Measurement of human vocal fold vibrations with laser triangulation

Göran Manneberg
Royal Institute of Technology
Dept of Physics, Optics
S-100 44 Stockholm, Sweden
E-mail: manneberg@optics.kth.se

Stellan Hertegård
Karolinska Hospital
Phoniatric Department
S-171 76 Stockholm, Sweden

Johan Liljencrantz
Royal Institute of Technology
Centre for Speech Technology
S-100 44 Stockholm, Sweden

Abstract. Laser triangulation is used to examine the vertical (longitudinal) movements of the vocal folds during phonation. A method for evaluating the blurry, elongated image of the laser spot by vibration amplitude is presented. The method is used *in situ* on humans and is calibrated with known vibrations of bovine tissue. © *2001 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.1396324]

Subject terms: medical optics; laser metrology; laser triangulation; vibration measurements; phoniatrics.

Paper 200351 received Sep. 5, 2000; revised manuscript received Mar. 29, 2001; accepted for publication Mar. 30, 2001.

1 Introduction

Direct optical inspection of the human vocal folds has been done for a very long time with the help of direct image relaying endoscopes,[1] and more recently also by connecting CCD cameras to endoscopes that are inserted into the pharynx. This equipment used in combination with stroboscopic light sources has facilitated the study of the transverse (horizontal) vocal fold movements during phonation. It has long been realized that vertical (parallel to the stream of air-longitudinal) movements (the so-called mucosal wave) also takes place. These vertical movements of the vocal fold mucosa facilitate vibrations during voice production.[1] Decreased mucosal wave amplitude is an important sign during medical examinations of hoarse patients due to inflammation in the larynx or due to mass lesions (e.g., nodules, polyps, or cancer) of the vocal folds as well as for cases with vocal fold paresis.[2] This is widely used during diagnostic procedures and also to monitor effects of various treatments.

It is, however, very difficult to quantify the vertical vocal fold movements during examinations. The reasons are the inaccessibility of the larynx and also the variations in distance from the endoscope lens to the vocal fold surface during examinations on awake human subjects. Previous examinations by us and others have shown that the amplitude of the movements are in the magnitude of 0.5 to 1 mm, which requires small measurement errors for any method for quantification.[3,4] We are not aware of any existing method for vertical amplitude measurements with a measurement error of <0.1 mm.

In this paper, we report the use of laser triangulation (see any book on metrology, e.g., Ref. 5) for *in vitro* mapping of the vertical movement of the vocal folds. Laser triangulation is normally used in conjunction with light centroid measuring devices (PSDs) and with CCDs, where the centroid position is found through intensity profile fitting to an assumed Gaussian shape of the illuminated spot.

In our case, we measured the vibration amplitude of the vocal folds by integrating the light distribution over time, giving an oval shape of the spot. The evaluation of the shape of this spot gives us the information concerning the vibration amplitude.

2 Principle of Measurement

The principle of measurement is, of course, very simple. The laser beam illuminates the vocal folds under an angle q measured to the symmetry axis of the observing optics. The angle and position of the beam are so chosen that $a=0$ at the medium observing distance, and is visible in the image for all relevant distances z.

The size and focal length of the camera is chosen as small as possible to minimize defocusing problems. Typical values in our setup are $z=50$ to 100 mm and $f=20$ mm. Note also that the optical system symbolized by a lens in Fig. 1 is a relaying optics system in the endoscope, and all distances refer to the relevant principal planes.

From equilateral triangles it is then obvious that $a/z = \delta/f$. Trigonometry further shows that $a+d=z\tan\theta$. Eliminating, a gives us

$$z=\frac{df}{f\tan\theta-\delta}, \tag{1}$$

which is an expression of the type

$$z=\frac{A}{B+\delta}, \tag{2}$$

meaning that A and B can easily be determined with great accuracy as part of a calibration process.

