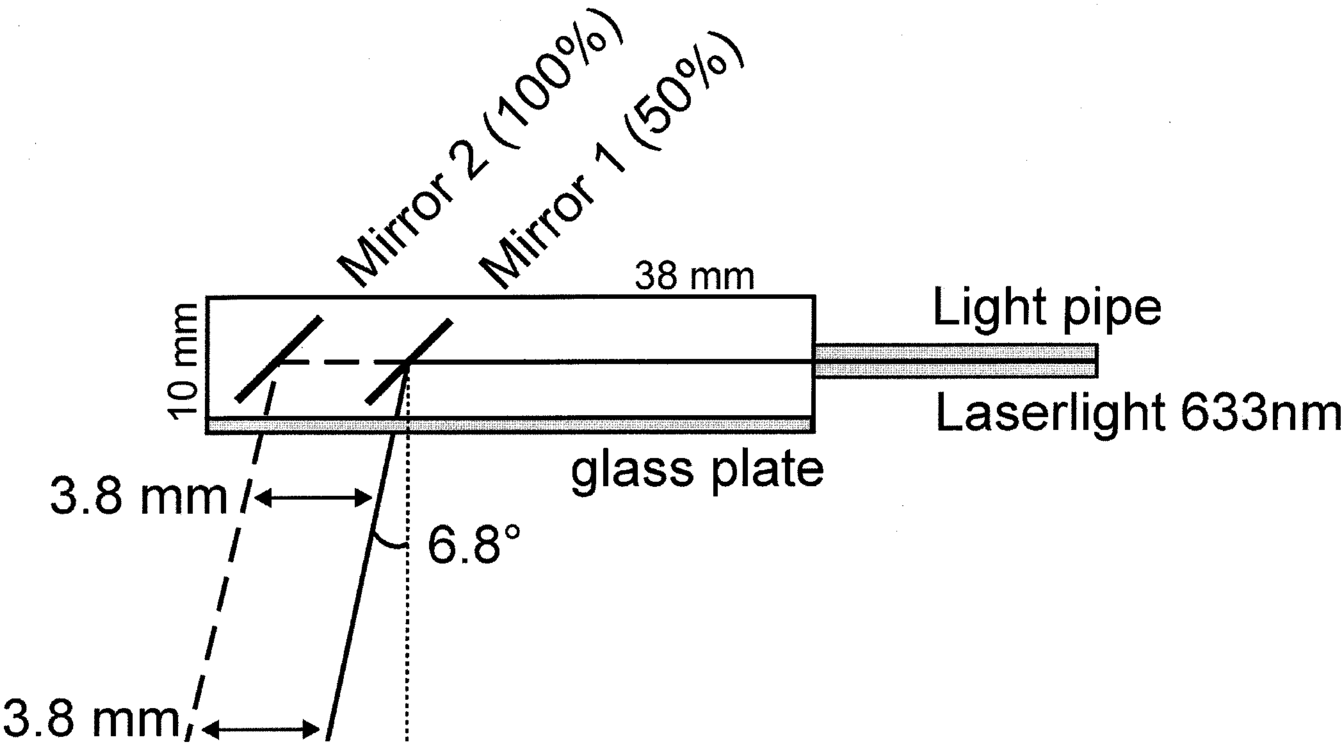

Fig. 2. A schematic image of the laser projection device shows its physical dimensions and the propagation of the laser light, which enters the device through a light pipe and is split into two beams by a mirror of 50% reflectivity. These parallel-adjusted laser beams leave the projection device under an angle of 6.8° and a distance of 3.8 mm.

during the glottal cycle at that moment when the glottal opening reached its maximum (i.e., when the glottal area was maximal). The glottal length tends to increase with increasing fundamental frequency. As shown in Figure 4, the frequencies are high in four cases. This is due to the fact that some subjects were slightly tensed during the investigation. However, the subjects had a complete healthy voice within a "normal" frequency range.

