Available online at www.sciencedirect.com

ScienceDirect

Medical Image Analysis 12 (2008) 300–317

MEDICAL IMAGE ANALYSIS

www.elsevier.com/locate/media

Calibration of laryngeal endoscopic high-speed image sequences by an automated detection of parallel laser line projections

Tobias Wurzbacher [a,*], Ingmar Voigt [b], Raphael Schwarz [b], Michael Döllinger [b], Ulrich Hoppe [c], Jochen Penne [d], Ulrich Eysholdt [b], Jörg Lohscheller [b]

[a] *Siemens Audiologische Technik GmbH, RD Signal Processing, Gebbertstrasse 125, 91058 Erlangen, Germany*
[b] *University Hospital Erlangen, Medical School, Department of Phoniatrics and Pediatric Audiology, Bohlenplatz 21, 91054 Erlangen, Germany*
[c] *University Hospital Erlangen, Medical School, Department of Audiology, Waldstraße 1, 91054 Erlangen, Germany*
[d] *Friedrich-Alexander University Erlangen-Nuremburg, Computer Science Department 5, Pattern Recognition, Martensstraße 3, 91058 Erlangen, Germany*

Received 27 February 2007; received in revised form 16 October 2007; accepted 15 December 2007
Available online 12 February 2008

Abstract

High-speed laryngeal endoscopic systems record vocal fold vibrations during phonation in real-time. For a quantitative analysis of vocal fold dynamics a metrical scale is required to get absolute laryngeal dimensions of the recorded image sequence. For the clinical use there is no automated and stable calibration procedure up to now. A calibration method is presented that consists of a laser projection device and the corresponding image processing for the automated detection of the laser calibration marks. The laser projection device is clipped to the endoscope and projects two parallel laser lines with a known distance to each other as calibration information onto the vocal folds. Image processing methods automatically identify the pixels belonging to the projected laser lines in the image data. The line detection bases on a Radon transform approach and is a two-stage process, which successively uses temporal and spatial characteristics of the projected laser lines in the high-speed image sequence. The robustness and the applicability are demonstrated with clinical endoscopic image sequences. The combination of the laser projection device and the image processing enables the calibration of laryngeal endoscopic images within the vocal fold plane and thus provides quantitative metrical data of vocal fold dynamics.

Keywords: Endoscopic high-speed imaging; Laser line detection; Calibration

1. Introduction

Vocal fold vibrations are the sound source needed for oral communication and are in normal voice quasi-periodic and symmetric, whereas hoarseness arises from irregularity and asymmetry (Eysholdt et al., 2003). In clinical routine, vocal fold vibrations are examined by laryngeal endoscopic techniques. A digital high-speed (HS) camera is coupled to the endoscope and enables real-time recordings of vocal fold oscillations during phonation (Wittenberg et al., 1995). The HS recording system generates image sequences and stores them into a PC for later visualization, analysis, and documentation.

The dynamics, especially the deflections, of the vibrating vocal fold edges are of interest for research and clinical purposes. Saadah et al. (1998), Yan et al. (2006), and Lohscheller et al. (2007) developed methods to track and segment the glottis in laryngoscopic image sequences. Fig. 1a shows the segmented vocal fold edges, which encloses the glottal area. The glottal axis is defined as the linear regression of this area. Vocal fold deflections are calculated as the orthogonal distance from the glottal axis to the segmented vocal fold edges. The resultant movements are illustrated in Fig. 1b. As demonstrated by Neubauer et al. (2001) asymmetries in vocal fold vibrations are not restricted to left/right asymmetries, but also longitudinal

[*] Corresponding author. Tel.: +49 9131 3083232; fax: +49 9131 3083406.
E-mail address: Tobias.Wurzbacher@siemens.com (T. Wurzbacher).

1361-8415/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2007.12.007

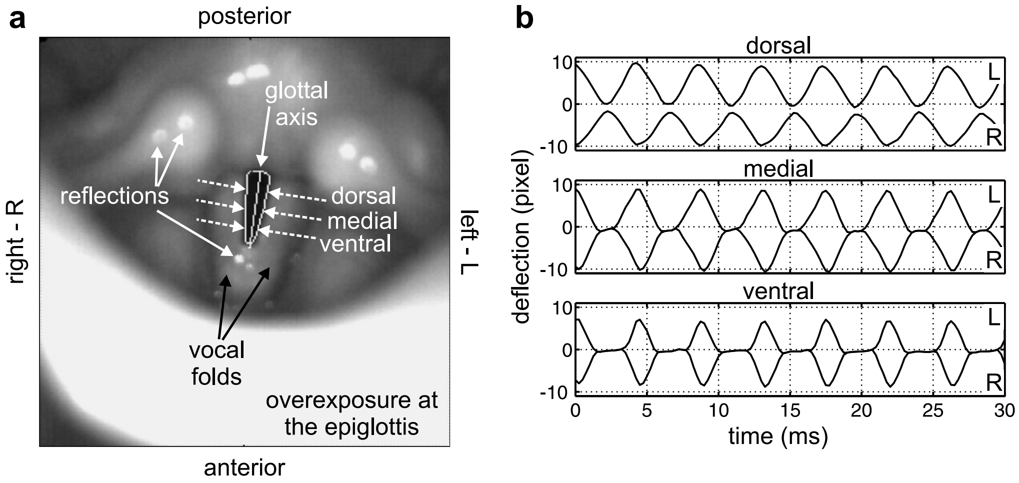

Fig. 1. (a) The image shows a frame of an endoscopic HS recording of the larynx. The vocal fold edges are segmented and the glottal axis is depicted. (b) For vocal fold vibration analysis the deflections of the vocal fold edges are evaluated at a dorsal, medial, and ventral position.

asymmetries in dorsal/ventral direction occur. These asymmetries can be evaluated by extracting the deflections of the vocal fold edges at a dorsal, medial, and ventral glottal third.

As the image data is delivered from the sensor of the HS camera in pixel units, only relative amplitude and velocity changes of vocal fold oscillations are measurable within a recording. However, the knowledge of absolute vocal fold velocity, respectively acceleration, is a clinically relevant measure, since it is suspected to be a possible origin for vocal fold nodules and polyps (Titze, 1994; Jiang et al., 2006). Quantitative analysis of the extracted vocal fold movements needs metrical calibration information in endoscopic images.

Laser devices can be used to project calibration patterns onto the vocal folds. Two types of calibration patterns are presented within the literature. The first method uses a single laser beam that produces a single spot (Hertegård et al., 1998; Manneberg et al., 2001; Larsson and Hertegård, 2004). However, before the one spot projector can be applied to laryngoscopic measurements, reference measurements outside the larynx have to be performed for gaging the system. The second method uses two parallel beams, which project laser spots with a known distance onto the vocal folds (Herzon and Zealear, 1997; Schuberth et al., 2002; Schade et al., 2004). It enables to scale the image data without any additional gaging measurements.

A two-spot laser projection system has been clinically applied by Hoppe et al. (2003) and Schuster et al. (2005). The major problem in calibrating HS image sequences is to identify the projected laser spots within the images. An automated finding of laser spots by image processing means may occasionally fail (Larsson and Hertegård, 2004) due to typical laryngeal endoscopic image artifacts (Fig. 1), as specular reflections of the mucosa, the overexposure induced by the endoscopic primary light source, or the inhomogeneous illumination of the larynx, which may hide the projected spots. The similarity in shape and intensity between the laser spots and the specular reflections within the image data leads to ambiguities. Consequently, an automated detection is difficult and time-consuming intervention by hand is necessary.

In order to make calibration by laser devices applicable the laser projections should be easy to identify in the image. Therefore, a laser calibration pattern should be more silhouetted against typical endoscopic image artifacts. In this paper, we propose two parallel laser lines as a projection pattern and a robust image processing algorithm to automatically detect and segment the laser lines within laryngoscopic image sequences. The line detection bases on a Radon transform (Leavers, 1992) approach and on a Hough transform (Illingworth and Kittler, 1988; Leavers, 1993), which is a special case of the former one. Both transforms map two dimensional images to a parameter domain $D(\rho,\theta)$, where lines can be identified as peaks (Bracewell, 1995; Toft, 1996a). Here, ρ is the distance of the straight line to the origin and θ denotes the angle of the corresponding normal vector. The application of the Radon/Hough transform is widely-used in medical image processing and ranges from needle detection (Okazawa et al., 2006) and calibration measurements (Rousseau et al., 2005) in ultrasound imaging over the segmentation of the optic nerve head in scanning laser tomography images of the retina (Chrastek et al., 2005) to surface and density reconstructions in tomography (Whitaker and Elangovan, 2002). Both transforms are robust against broken line segments, uncorrelated noise, and correlated noise, as long as the correlation length of the background noise is less than the line length (Beyerer and León, 2002).

The stability of the laser line detection depends on the line length, which in laryngoscopy is determined by the vocal fold width. In a single HS recording image (Fig. 1) the vocal fold width is only a fraction of the image width and thus the laser lines are short within a single image frame. An elongated representation of the laser lines is obtained by processing sufficient subsequent frames of the HS recording together and initially performing the line detection along the time axis of the recorded image sequence. The applicability of the laser projection device and the robustness of the laser line detection algorithm for calibration purpose are demonstrated with clinical HS endoscopic recordings of the larynx.