By differentiating we can further see the resolvable change in $z(=\Delta z)$ depends on the geometry variables as

Principle of measurement

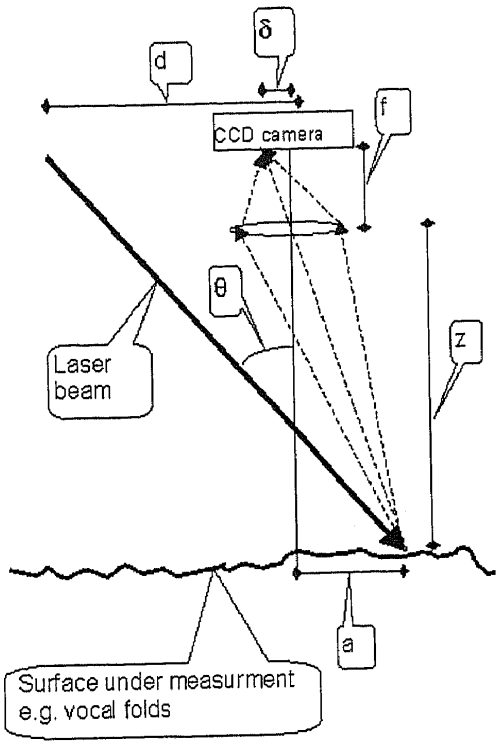

Fig. 1 Experimental setup.

$$\Delta z = \frac{df}{(f \tan\theta - \delta)} \Delta\delta = \frac{z^2}{df} \Delta\delta, \tag{3}$$

where $\Delta\delta$ is the smallest resolvable shape change, and can be first approximated by the pixel size. In this expression, z is rather given by the geometry of the throat and f is restricted by the focus depth condition. The only parameters available to increase resolution are the pixel size ($\Delta\delta$) and the triangle base d.

An increase of d can be done only at the detriment of the patients' comfort as it requires accepting a wider instrument in the throat. With a pixel size of 5 μm and $z = 50$ mm, $d = 18$ mm, and $f = 20$ mm, we have a resolution of approximately 30 μm.

The size of the spot is typically 300×600 μm on the tissue, giving a spot on the CCD of tens of pixels in both directions. Hence discretization of the detector should not pose a problem if the pixel data is treated correctly. To do that, several steps must be taken.

First there is a large amount of stray light due to the diffusely scattering nature of the tissue in the cavity. Thus the stray light level must be determined from areas away from the laser spot and then must be subtracted from the pixel values at the spot.

Second, there are glistening specular reflections on the surfaces of liquid and mucosal overlays on the diffusely reflecting surface of interest. These are controlled (at least partially) by identifying the peak corresponding to the laser spot and then masking everything outside a circle corresponding to the largest possible spot.

Then the center of the illuminated spot is determined to obtain the average position. The amplitude of the vibration is then determined from the shape of the spot at the intersection between the image plane and the plane of incidence of the laser beam. The beam waist ω_{0x} is carefully measured in that direction, and care is taken in the beam delivery system to maintain correct orientation of the (astigmatic) semiconductor laser beam.

The movement of the laser spot will yield an intensity profile in the time-integrated image, which is the convolution between the Gaussian profile of the spot and the time increment (Δt) spent at every position increment (Δx). If the vibration were sawtooth-like, this would be the familiar convolution with a recut well known from photographing objects moving with constant speed.[6]

In our case, we assume a sinusoidal vibration of the tissue, giving (in relevant approximation) a sinusoidal vibration of the laser spot. This means that time as function of position during one half vibration cycle is given by

$$t = \frac{1}{\omega} \arcsin\left(\frac{x}{x_0}\right), \tag{4}$$

where x_0 is the vibration amplitude. Thus, the convolution function mentioned earlier will be

$$\frac{\Delta t}{\Delta x} \cong \frac{dt}{dx} = \frac{1}{\omega} \frac{1}{\sqrt{(x_0^2 - x^2)^{1/2}}}. \tag{5}$$

The expected intensity profile will hence be given by

$$I(x) = \frac{I_0}{\omega} \left\{ [\exp(-x^2/\omega_{0x}^2)] \otimes \left[\frac{1}{\sqrt{(x_0^2 - x^2)^{1/2}}} \right] \right\}. \tag{6}$$

This function is plotted for $x_0 = \omega_{0x}$, $2\omega_{0x}$, and $3\omega_{0x}$, respectively, in Fig. 2.