Amplitude and Velocity of Vocal Folds

The distance of the vocal fold edges from the glottal midline can be extracted from the high-speed recordings using the previously mentioned image processing algorithm. By applying this algorithm for each image, one can extract the vocal fold amplitudes as a function of time for each point of the vocal fold edge. Obviously, these trajectories depend on the exact position on the vocal fold. Usually, they exhibit maximum amplitudes at half the distance from anterior to posterior along the vocal fold edge. In the current study, only these trajectories for the mid position were evaluated.

Figure 5 shows the trajectories of the right-side vocal folds for seven subjects as a function of time over a time period of 25 milliseconds. Because the trajectories for the left and right vocal folds do not exhibit differences within the accuracy of measurement, only the right side is shown.

The trajectories differ primarily because of the different fundamental frequencies in the subjects (120–242 Hz). They allow direct determination of the opening and closing times and other oscillation parameters. Within these subjects, vibratory amplitudes of the vocal folds are approximately 1 mm and across all subjects quite similar. In one subject (subject G), no glottal closure occurs during phonation as indicated by the shifted trajectory. In another subject (F), the closure is not achieved in any glottal cycle.

In addition to the vocal fold amplitudes, the corresponding velocities are shown in Figure 5. They were calculated as the time-derivative of the trajectories. Velocities vary during the glottal cycle. Maximum velocities are approximately 1.4 meter/sec, and the average over all maximal velocities over the subjects is 1.0 meter/sec. The maximal velocities allow an estimation of the velocity at the instant of impact of the vocal folds. Although the velocities during the opening phase (positive velocities) are quite similar between these subjects, they differ substantially during the closing phase (negative velocities).

Precision

The precision of the system is mainly limited by three properties. First, the glottal plane might not be orthogonal to the optical axis. Second, the laser beams are not completely parallel adjusted. Third, the pixel resolution of the digital images is limited. In this section, the three error sources are discussed.

The compensation angle of 6.8° between laser beams and laser projection system (Fig. 2) was introduced to achieve an angle of 90° between the optical axis and the