2. Laser calibration hardware

2.1. Laser line projection system (LLPS)

The projection unit of the LLPS is housed in a case of 55 × 7.5 × 9 mm and is clipped to a 90° rigid endoscope of 9 mm external diameter as depicted in Fig. 2a and b. The projection unit is connected via a light pipe to the laser source. The source radiates laser light of the wavelength 635 nm (GaAs) with a measured power output of 22 mW. The principle of the laser line projection is illustrated in Fig. 2c. A glass cube with 50% reflectivity splits up the laser into two beams in the projection unit. The first beam is diverted by the glass cube. The second beam is diverted by a prism behind the glass cube. The laser beams are focused for a projection height $h \approx 70$ mm and are radiated in an angle of 83.2° from the projection unit. According to Schuberth et al. (2002) this angle reduces the tilt between the planes of the projection unit and the vocal folds, which results in a measurement error less than 1% due to non perpendicular projection. Both laser beams propagate through a diffraction plate at the bottom of the projection unit, which generates a laser line out of each beam by constructive interference. The metrical interline distance d_m has to be smaller than the vocal fold length. It is $d_\mathrm{m} = 5.4$ mm, which is 0.7 times less than the minimal vocal fold length in female subjects (Schuster et al., 2005).

2.2. LLPS accuracy

For calibration purpose the interline distance d_m has to be constant for the laser lines independent of the projection height h. The LLPS accuracy depends on the beam divergence of the laser lines and the spatial resolution of the image sensor. Measuring the LLPS accuracy serves to validate the parallelism specifications of the projection hardware and determines, which precision can be achieved with the LLPS in principle for high-resolution images.

For accuracy measurements, the laser lines are projected on a planar surface from different heights h between 49 mm and 87 mm around the assumed working height of 70 mm. Photographs are taken with a high-resolution camera with six million pixels that is mounted on a tripod. A workpiece of 41.5 × 27.5 mm manufactured with a precision of 10 μm is put beside the projected laser lines as a reference scale as illustrated in Fig. 2d. The scaling information of the laser line calibration is compared with the reference scale. Two scaling factors are calculated from each photograph: First, the side-length in pixels of the high-precision workpiece is manually determined. This is repeated ten times to average out measurement errors. The resulting scaling factor γ_r is used as reference. Second, the scaling factor γ of the laser lines is determined. For this, the distance in pixels d_p between the two laser lines is calculated by taking the Radon transform. The line distance d_p is given by the

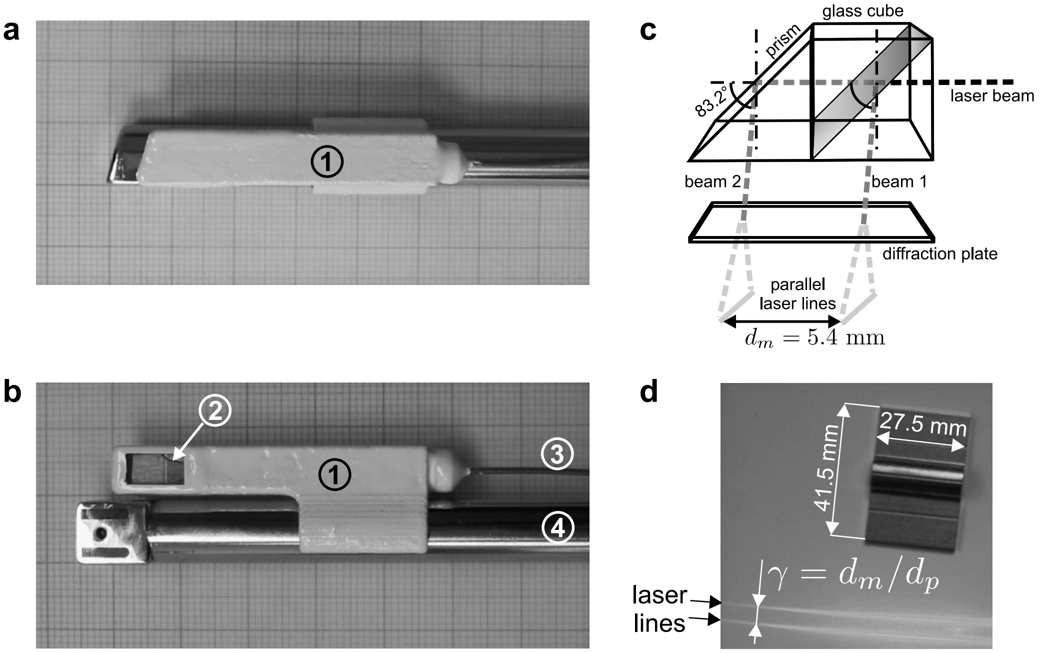

Fig. 2. Photograph of the LLPS mounted on a rigid endoscope in (a) side view and (b) in view from below. ① Casing of the projection unit. ② Diffraction plate. ③ Laser light pipe. ④ The endoscope has a diameter of 9 mm and a 90° optic. The graph paper in the background has a 1 × 1 mm scale. Subfigure (c) illustrates the principle of the laser line projection. Subfigure (d) shows an image of the LLPS accuracy measurements with the calibration workpiece and the laser lines.

two highest peaks in the Radon space $D(\rho,\theta)$, from which the laser scaling factor γ is derived. Additionally, the Radon space $D(\rho,\theta)$ provides information about the line parallelism within the projection plane. It is measured as the difference $\Delta\theta$ of the angles that represent the two dominant peaks.

3. HS image sequences

HS image sequences of the subjects' vocal fold vibrations are acquired with the HS ENDOCAM 5560 (Richard Wolf Corp. Knittlingen, Germany) with a spatial resolution of 256 × 256 pixels and a temporal resolution of $f_s = 4000$ Hz. The LLPS is equipped to the HS ENDOCAM and projects laser lines across the vocal folds. The unprocessed image series is denoted as $\widetilde{I}(x,y,t)$. Simultaneously to the HS image series, the acoustical signal and the sound pressure level is recorded. The microphone signal is sampled with 44.1 kHz and quantized with 8 bit. An overview of the HS recording technique is given in Eysholdt et al. (2003).

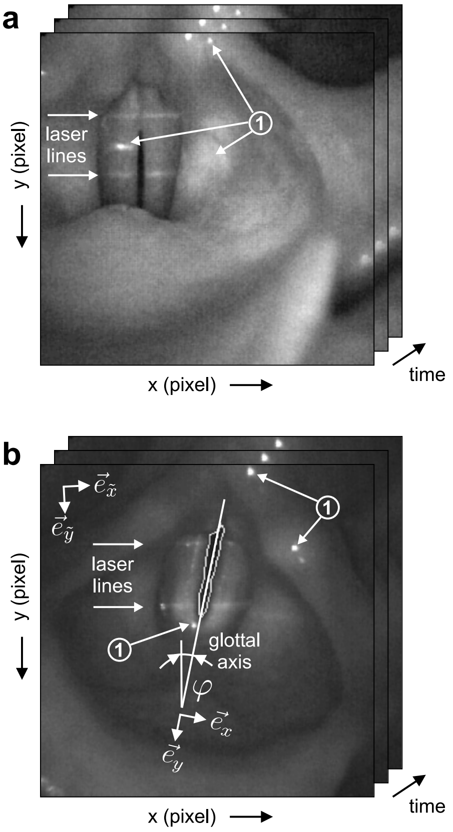

Fig. 3. Two image frames $\widetilde{I}_n(x,y)$ with the parallel laser lines on the vocal fold tissue are depicted for two different HS recordings (a) and (b). The recording quality depends on the light reflections of the primary light source needed for endoscopy. The mark ① denotes spot-like and diffuse reflections inside and outside the vocal fold tissue. In (b) the segmented vocal fold edges are illustrated. The glottal axis has an angle φ to the y-axis of the image.

Fig. 3a and b depict a single image frame $\widetilde{I}_n(x,y) := \widetilde{I}(x,y,t_n)$ of two HS recordings of the larynx. Small spot-like reflections are found across the images. The laser lines with a typically width of $w = 5$ pixels can be distinguished from the background by higher pixel intensities. In the average, the laser line intensities have a gray level scale of 38 ± 11, which is ~6 units greater than the vocal fold tissue in the neighborhood of the laser lines. The subglottal walls of the trachea are not illuminated. Therefore the glottal area remains dark, i.e. the gray level keeps low. Hence, the vocal fold edge segmentation is not influenced by the additional projected laser lines. Within the segmented Fig. 3b there is an angle φ between the glottal and the y-axis of the image. This angle is given by

$$\cos\varphi = \langle \vec{e}_{\tilde{y}}, \vec{e}_y \rangle, \tag{1}$$

where $\vec{e}_{\tilde{y}}$ is the y-unit vector with respect to the $\widetilde{I}(x,y,t)$ coordinate system and $\vec{e}_y$ is the corresponding unit vector of the glottal axis coordinate system as determined by the glottal area segmentation (Lohscheller et al., 2007). Here, $\langle\cdot,\cdot\rangle$ indicates the scalar product. An angle correction is applied by rotating $R_\varphi\{\cdot\}$ the unprocessed image series

$$I(x,y,t) := R_\varphi\left\{\widetilde{I}(x,y,t)\right\}, \tag{2}$$

in order to get the glottal axis parallel to the y-axis. The angle correction standardizes the orientation in terms of the glottal axis and supports a simplified determination of a region of interest. All line detection steps are applied to $I(x,y,t)$.

Besides the laser lines, there are periodically returning light spots on the vocal folds in the endoscopic video. In Fig. 4, every second image is grabbed out of the glottic cycle. During the opening phase the vocal folds move apart and thereby minimize the tissue area from which the laser lines are reflected. Thus, the shortest laser lines are found in the maximal open glottis. Short laser lines correspond to less calibration information and therefore increase the uncertainty of the line detection in the corresponding image frames. The spot-like reflections can be mistaken for short laser lines and hence they are an error source for the line detection.

4. Detection of parallel laser lines

As a precondition for calibration, the laser lines have to be automatically detected, especially when applied to a large number of clinical recordings. In the following an automated laser line detection is presented that calculates a metrical scaling factor γ for HS image data. An overview of the algorithm is schematized in Fig. 5.