The measured intensity profile is then matched to the profile corresponding most closely to the actual situation with x_0 and l_0 being the free parameters. The latter is, of course, totally uninteresting. This gives us the vibration amplitude in the CCD plane, which is then transformed to a real vibration in the object plane with Eq. (3) and the z value obtained earlier.

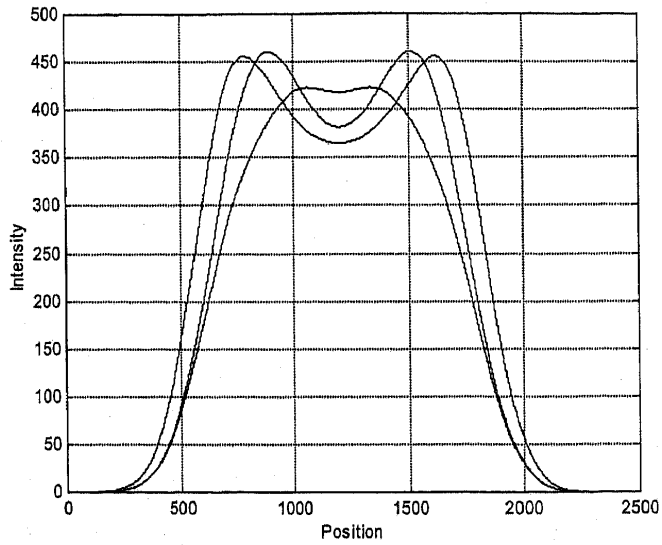

Fig. 2 Convolution function for three different amplitudes.

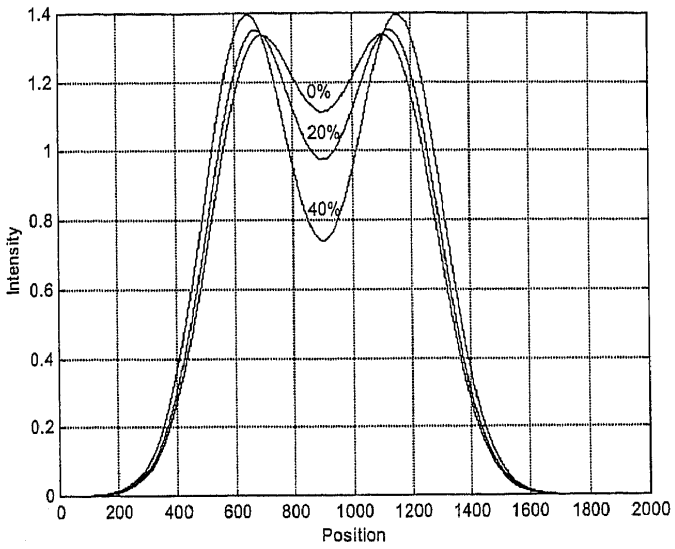

Fig. 3 Effect of superimposed harmonics on the convolution function.

The question can also be raised whether the assumption of sinusoidal motion of the vocal folds distorts the results. To check that, we numerically simulated the curve corresponding to the widest of those in Fig. 2 with an increasing amount of the third harmonic in the movement. We checked it for 0, 20, and 40 percent, and the result is that it distorts the results according to Fig. 3. Applying the new curves to the same curve fitting as for the curves in Fig. 2 shows the same result in vibration amplitude (within ±10 percent), but with worse fitting.

3 Calibration and Validation of the Method

The calibration of the instrument is very straightforward as least square fitting of the parameters A and B in Eq. (2).