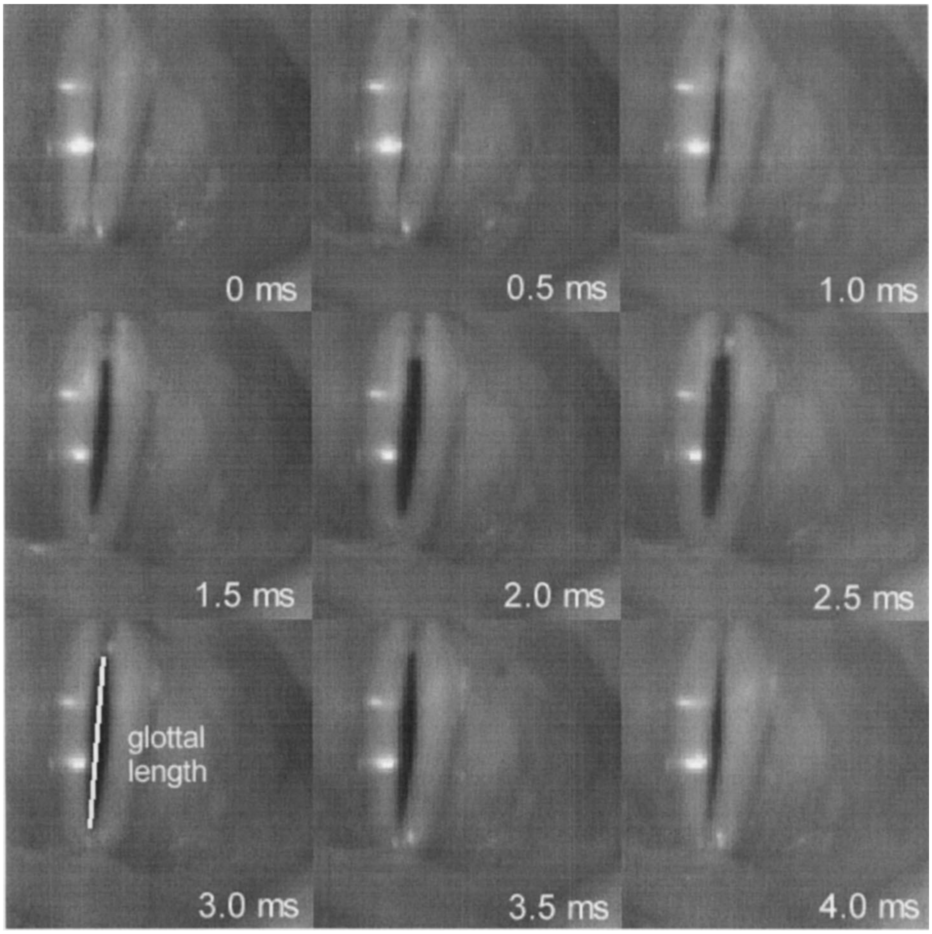

Fig. 3. A time sequence of nine images of a high-speed recording with the laser spots clearly visible on the subject's right vocal fold. The white line in the left lower corner at time index 3.0 ms marks the glottal length. This line corresponds to the parts of the vocal folds that do not remain in contact during phonation.

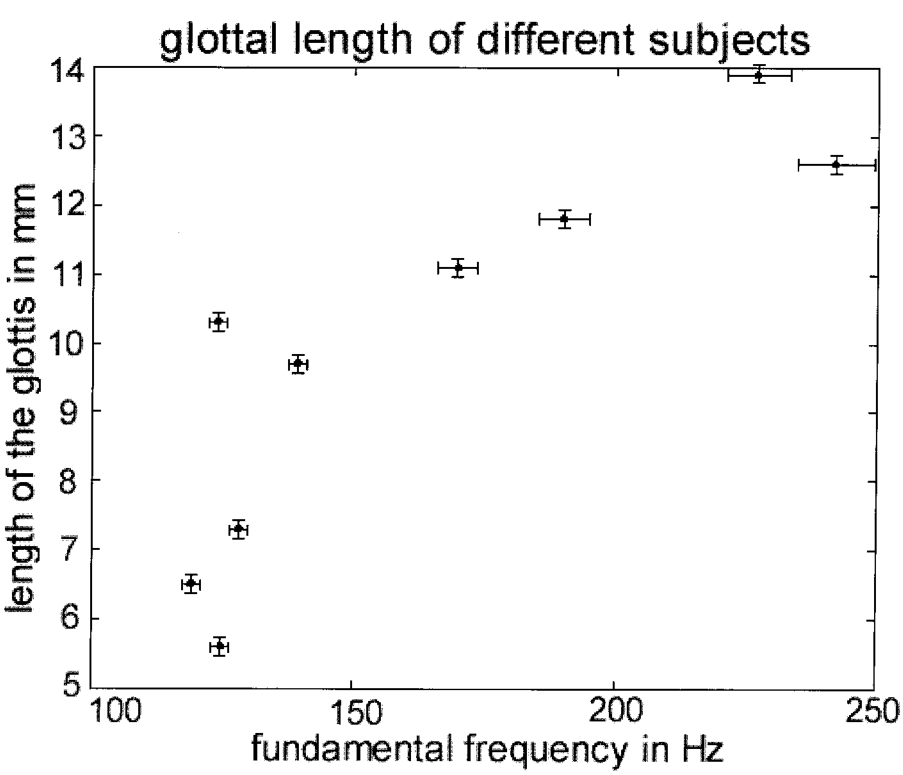

Fig. 4. The glottal length in millimeters as a function of the fundamental frequency in hertz in male subjects. The error bars indicate uncertainty of frequency attributable to the limited temporal sample rate and uncertainty of length of the glottis attributable to the limited spatial resolution of the CCD array in the high-speed camera.