4.1. Data reduction: region of interest (ROI)

Laser lines can only be used for calibration purposes if they are directly located on the vocal folds. Hence, it is valuable to limit the search space to the vocal fold region within the image data. The segmented vocal fold edges are used to define two ROIs – one for the left and one

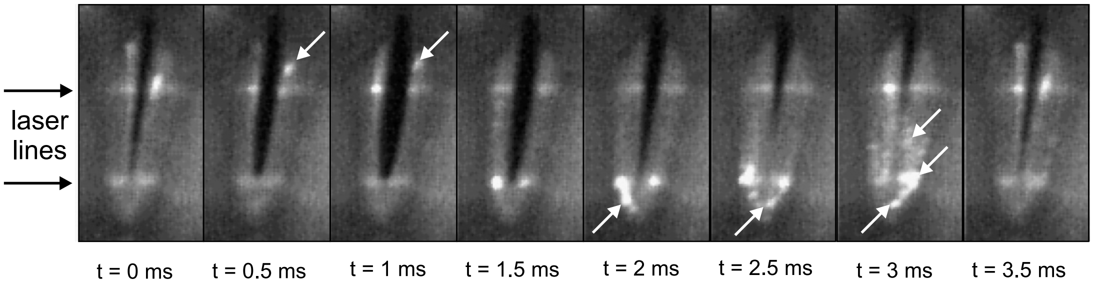

Fig. 4. One period of vocal fold oscillations is shown for every second image $\tilde{I}(x,y,t)$ of the glottal section. The vocal fold tissue reflects the parallel laser lines. There are also spot-like reflections, which are marked by white arrows.

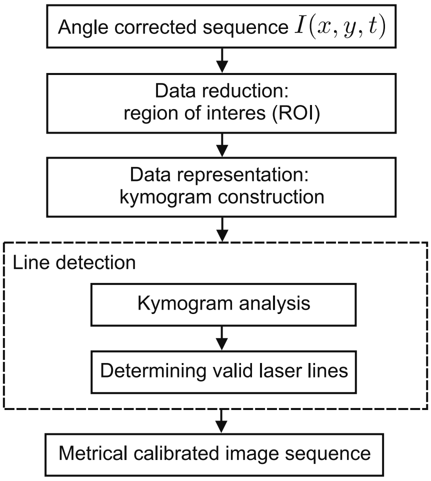
Angle corrected sequence $I(x,y,t)$
Data reduction: region of interes (ROI)
Data representation: kymogram construction
Line detection
Kymogram analysis
Determining valid laser lines
Metrical calibrated image sequence

Fig. 5. Overview of the algorithm used for laser line detection.

for the right vocal fold. The ROIs are two rectangles as illustrated in Fig. 6. The y_{ROI}-range is calculated from the length $y_l := y_v - y_d$, which is the difference of the extremal ventral and dorsal y-coordinate of the glottis in the complete open state. A safety margin of one third of y_l is added to y_{ROI} to assure that the ROI length covers the vocal fold length. The ROI width x_w for one vocal fold is estimated by a ROI width/length relation of 0.3, which is in the same order of magnitude as the measured width/length quotients of the glottal area (Arndt and Schäfer, 1994). The maximal x_r and minimal x_l positions of the segmented right and left vocal fold edges are expanded by x_w in order to obtain the x_{ROI}-range.

The restriction to the ROIs partly eliminates common disturbances of the endoscopic recordings. Especially, the influence of overexposure is suppressed, which primarily appears at the edges of the image data. Additionally, the reduction to the ROIs speeds up computation time.

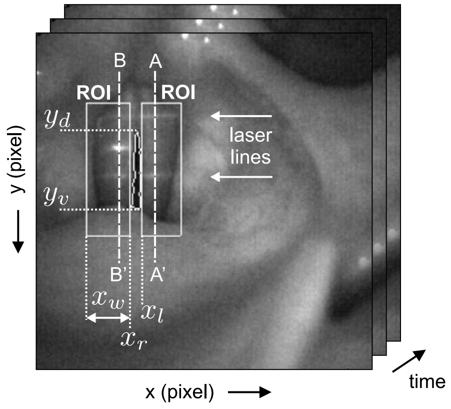

Fig. 6. Two rectangles define the ROIs for the laser line detection. The glottal gap between the left and right vocal fold is excluded, since the laser lines are only reflected from the vocal fold tissue. The ROIs are calculated from the coordinates of the segmented vocal fold edges. Two vertical scan lines AA′ and BB′ for the kymogram construction are exemplarily drawn.

4.2. Data representation: kymogram construction

Within HS recordings succeeding frames are highly correlated. An analyzing method that makes use of this HS characteristic is a kymogram representation of the video data. For example Wittenberg et al. (1995) and Mergell et al. (2000) used the kymographic principle for displaying and investigating the motion of the vocal fold edges in detail. To do so, the kymogram is constructed from the recorded video $I(x,y,t)$ by reducing the dimensionality of the original data, to a two-dimensional subspace by keeping the y-coordinate at a fixed position. Hence, a kymogram provides a localized view on the recorded sequence.

Here, the data representation by a kymogram is utilized to transfer the line detection problem to the ty-plane of the HS image sequence $I(x,y,t)$. For this, a vertical scan line (constant $x = X_i$-coordinate) is taken from a frame of the endoscopic image series at the time $t = t_0$. Further $N-1$ vertical lines are concatenated side by side from subsequent frames each at the same X_i-coordinate. A new image builds up, the kymogram $K_{X_i}(t,y)$, in which the abscissa is a time axis:

$$K_{X_i}(t,y) = [I(X_i,y,t_0), I(X_i,y,t_1), \ldots, I(X_i,y,t_{N-1})]. \tag{3}$$

In order to facilitate the visual assessment and to make use of the entire range of intensity values a contrast enhancement is applied. The input pixel intensity p_{in} is linearly stretched to the output pixel intensity p_{out} by

$$p_{\text{out}} = (p_{\text{in}} - a)\frac{2^z - 1}{b - a}, \tag{4}$$

with $a := \min\{K_{X_i}(t,y)\}$, $b := \max\{K_{X_i}(t,y)\}$, and z denotes the number of bits for the graylevel scale. Kymograms are generated for each x_{ROI} -coordinate. They compose a set of $2x_{\mathrm{w}}$ kymograms.

The kymogram representation is beneficial for the laser line detection, if the number of frames on which the laser lines are visible on the vocal folds exceeds the line length within a single frame of the HS image series. Then $K_{X_i}(t,y)$ results in an elongated laser line representation, which facilitates a robust detection of the lines. A time span of 32 ms, which corresponds to $N = 128$ (32 ms · 4000 Hz) frames, ensures on the one hand an elongated line representation compared to typical laser line lengths of <70 pixels in a single frame for the used HS camera. On the other hand, 32 ms is in the same order of magnitude as typical time constants for muscle contraction and mechanical delays (Perlman and Alipour-Haghighi, 1988) and thus sufficiently short to provide a constant distance between the endoscope and the vocal fold plane. The benefit of a constant distance is that the laser lines possess a horizontal orientation within a kymogram $K_{X_i}(t,y)$ regardless of the orientation in the angle corrected image frames $I_n(x,y)$.

Fig. 7a illustrates the kymogram construction for two vertical scan lines AA′ and BB′ each at a different $x = X_i$-coordinate. Fig. 7b depicts the kymogram $K_{\mathrm{AA}'}(t,y)$. The lines can be viewed as the two brightest horizontal lines. The laser line lengths are longer in the kymogram representation compared to the lengths within a frame $I_n(x,y)$ of the recording. Fig. 7c shows the kymogram $K_{\mathrm{BB}'}(t,y)$. The scan line BB′ crosses a reflection, which is located between the laser lines. The reflection dominates the intensity values in $K_{\mathrm{BB}'}(t,y)$, which is an artifact.

4.3. Line detection step 1: kymogram analysis

As the calibration marks are lines the Radon transform is suitable to detect them (Leavers, 1992). The Radon transform maps two dimensional images to a parameter domain $D(\rho,\theta)$, where lines can be identified as peaks. Here, ρ is the distance of a straight line to the origin and θ denotes the angle of the corresponding normal vector in the Hessian normal form

$$\rho = t\cos\theta + y\sin\theta. \tag{5}$$

Since the laser lines are known to be horizontal in a kymogram $K_{X_i}(t,y)$, the parameter range for the transformation angle can be fixed to $\theta = 90°$. Hence, the normal form of Eq. (5) reveals the identity of $\rho = y$ for $\theta = 90°$. Consequently, the Radon transform of $K_{X_i}(t,y)$ is given by its row sum

$$D_{X_i}(\rho = y, \theta = 90°) := \sum_{n=0}^{N-1} K'_{X_i}(t,y)\bigg|_{t=n\Delta T_s}. \tag{6}$$

The normalization of Eq. (6) to the row length N yields the temporal mean of the Radon peak profile

$$r_{X_i}(y) := \frac{1}{N} D_{X_i}(\rho = y, \theta = 90°). \tag{7}$$

The Radon peak profile for the two kymograms $K_{\mathrm{AA}'}(t,y)$ and $K_{\mathrm{BB}'}(t,y)$ is shown in Fig. 8. Positive and negative peaks form out in $r_{\mathrm{AA}'}(y)$ and $r_{\mathrm{BB}'}(y)$ and correspond to bright and dark line structures within the kymograms. Detecting laser lines means finding appropriate positive peaks. A Radon peak profile contains many local extrema and it has to be decided, which maxima belong to laser lines. On the basis of the peak values alone a decision is inappropriate, since the laser lines do not correspond mandatorily to the largest maxima. Within a kymogram also artifacts occur, which can dominate the peak profile, exemplarily shown in $r_{\mathrm{BB}'}(y)$. In order to facilitate the peak detection in the transformation domain following preprocessing steps are applied to each kymogram $K_{X_i}(t,y)$.