Validation of the method *in situ* is, as far as we understand, impossible as present at there is no other method for (simultaneous) vibration measurement. As a consequence, we must create circumstances that are as "human-throat-like" as possible in an environment where controlled vibrations can be created. We chose to mount a slice of bovine soft tissue (fascia) on a loudspeaker with well-defined vibration amplitudes (Fig. 4). The slice was trimmed to a thickness of approximately 0.5 mm (which is similar to the thickness of the superficial vocal fold mucosal layer that vibrates during phonation. This tissue (of 0.5×0.5-cm size)

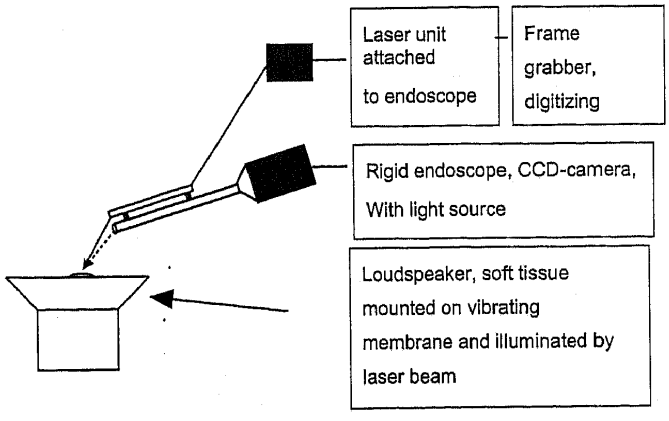

Fig. 4 Experimental setup for validation of the method.

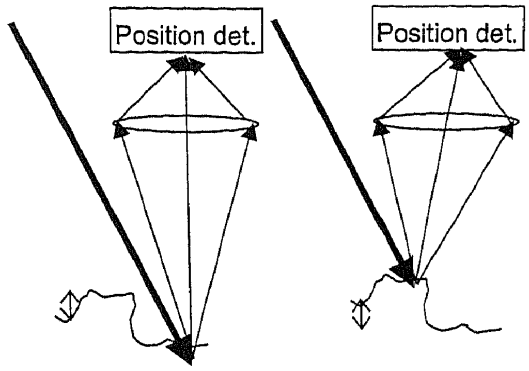

Fig. 5 Rough surface problem.

was kept moist with saline and was adhered to the central part of the loudspeaker membrane with its own moisture. The vibration amplitude was measured with a custom-made laser unit consisting of a semiconductor laser (643 nm) attached to the distal end of a rigid endoscope (Hopkins 70 deg, Karl Storz, Germany). The halogen light source of a Brüel&Kjær (4914) stroboscope was used as the illumination source. The endoscope was connected to a Panasonic (KS 152) CCD video camera. Recordings were made on a S-VHS Panasonic (Model 7350) video. The video recordings were fed to a PC with a Matrox Inspector (1.7) frame grabber (Matrox Electronic Systems Ltd.) and images were saved in bitmap format for later digitized image analysis.

The root mean square (rms) error in the image plane was seen to be almost constantly equal to approximately 4 to 5 μm, which can then be transformed to a corresponding error in the object plane using Eq. (3) expressed in the least-squares determined coefficients A and B:

$$\Delta z = \frac{A}{(B+\delta)^2}\Delta\delta = \frac{z^2}{A}\Delta\delta. \tag{7}$$

The typical error in amplitude is 30 μm, giving double that for the hill-to-crest movement.

Another potential source of error is bulk scattering of the laser light in the tissue of the vocal folds. A calculation of the effect of this can only be made with measurements on exactly that sort of tissue, and so it was not done in this work. The human vocal folds are, however, rather opaque and white, and so do not scatter the light extensively. An indication that this is not a serious problem comes from the fact that the beam waist in the direction perpendicular to the beam vibration plane measured in the image is the same as that measured directly in the beam with a razor edge test.

4 Relevance of the Measurement

It has to be noted that the measured vibration amplitude is the amplitude of movement of the illuminated spot. If the illuminated surface in reasonably flat on a scale comparable to the spot size of the laser beam, this will reflect the actual movement of the surface that invokes the longitudinal movement of the air column. Roughness with a wavelength smaller than the size of the spot will be averaged out and roughness much coarser than the spot will not contribute to any error as the movement of the spot will cover only a fraction of a wavelength.