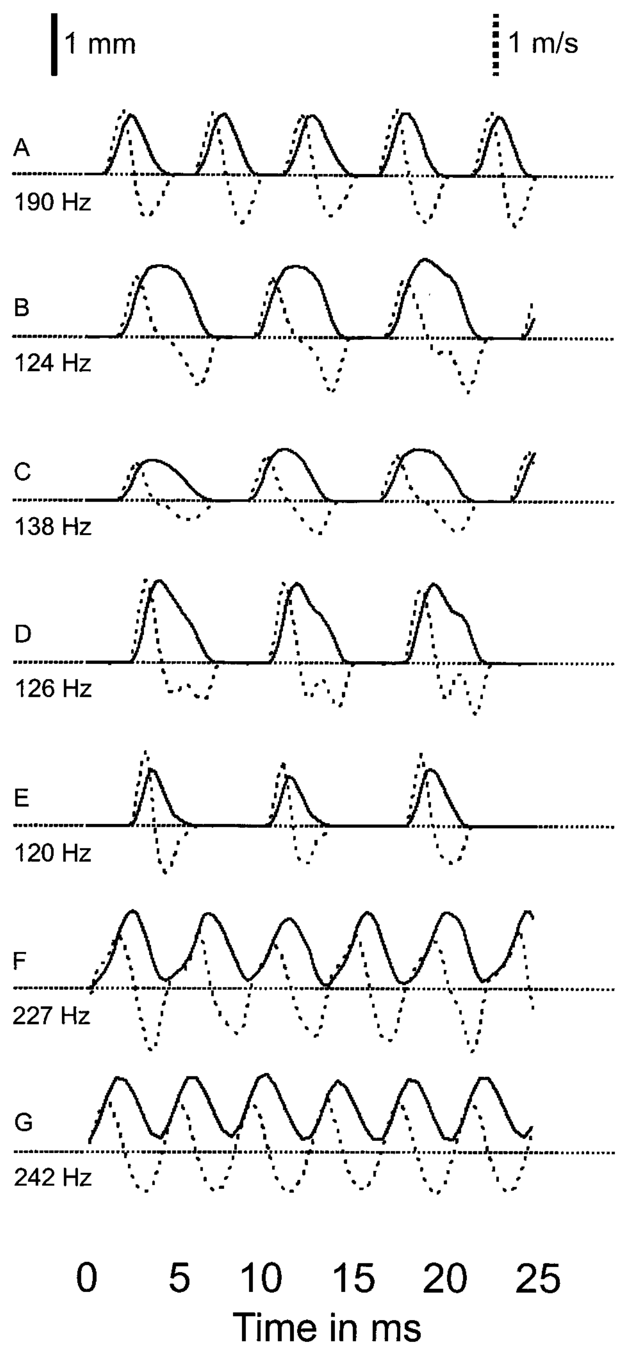

Fig. 5. The trajectories as a function of time (solid line) and velocities (dashed line) of the vocal folds at the glottal midline. At the top, the scales (millimeters and meter/sec) are shown. The group of velocities shows significantly more variation than does the group of trajectories. This suggests the use of trajectories and velocities for diagnosis.

glottal plane. Under standard conditions of examination the subject has to bend slightly forward in to grant optimal view on the larynx during endoscopy while the endoscope itself remains horizontal. No studies were found about this fact, but from our own observations we estimated an angle of 5° to 10°. The difference between the compensation angle of 6.8° and the actual angle during examination is called the tilt angle and produces an error in measurement as shown in Figure 6.

As shown in Figure 6, a tilt angle of less than 8° results in an error of less than 1% for the metric units. Additional errors might occur because of nonparallel laser beams. As mentioned earlier in the present study, this error is smaller than the beam divergence of 0.17°. Assuming an average distance of 8 cm from the endoscope to the glottal plane, this represents maximum error of approximately 0.6% (i.e., 0.22 mm) in the distance between the two laser spots.

An additional uncertainty occurs because of the limited pixel resolution of the image. The standard resolution of the image was 128×64 pixels, and the recorded area ranged from 8×8 to 15×15 mm. With this specification, an absolute error of 0.12 mm horizontally and 0.23 mm vertically arises. This uncertainty can be improved by better pixel resolution of the CCD array of the high-speed camera.

The total error can be roughly estimated as follows: The tilt angle is usually less than 8°. Hence, the corresponding error is less than 1%. The error from nonparallel beams is approximately 0.6 %. Therefore, the average relative error can be estimated to be 1.6%. For example, a 16-mm vocal fold would have an average absolute error of 0.26 mm. In the worst case, the spatial resolution contributes an error of 0.23 mm. Therefore, the apparatus allows measurements with a precision of better than 0.49 mm.

DISCUSSION

We presented an LPS that allows determination of metric units in endoscopic recordings. The LPS was combined with a high-speed apparatus to investigate vocal fold vibrations during phonation.