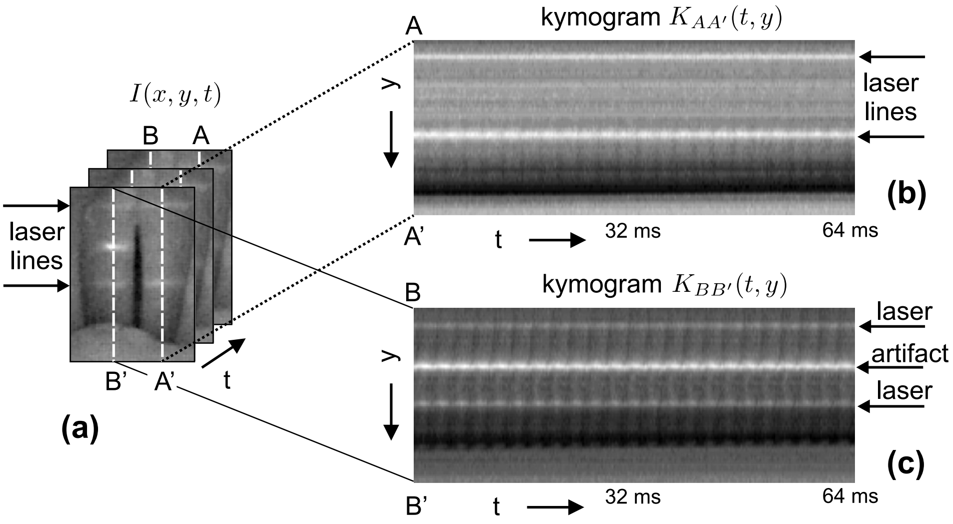

Fig. 7. A zoomed part of the glottis with the vertical scan lines AA′ and BB′ is depicted in subfigure (a). Contrast enhanced kymograms are shown in the subfigures (b) and (c). The time course of the pixel intensity can be viewed. In the kymogram representation the laser line length is longer than in single frames of the image data. In (b) the two brightest horizontal lines correspond to the laser lines. In (c) a disturbing reflection between the two laser lines is visible.

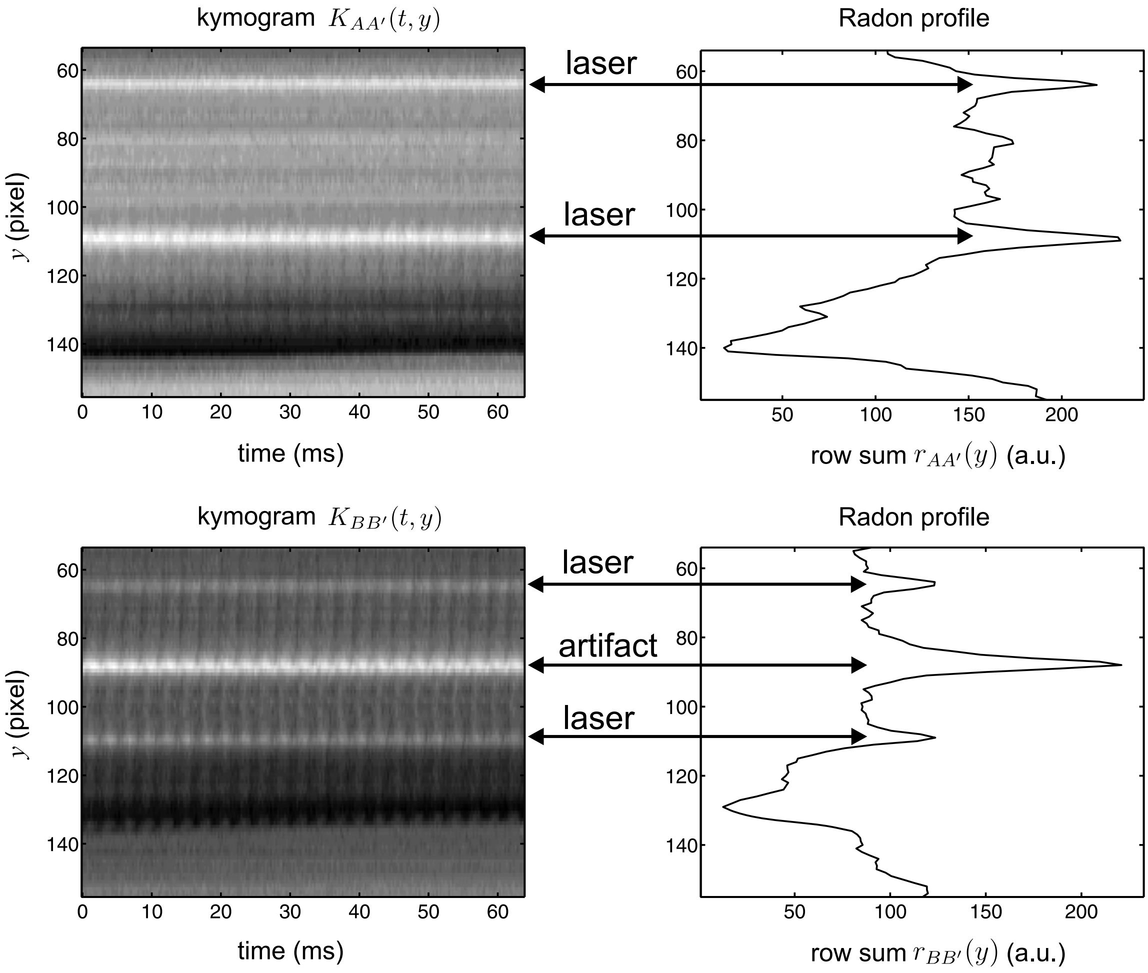

Fig. 8. The kymograms $K_{\mathrm{AA'}}(t,y)$ and $K_{\mathrm{BB'}}(t,y)$ and the corresponding Radon peak profiles. Positive peaks identify bright line structures within the kymograms and negative peaks identify dark line structures.

4.3.1. Smoothing

As a first step, periodically returning reflections within a kymogram $K_{X_i}(t,y)$ are reduced by applying a two-dimensional (space × time) median filter. The filter size is $3 \times f_s/(4 \cdot f_0)$, where f_0 is the fundamental frequency of the vocal fold oscillation. The spatial filter length of 3 pixels is less than the typical line width $w = 5$ pixels and thus prevents to smooth out laser induced intensity changes along the y-axis. The temporal filter length is coupled to the fundamental frequency f_0, since the periodicity of the light spots depends on the frequency of the vibrating vocal fold tissue. To suppress the light spots the temporal filter length is set to a fourth of the oscillation period. The median filtering preserves the line edges and smooths out isolated light spots in $K_{X_i}(t,y)$. Fig. 9 shows the smoothed versions of the two examples $K_{\mathrm{AA'}}(t,y)$ and $K_{\mathrm{BB'}}(t,y)$.

4.3.2. Gradient information

Since laser lines exhibit a characteristic progression of an increase followed by a decrease in the intensity values along the y-axis, derivative information of $K_{X_i}(t,y)$ can be used to increase the detection robustness. The intensity changes along the y-axis are emphasized by calculating the derivative of the median filtered kymogram $K_{X_i}(t,y)$ as

$$K'_{X_i}(t,y) := \frac{\partial}{\partial y} K_{X_i}(t,y) = \frac{1}{w}\left[K_{X_i}(t,Y_{j+s}) - K_{X_i}(t,Y_{j-s})\right], \qquad (8)$$

where $K_{X_i}(t,Y_j)$ is the Y_jth pixel of the median filtered kymogram and $s = (w-1)/2$ is the laser line half-width. A similar procedure to incorporate gradient information for making a Hough-based needle detection robust in ultrasound images was presented in Okazawa et al. (2006). In Fig. 10 the derivative data of the two examples $K_{\mathrm{AA'}}(t,y)$ and $K_{\mathrm{BB'}}(t,y)$ is shown. Line structures can be identified by a positive peak immediately followed by a negative peak.

4.3.3. Thresholding

The peaks in the derivatives $K'_{X_i}(t,y)$ are disturbed by background noise. A threshold operation is applied to separate the noise from the signal components. It is assumed, that the highest and lowest peaks in the derivative $K'_{X_i}(t,y)$ correspond to line structures, whereas smaller intensity