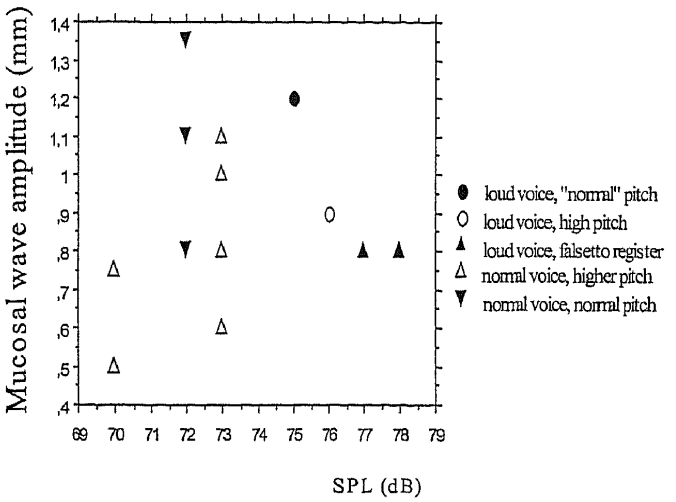

Fig. 6 Mucosal wave amplitude (vertical vocal fold vibration amplitude) for a normal human subject during phonation at various pitch and intensity levels.

If on the other hand, the spot happens to illuminate a region where the topography resembles that of Fig. 5, a large error in the measurement will occur. The vibration will be exaggerated several orders of magnitude. The human vocal folds are rather flat, but when there are pathological changes it is vital for the measurement to measure at several locations at the same time and then exclude measurements that are incoherent with the rest. Furthermore if the laser beam is directed to the medial border of a vibrating vocal fold, different parts of the mucosa will be hit during a vibratory cycle (which is typically 10 ms for male and 5 ms for female phonation). This is caused by the combined horizontal and vertical mucosal movements at this part of the vocal fold.

Note also that the amplitude measured is the component in the direction of the line of sight. Transverse vibrations are not included in this study, but have long been investigated, as they can be seen with direct inspection with any endoscope. Finally, also note that only the amplitude, and not the waveform, can be measured with this scheme.

5 Examples of Measurement from Human Subjects

Figure 6 shows vertical vocal fold amplitude measurements (mucosal wave amplitude) for a normal subject phonating at different intensity levels and at different pitch levels. Amplitude (top to bottom is around 1 mm). Roughly, the amplitude increases with intensity [Sound Pressure Level (SPL)] except for high-pitched falsetto phonation. Thus the SPL and pitch must be monitored when performing measurements in order to be able to compare successive measurements.

6 Conclusions

We demonstrated a method for examining the amplitude of the vibration of the human vocal fold during phonation. The errors in the measurement lies entirely below that of the pixel resolution in laboratory tests on a loudspeaker with adhered biological tissue. Systematic errors can occur if the tissue under measurement is not opaque enough, but for human vocal folds that does not seem to be a problem.

For a discussion of the medical relevance of the result please see the papers in the medical press.[3]

A natural extension of this work is to project a number of similar spots in a given pattern to facilitate a simultaneous continuous mapping of the entire surface of the fold.

References

1. I. R. Titze, "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.* **83**(4), 1536–1552 (1988).
2. P. Kitzing, "Stroboscopy—a pertinent laryngeal examination," *J. Otolaryngol.* **14**(3), 151–157 (1985).
3. S. Hertegård, G. Björck, and G. Manneberg, "Using laser triangulation to measure vertical distance and displacement of laryngeal mucosa," *Phonoscope* **1**(3), 179–185 (1988).
4. Y. Kakita, M. Hirano, H. Kawasaki, and K. Matsuo, "Stereolaryngoscopy: a new method to extract vertical movement of the vocal fold during vibration" in *Biomechanics, Acoustics, and Phonatory Control*, I. R. Titze and R. C. Scherer, Eds., pp. 191–201, Denver Center for the Performing Arts, Denver (1983).
5. K. J. Gasvik, *Optical Metrology*, Wiley, New York (1995).
6. Goodman, *Fourier Optics*, 2nd ed, Chapter 8, pp. 217–295, New York (1996).