The robust design of the LPS proved reliable during the examinations for two reasons. First, mirrors cannot be displaced by accidental contacts with the tissue, and antifog agents can be used for reducing the misting of optical components during the examination. Because of the closed construction, it is not necessary to apply the agents directly on the mirrors but only on the glass plate. Second, the closed construction allows disinfection of the LPS with no risk for the optical elements.

The LPS was kept small. A small size is wanted for reasons of practicability. However, the size is crucial for both light efficiency and precision of the measurement. The distance between the two laser spots forms the scale of the image. It was chosen to be approximately 4 mm (3.8 mm) to be appropriate for both male and female subjects. This value is a compromise between the practicability of exact positioning of the laser spots onto the vocal folds during the phonation and the accuracy of the absolute measurement of the glottis.

Because the two laser beams are parallel, no calibration of the system is necessary. The distance of the laser spots represents the same length within each image for any focal length. In principle, the scaling of the images is possible with only one beam.[8] However, this method requires calibration for each image.

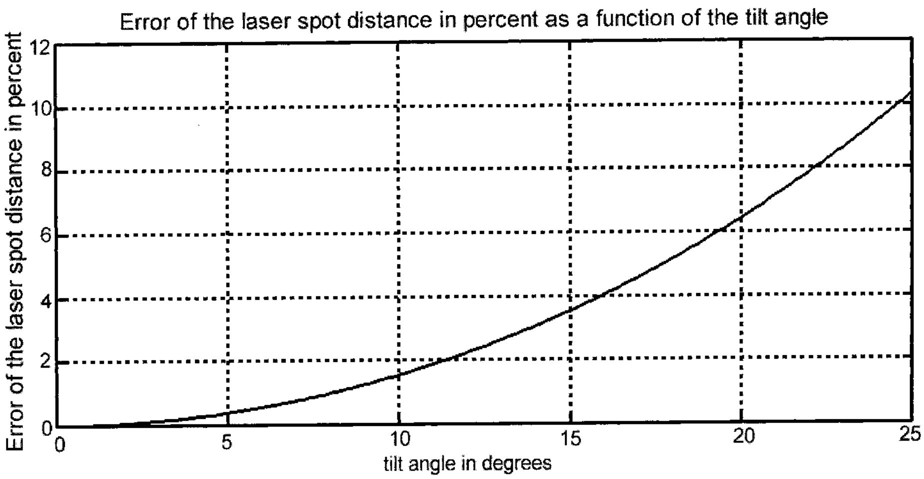

Fig. 6. The error of the distance of the two laser spots (percentage) as a function of the tilt angle, which is defined as the angle between the glottal plane and the laser beams.

The applicability of the LPS was demonstrated in investigations of endoscopic recordings of male subjects. The glottal length could be determined for each subject. The results agree with former measurements, in which the vocal fold length was determined by other x-ray[9–11] and stereo endoscopy[2,12,13] methods. However, the methods used in the former studies implied errors of several millimeters, which is 1.5 orders worse than the method presented in the current study.

By combining the LPS with endoscopic high-speed imaging, it is possible to measure the velocities of the vocal folds in units of meter/sec as shown in Figure 5. These velocities are the time derivative of the vocal folds' trajectories along an orthogonal direction to the glottal midline. The velocities may provide additional information for diagnosis of functional voice disorders. In addition, the collision forces between the vocal folds at the initial closure can be estimated from the velocities and elasticity parameters of the vocal fold tissue.

Hanson et al.[14] described similar measurements of vocal fold displacements and mucosal wave velocities. They used standard stroboscopic recordings extended by a two-point laser projection system similar to the one presented by us. They focused their evaluation on mucosal wave and calculated both amplitudes and velocities. However, the system presented by Hanson et al.[14] was based on stroboscopy and yielded reliable results only in cases of stationary vocal fold oscillations.

Because of the limited spatial resolution of the high-speed system, the identification of mucosal waves is difficult. Therefore, we focused on a direct measurement of vocal fold elongation and velocity. These data are an important precondition for comparisons with biomechanical models.[15] Detailed measurements including mucosal wave amplitude and velocity are part of a current project using an enhanced high-speed system.