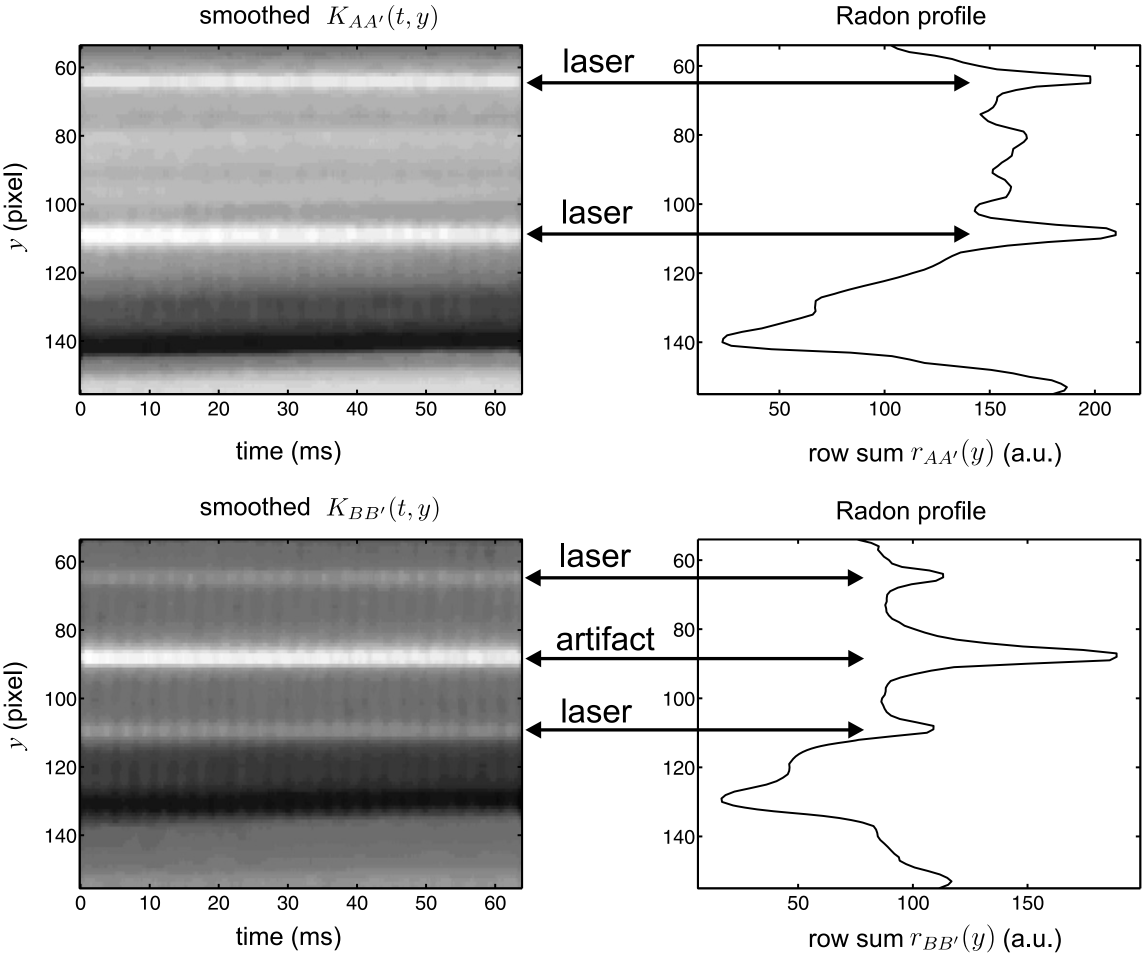

Fig. 9. Median filtered kymograms $K_{\mathrm{AA'}}(t,y)$ and $K_{\mathrm{BB'}}(t,y)$ and the corresponding Radon peak profiles. Local extrema are reduced compared to Fig. 8.

changes result from the background noise. The less frequent positive and negative peaks should be preserved and the more frequent noise values are eliminated. For this purpose, a lower and upper bound for the thresholding is determined by calculating the 0.15 and the 0.85 percentile of the values in $K'_{X_i}(t,y)$. The values that lie within these threshold levels belong to noise and are set to zero. The values <0.15 belong to negative peaks and values >0.85 belong to positive peaks. Fig. 11 illustrates the effect of the threshold operation.

4.3.4. Identifying line structures

Line structures in the threshold-clipped derivative $K'_{X_i}(t,y)$ are characterized by positive peak values $r_{X_i}(Y_j^+)$ immediately followed by negative peak values $r_{X_i}(Y_j^-)$. The index j numbers the individual peaks within one kymogram. Other intensity changes for example light overexposure at the epiglottis do not possess this characteristic. Hence, intensity changes that do not match a laser line profile can be filtered out if the following inequality is not fulfilled

$$\arg\left(r_{X_i}(Y_j^-)\right) - \arg\left(r_{X_i}(Y_j^+)\right) < 2w. \tag{9}$$

This is demonstrated in Fig. 12a and b for the two Radon peak profiles of Fig. 11. Contemplable regions of width $2w$ are marked by gray color and the corresponding positive and negative peaks by ×-marks. The artifact reflection within the kymogram $K_{\mathrm{BB'}}(t,y)$ is still existent and needs to be eliminated in further processing steps. From the remaining peaks possible candidates for being a point of a laser line are calculated as

$$c_{X_iY_j} := \frac{1}{2}\left[\arg(r_{X_i}(Y_j^+)) + \arg(r_{X_i}(Y_j^-))\right]. \tag{10}$$

The candidates $c_{X_iY_j}$ of a single kymogram $K_{X_i}(t,y)$ serve as initial segmentation result of the associated X_i-coordinate in the HS image series $I(x,y,t)$.

The candidate detection is not only applied to one or to two specific kymograms $K_{X_i}(t,y)$. It is applied to all $2x_w$ kymograms of the x_{ROI}-range.

4.4. Line detection step 2: determining valid laser lines

4.4.1. Candidate assignment

The described line detection exploits only the information in the spatio-temporal ty-plane of the HS video to

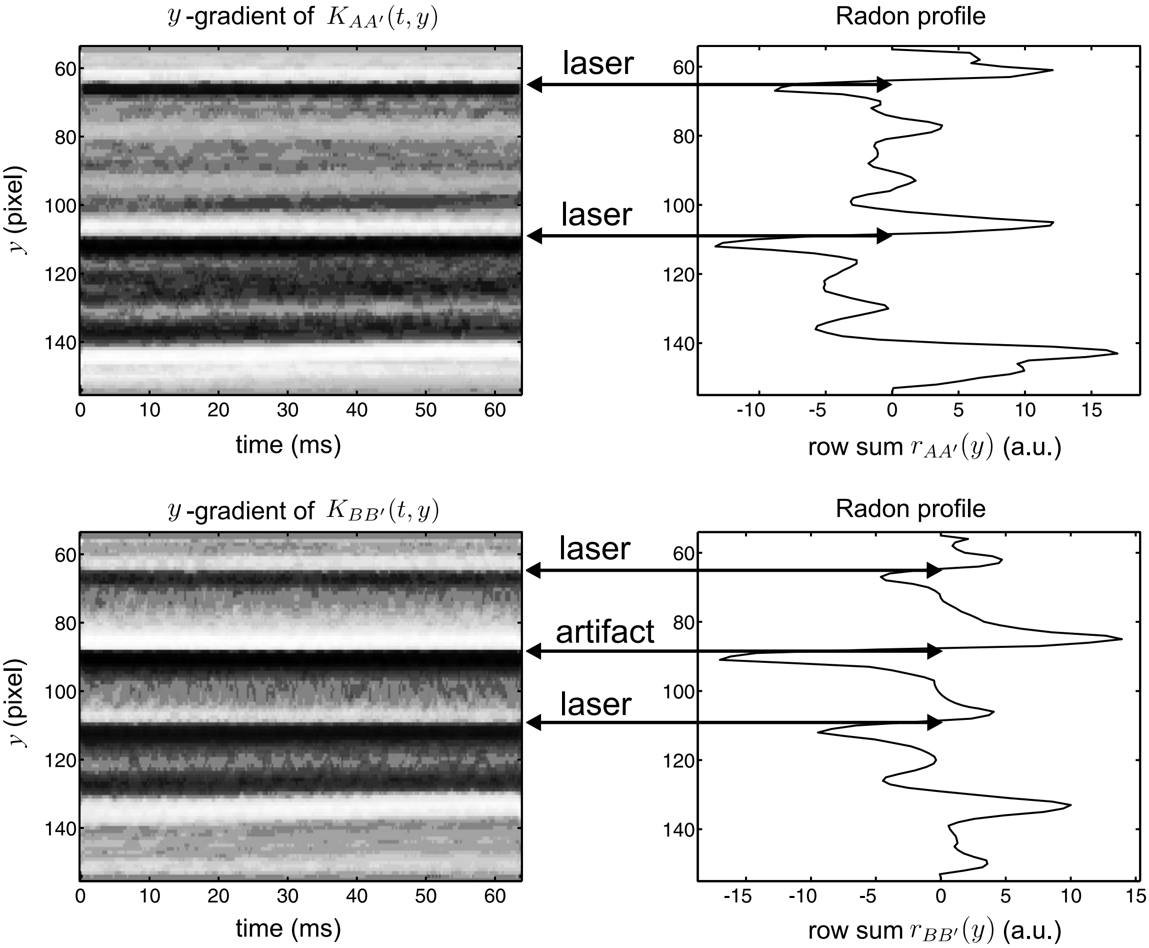

Fig. 10. Derivative of the kymograms $K_{AA'}(t,y)$ and $K_{BB'}(t,y)$ and the corresponding Radon peak profiles. Line structures are apparent as a change from a positive peak immediately followed by a negative peak.

obtain the candidates $c_{X_iY_j}$. Now, the information in the xy-plane is used to connect the candidates $c_{X_iY_j}$ in spatial dimension, which were determined independent from each other for the different X_i-coordinates. The aim is to assign candidates $c_{X_iY_j}$ to one of the laser lines or to reject them as artifacts or outliers. The candidates $c_{X_iY_j}$ are gathered in a binary image

$$B(x,y) := \begin{cases} 1 & \text{if } c_{X_iY_j}\exists \text{ for } x = X_i \wedge y = Y_j \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

for all x and y-coordinates. Adjacent candidates form out line paths, whereas artifact and outlier candidates have a more clustered shape within $B(x,y)$. This is illustrated in Fig. 13. The dominant line paths in $B(x,y)$ are assumed to be the laser lines. They are identified by a Hough transform, which maps each image point of the binary image $B(x,y)$ to a parameter domain $D(\rho,\theta)$, like the Radon transform does for graylevel images. In order to make the Hough transform robust against lines with wiggles (Toft, 1996b), the width of the transformation core is raised from one pixel to a width of $2k+1$ as proposed by Deans (1983). Here, k denotes the additional one-sided bandwidth and it is set to $k=1$ pixel.

The estimation of the line parameters is limited by the discretization of the calculation of the Hough transformation. By applying a least-square line fitting this could be circumvented. Although, the Hough transformation does explicitly not estimate parameters in the least-square sense, it is reasonable that considering the lateral resolution of the images and the accuracy of the parallel projection of the laser lines the benefit of a least-square estimate of the line parameters is doubtable. Using the parameters estimated by the Hough transformation furthermore provides a very robust estimate of the line parameters: Rousseeuw and Leroy (1987) reports an excellent robustness of the Hough transformation in case of low-level noise. The breakdown occurred at 90%, i.e. stable parameter estimation was obtained until 90% of the data set was contaminated by outliers. Most other techniques do not attain a breakdown point of 30%.

The global maximum in the parameter domain $D(\rho,\theta)$ represents the dominant line path of $B(x,y)$. The corresponding coordinates ρ_1 and θ_1 are obtained as

$$(\rho_1,\theta_1) = \arg\max_{\rho,\theta} D(\rho,\theta). \tag{12}$$

The parallelism of the laser lines implies that the normal angle θ_2 of the second line is identical to the first line

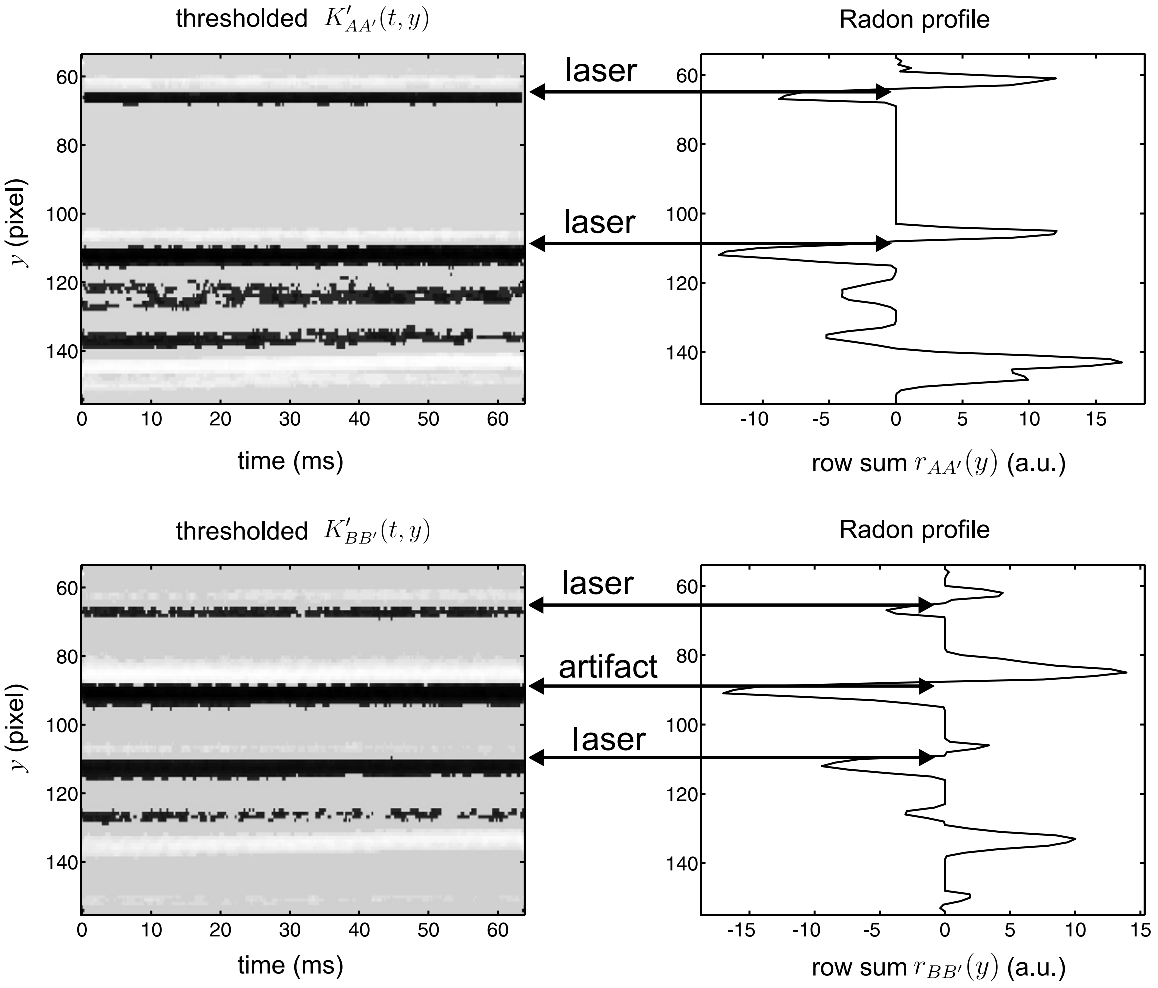

Fig. 11. Thresholded derivative of kymograms $K_{\mathrm{AA'}}(t,y)$ and $K_{\mathrm{BB'}}(t,y)$ and the corresponding Radon peak profiles. The threshold operation suppresses the frequent values belonging to noise, whereas the dominant peaks are kept.

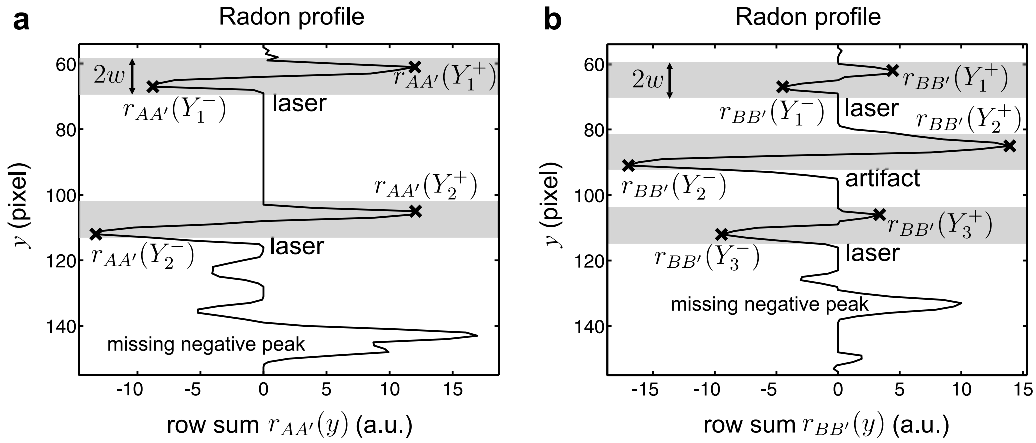

Fig. 12. Temporal mean of the Radon peak profile $r_{X_i}(y)$. The subfigures (a) and (b) depict the Radon profile belonging to the kymograms constructed along the AA′ and BB′ scan line. Lines are characterized by an increase of $r_{X_i}(y)$ to a positive peak followed by a decrease to a negative peak within the width of $2w$. These regions are marked by gray color.

($\theta_2 = \theta_1$). Thus, the line parameter ρ_2 is found at the second largest local maximum along the angle θ_1

$$\rho_2 = \arg\max_{\rho,\rho\neq\rho_1} D(\rho,\theta_1). \tag{13}$$

To derive an absolute scale from the endoscopic image sequence $I(x,y,t)$ the pixel distance d_{p} and the metrical distance d_{m} between both laser lines have to be related to each other. The distance d_{m} is known from

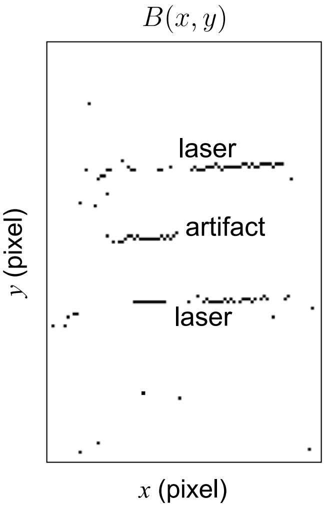

Fig. 13. Binary image $B(x,y)$ of the laser line candidates $c_{X_iY_j}$ for the kymogram example $K_{\mathrm{BB}'}(t,y)$, which contains an artifact due to a prominent and wide-spread reflection.

the LLPS specifications. The distance d_p is calculated as

$$d_\mathrm{p} = |\rho_1 - \rho_2| \tag{14}$$

from the determined line parameters. The quotient between the metrical and the pixel distance, yields the scaling factor

$$\gamma = d_\mathrm{m}/d_\mathrm{p} \tag{15}$$

for the observed image frames $I(x,y,t)$.

4.4.2. Consistency check

The final step is a consistency check on the calculated parameters. It decides whether to accept or reject the results as valid for calibration. The decision bases on the characteristics of the determined candidates $c_{X_iY_j}$ and on a priori information about the LLPS as well as the HS recording specifications.

Based on the estimated parameters (ρ_1,θ_1) and $(\rho_2,\ \theta_2)$ the candidates $c_{X_iY_j}$ are separated into two pixel sets $L_s(x,y)$ with $s \in \{1,2\}$ for the first and second laser line, respectively. A candidate $c_{X_iY_j}$, represented by $B(x,y)$, is assigned to one of the laser line sets $L_s(x,y)$, if its $(X_i,\ Y_j)$-coordinate corresponds to one of the estimated line parameters (ρ_s,θ_s):

$$L_s(x = X_i, y = Y_j) = \begin{cases} B(X_i, Y_j) \text{ if} & \rho_s = X_i\cos\theta_s + Y_j\sin\theta_s \\ \text{reject} & \text{otherwise} \quad \forall i,j; s \in \{1,2\} \end{cases} \tag{16}$$

This criterion sorts out artifacts and outliers and keeps only those candidates $c_{X_iY_j}$ that contributed to the Hough line estimation. This is exemplarily demonstrated in Fig. 14a–c. Fig. 14a shows the glottal section of a frame $I_n(x,y)$ with two laser lines and an artifact reflection. The laser line candidates $c_{X_iY_j}$, gathered in $B(x,y)$, are overlaid in Fig. 14b. Most of the candidates $c_{X_iY_j}$ mark the brightest path of the laser lines along the x-coordinate. However, there are also candidates $c_{X_iY_j}$ that label the artifact reflection. There are further candidates $c_{X_iY_j}$, the so-called outliers, as exemplarily indicated by two white arrows. Fig. 14c displays the effect of the artifact and outlier elimination given in Eq. (16). From the original candidates $c_{X_iY_j}$ only those points survived that mark the brightest path of each laser line.