Considering the different error sources that may contribute to the measurement, the maximum error was estimated to be below 0.49 mm. However, this value can be substantially reduced during examination. For example, the angle between the endoscope and the glottal plane can be reduced to 0.32 mm by varying the tilt angle in such a way that the distance between the two laser points within the endoscopic image is minimal.

The results demonstrate the great intersubject variability. This phenomenon may be explained by anatomical differences and by individual techniques of phonation. Systematic investigations in larger groups of subjects including female voices are in progress.

CONCLUSION

The presented LPS allows the determination of absolute measures in endoscopic laryngeal images. By combining the system with a digital high-speed camera, fully quantitative data during phonation can be obtained. Particularly, glottal and vocal fold lengths and vocal fold velocities can be determined precisely during respiration and all stages of phonation (e.g., phonation onset). Quantitative data are useful for the comparison of different endoscopic investigations. Therefore, the presented method can be used for voice diagnosis and for monitoring during voice therapy.

BIBLIOGRAPHY

1. Mergell P, Herzel HP, Titze IR. Irregular vocal-fold vibration: high-speed observations and modeling. *J Acoust Soc Am* 2000;108:2996–3002.
2. Wittenberg T, Spinnler K, Tigges M, Schuberth S, Schmidt R. Stereo-Hochgeschwindigkeitsaufnahme zur dreidimensionalen Vermessung und Berechnung von Stimmlippenschwingungen. In: Kruse E, ed. *Deutsche Gesellschaft für Phoniatrie und Pädaudiologie, 17. Wissenschaftliche Jahrestagung.* Tübingen: Deutsche Gesellschaft für Phoniatrie und Pädaudiologie, 2000:51.
3. Eysholdt U, Tigges M, Wittenberg T, Pröschel U. Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia Phoniatr Logop* 1996;48:163–170.
4. Wittenberg T, Moser M, Tigges M, Eysholdt U. Recording, processing and analysis of digital high-speed sequences in glottography. *Machine Vision Applications* 1995;8:399–404.
5. Hirano M, Kurita S, Nakashima T. The structure of the vocal folds. In: Stevens K, Hirano M, eds. *Vocal Fold Physiology.* Tokyo: University of Tokyo Press, 1981:33–34.
6. Wittenberg T. *Wissensbasierte Bewegungsanalyse von Stimmlippenschwingungen anhand digitaler Hochgeschwindigkeitsaufnahmen.* Aachen: Shaker Verlag, 1998:104–123.

7. Hoppe U. *Mechanisms of Hoarseness: Visualisation and Interpretation by Means of Nonlinear Dynamics*. Aachen: Shaker Verlag, 2001.
8. Larsson H, Hertegård S, Mannerberg G. Calibration of high speed imaging by laser triangulation. In: Schutte H, ed. *Fifth International Conference: Advances in Quantitative Laryngology* [conference on CD-ROM]. Groningen: Groningen Voice Research Laboratory, 2001.
9. Hollien H. Some laryngeal correlates of vocal pitch. *J Speech Hear Res* 1960;3:361–371.
10. Sonninen A. Is the length of the vocal cords the same at all different levels of singing? *Acta Otolaryngol Suppl* 1954;188:219–231.
11. Godfrey E, Arnold MD. Physiology and pathology of the cricothyroid muscle. *Laryngoscope* 1961;71:687–753.
12. Nishizawa N. Stereoendoscopic observation of the larynx vocal fold length in respiration and in phonation with vocal pitch change. *J Otolaryngol Jpn* 1989;92:1244–1252.
13. Sawashima M, Miaazaki S. Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberscopes. *Ann Bull Res Inst Log Phon* 1974;8:7–12.
14. Hanson DG, D'Agostino MM, Jiang J, Herzon G. Clinical measurement of mucosal wave velocity using simultaneous photoglottography and laryngostroboscopy. *Ann Otol Rhinol Laryngol* 1995;104:340–349.
15. Doellinger M, Hoppe U, Hettlich F, Lohscheller J, Schuberth S, Eysholdt U. Vocal fold parameter extraction using the two-mass-model. *IEEE Trans Biomed Eng* 2002 [*submitted*].