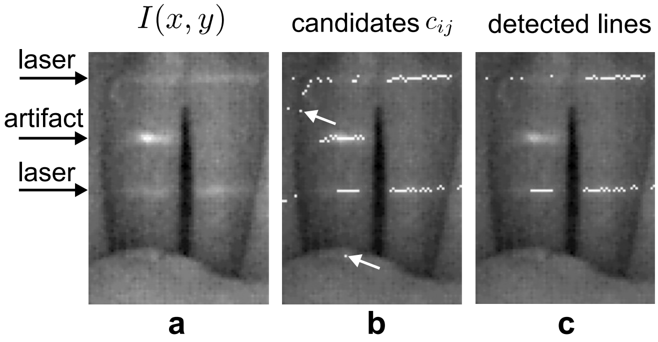

Fig. 14. A part of the image frame $I_n(x,y)$ that contains two laser lines and an artifact reflection. (a) Depicts the raw image. (b) The candidates $c_{X_iY_j}$ are overlaid on $I_n(x,y)$. They label the laser pixels but also the reflection and some outliers as indicated by the white arrows. (c) After an estimate of the line positioning via the Hough transform, only those points are kept that contribute to the line.

Following assumptions are made for the consistency check: The ROI has the width $2x_\mathrm{w}$ and covers the vocal folds. Ideally, the projected laser lines and the related pixels $L_s(x,y)$ cover the entire ROI width, while specular reflections on the vocal fold tissue are spatially limited. Further on, line structures differentiate from specular reflections in $L_s(x,y)$ by a larger number of connected pixels C_s in the neighborhood. Falsely incorporated light reflections have normally a reduced number of connected neighbor pixels C_s. Thus, the number of pixels $|L_s(x,y)|$ related to the ROI width multiplicated with the connection length C_s serve as consistency criterion

$$Z_s = \frac{|L_s(x,y)|}{2x_w} \cdot C_s, \quad s \in \{0,1\}. \tag{17}$$

Classification on data sets with a decision tree (Mierswa et al., 2006; Hand et al., 2001) revealed to discard line sets $L_s(x,y)$ for values $Z_s < 1.09$. Beyond the characteristic of the line set $L_s(x,y)$, the knowledge about the recording specifications are incorporated to reject unreasonable pixel distances d_p. The result is only accepted if d_p is within a 15% region around the expected value. Otherwise the distance d_p is discarded and the calibration is marked as invalid.

5. Validation

The line detection method was validated by a manual segmentation of the laser lines done by experts within HS recordings of the vocal folds. A group of 12 experts consisting of medical doctors and image processing practitioners familiar with HS image series were asked to mark both laser lines in a single frame of ten different HS videos. The interline distances d_p^E as determined by the individual experts were evaluated for (1) the inter-rater reliability and (2) the group average was compared to the outcome of the interline distance d_p that was automatically computed from the proposed line detection.

6. Application to clinical data

In order to demonstrate the robustness as well as the clinical applicability of the LLPS and the proposed line detection, the algorithm was evaluated within 10 HS recordings of volunteers with a normal voice production. In order to validate the decision performance of the proposed algorithm 25 HS sections are chosen from the 10 HS recordings. The 25 HS sections are comprised of three types. In the first ten sections #1–#10 both laser lines are reflected from the vocal fold tissue as confirmed by visual inspection. The next five sections #11–#15 contain only one laser line and in the last ten sections #16–#25 no laser line is reflected from the vocal fold tissue.

7. Results

Firstly, the results of the LLPS accuracy measurements with a high-resolution camera are presented. Secondly, the results of the cross validation between the line detection and the manual segmentation are given. Thirdly, the applicability of the proposed algorithm is demonstrated with 25 sections of HS recordings. Finally, the laryngeal magnitudes, glottal area, vocal fold length, oscillation amplitude, velocity, and acceleration are given in metrical units for LLPS calibrated HS recordings.

7.1. LLPS accuracy

The accuracy measurements serve to validate the parallelism specifications of the LLPS hardware and determine which precision can be achieved with the LLPS in principle for high-resolution images. Table 1 summarizes the accuracy measurements for different projection heights h. The interline distance is in the average $d_m = 5.4 \pm 0.08$ mm and is identical to the given specification value of the LLPS. The parallelism of both laser lines in the xy-plane is measured with the $\Delta\theta$ beam divergence term. The average value is $\Delta\theta = 4.94 \cdot 10^{-2°}$. Therefore, the laser lines of the LLPS are regarded as parallel in the projected plane. The difference between the two scaling factors γ_r and γ gives the attainable accuracy of the LLPS for high-resolution images. The mean difference is $(0.84 \pm 0.49) \cdot 10^{-3}$ mm/pixel. Thus, the LLPS hardware allows precise metric calibration down to the μm region.

Besides the production tolerance of the LLPS the spatial resolution of the recording unit influences the attainable accuracy. The spatial resolution of the HS camera sensor is with 256 × 256 pixels less than the high-resolution images used for the accuracy measurements. An average height h of 70 mm from the LLPS to the glottal plane results in a recording area of 28 × 28 mm. Hence, the attainable accuracy reaches $28/256 \approx 0.1$ mm/pixel, which is by a factor of $0.1/0.84 \cdot 10^{-3} \approx 119$ worse than the accuracy as determined with the high-resolution images, but still in the submillimeter region. Besides the spatial resolution also the tilt angle contributes to measurement errors. According to Schuberth et al. (2002) the tilt angle error is in the range of 1%. Since the glottal length is the largest magnitude an upper bound for the tilt error is calculated as 0.1 mm for a typical 10 mm visible glottal length. The total error for glottal length measurements is the sum of the spatial resolution and the tilt angle error and is less than 0.2 mm.

Table 1
Measurement accuracy of the LLPS for high-resolution images

h	d_m	$\Delta\theta \cdot 10^{-2}$	$\lvert\gamma_r - \gamma\rvert \cdot 10^{-3}$
49	5.43	5.93	0.9
58	5.53	8.06	0.4
68	5.33	4.43	0.9
79	5.37	3.46	1.6
87	5.34	2.84	0.4
mean	5.40 ± 0.08	4.94 ± 2.10	0.84 ± 0.49

The projection height h and the determined laser line distance d_m are given in mm. $\Delta\theta$ measures in degree the parallelism of the laser lines in the xy-plane. The unit of the scaling factor γ_r of the high-precision workpiece and the metrical scaling factor γ as determined with the LLPS is mm/pixel.

7.2. Validation

For evaluation purpose the outcome of the automatic line detection to the manual segmentation results was compared. Fig. 15 illustrates the interrater reliability for ten images taken from different clinical HS recordings. The expert deviation of the interline distance d_p^E are characterized as boxplots. The between expert variance amounts 0.84 pixel in the average. In addition, the calculated line distances d_p of the detection algorithm are drawn in as $*$-marks. The difference between d_p and the mean estimates

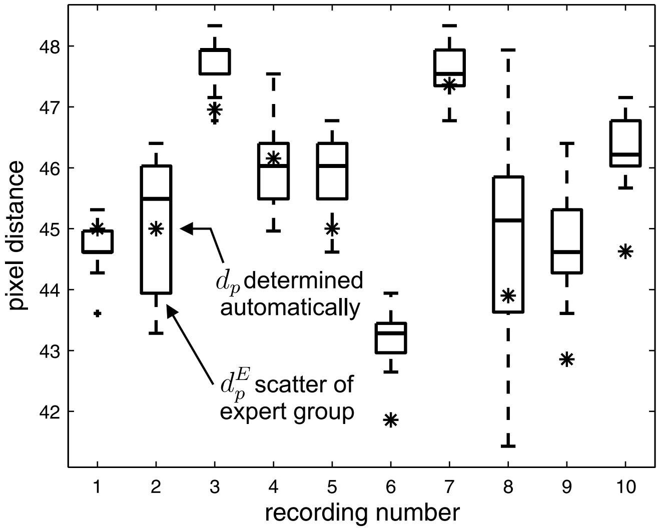

Fig. 15. Comparison of the interline distance d_p^E as manually determined by an expert group and the interline distance d_p as automatically determined from the line detection algorithm.

of the manual segmentation d_p^E amounts 0.83 pixel with a variance of ±0.74 pixel. Hence, the line detection algorithm differs only in the order of 1 pixel from the expert group. This results in a total relative error of 1.8% between the manual and the automatic segmentation yielding a difference of 0.02 mm referring to typical vocal fold deflections of ≈10 pixels. Since the line detection finds values close to the manual segmentation its application to HS recordings yields valid results.

7.3. LLPS calibrated clinical recordings

The results concerning the consistency check is illustrated in Fig. 16. Each HS recording has two bars associated with the first (dark gray) and the second (light gray) laser line. An identification is valid if the consistency for both laser lines is greater than the decision tree determined boundary $Z_s > 1.09$ and the distance error is less than 15%. These boundaries are sketched as planes in Fig. 16 and separate a region of acceptance from the illustrated space. Only the HS recording sections #1–#10 that really contain two laser lines on the vocal fold are within the region. All other recordings #11–#25 are correctly discarded from calibration, due to the associated consistency values Z_s and/or the distance error.

Figs. 17 and 18 show the line detection results of the HS sections #1–#10. All of the images are contrast enhanced and a Gamma correction was applied to increase the visibility condition. Column A depicts an angle corrected image frame $I_n(x, y)$ of the HS recording and displays information about the line detection. Values given are the number of detected lines, the scaling factor γ, the rotation angle of the glottal axis φ, the pixel distance d_p between the laser lines, and the consistency values Z_s for both laser lines. The rectangle covers the ROIs and additional the omitted glottal gap. A zoomed version of this area is depicted in Column B. The area mainly encloses the vocal folds. In column C the initial laser point candidates $c_{X_iY_j}$ as derived from the kymograms $K_i(t, y)$ are overlaid as white points on the zoomed image frame. The main portion of the points represents the brightest path of the two laser lines. The remaining points are outliers and artifacts. They mark intensity changes that are similar to the laser line characteristic. Column D shows the line positions and the pixel line distances as calculated with the Hough transform of the point set $B(x_i, y_j)$. For all 10 HS sections #1–#10 the detection result for both laser lines is reasonable as indicated by visual inspection of Figs. 17 and 18.

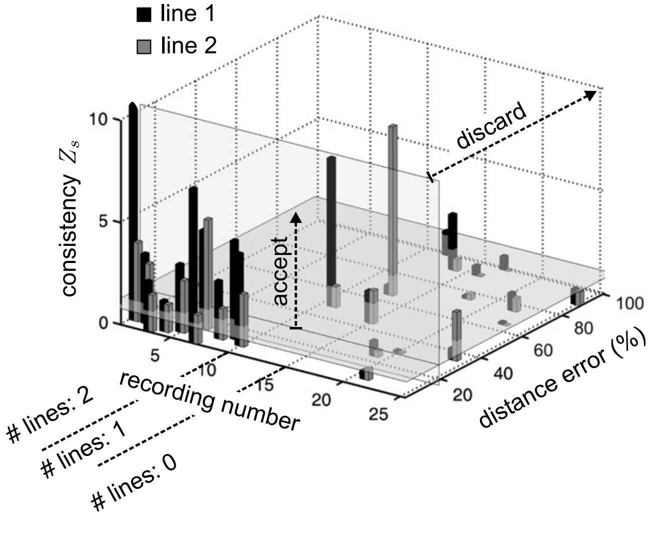

Fig. 16. The space to accept or discard the laser lines for the 25 HS sections. The decision to reject located lines depends on the distance error and the consistency value Z_s. The sections #1–#10 have two laser lines on the vocal folds. The determined calibration marks fulfill the requirements of a line and are in the region of acceptance. The sections #11–#25 are correctly discarded from calibration, since they contain only one or no laser line calibration mark on the vocal folds.

For the cases of a single and no laser line on the vocal folds, the line detection result is exemplarily depicted for HS sections #13 and #16 in Fig. 19. The line detection correctly estimates the number of laser lines to one for HS section #13 and identifies no laser line in the HS section #16.

7.4. Metrical units of laryngeal magnitudes

Vocal fold deflections and velocities of a 30 ms time span are exemplarily illustrated in Fig. 20 for the LLPS calibrated recording number #5. The velocities are calculated as time derivatives of the deflections. For the left side the velocities are positive for the opening phase, while the velocities are negative for the closing phase. For the right side it is vice versa. From the graph the impact velocities can be extracted, which are the velocities at the beginning of collision between the left and right vocal folds. This velocity may play a role in the genesis of vocal fold nodules.

In Fig. 21a and b a deflection–velocity plot is presented for the recording numbers #8 and #10. The range of values differs between the two individuals. For each subject the dorsal, medial, and ventral parts exhibit their own characteristic in the plot. This demonstrates the variability of vocal fold oscillation patterns even for subjects with a normal voice production.

For the HS sections #1–#10, the maximum deflection in mm of the vocal fold edges from the glottal axis and the associated peak velocity in m/s at three different positions along the vocal folds are plotted in Fig. 22. There is a linear trend to an increased velocity for an increased deflection at all three vocal fold positions in Fig. 22a. In the average the deflection amounts 0.49 ± 0.21 mm, 0.57 ± 0.22 mm, and 0.51 ± 0.16 mm for the dorsal, medial, and ventral position. The averaged corresponding peak velocity is 0.84 ± 0.24 m/s, 1.06 ± 0.23 m/s, and 1.01 ± 0.33 m/s. The ordinate of Fig. 22b depicts the maximum deflection and the abscissa designates the vibration frequency of the vocal folds. The frequency ranges from 180 Hz to 473 Hz and is identical for the dorsal, medial and ventral positions in the analyzed HS sections. The maximum deflection dif-

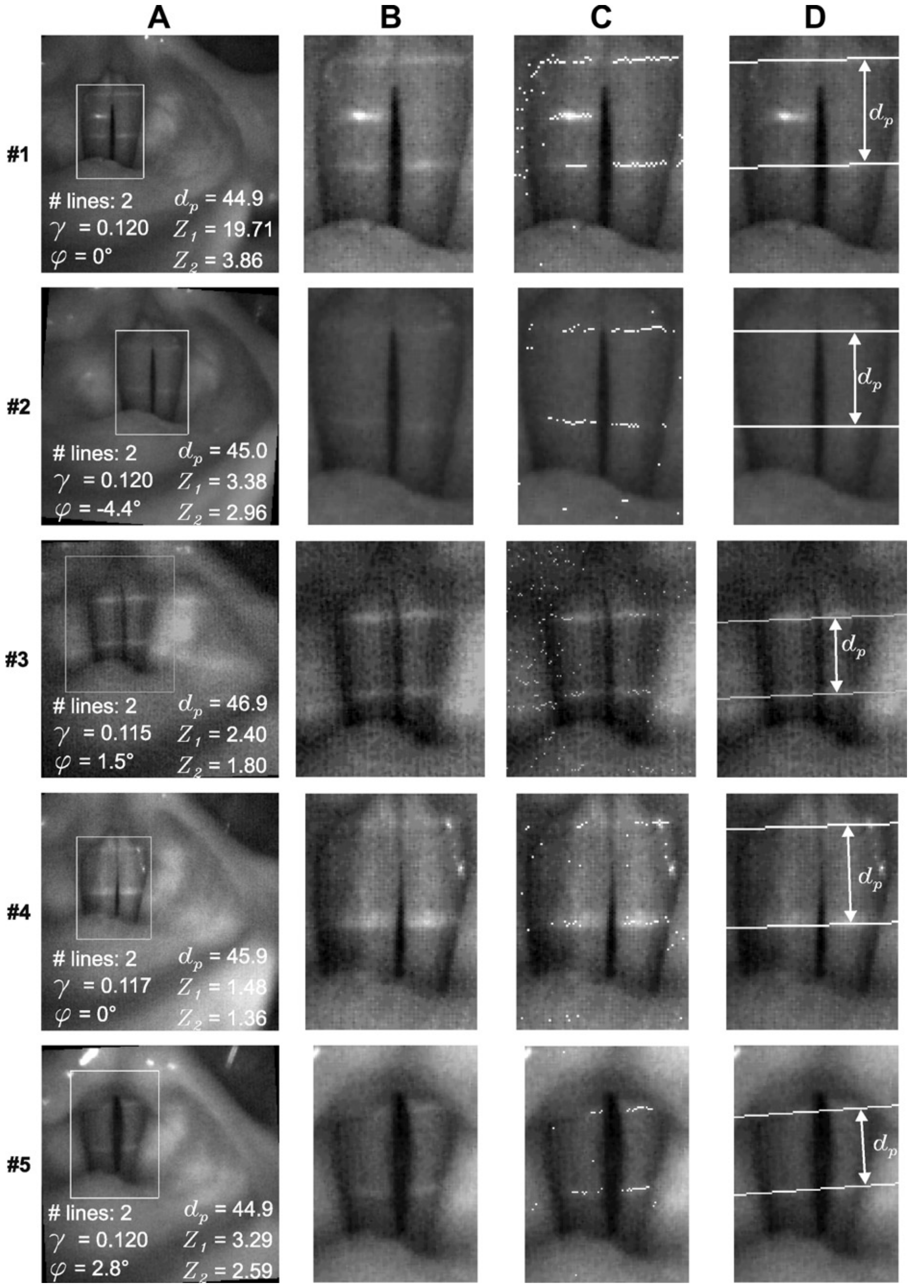

Fig. 17. The first five (#1–#5) results of the laser line detection. Column A shows a single frame $I_n(x, y)$ of the HS recording. The white rectangle describes the calculated ROI with the gap of the glottis. The lower part of the image contains following information: number of detected lines, the scaling factor γ, the rotation angle of the glottal axis φ, the pixel distance between the lines d_p, the consistency values Z_1 and Z_2 to reject laser lines. Column B is a zoomed version of the framed area in column A. Column C depicts the candidates $c_{X_iY_j}$ overlaid on the image of Column B. Column D illustrates the location of the laser lines as determined via Hough transform on the binary image $B(x_i, y_j)$.

fers for the three vocal fold positions. The sound pressure level is given in dB and the measured values for each HS section are depicted in Fig. 22b. The sound pressure level ranged from 70 dB to 83 dB. The deflection shows a tendency to decrease with an increased fundamental frequency and to increase with an increased sound pressure level.

The acceleration of the vocal fold edges is measured besides the deflection and velocity. The root mean square (RMS) acceleration for the dorsal, medial, and, ventral position amounts $829 \pm 232\ \mathrm{m/s^2}$, $1029 \pm 288\ \mathrm{m/s^2}$, and $981 \pm 270\ \mathrm{m/s^2}$. The averaged visible glottal length is measured to $y_1 = 8.55 \pm 1.03$ mm and averaged the glottal area for the maximum open condition during phonation amounts $9.0 \pm 2.77\ \mathrm{mm^2}$.

8. Discussion

The ability of an individual to produce speech becomes more important in modern service-oriented societies (Ruben, 2000). Hoarseness – as a common symptom of a voice disorder – restricts the communication skills. Laryngoscopic HS recordings provide the basis for the diagnosis

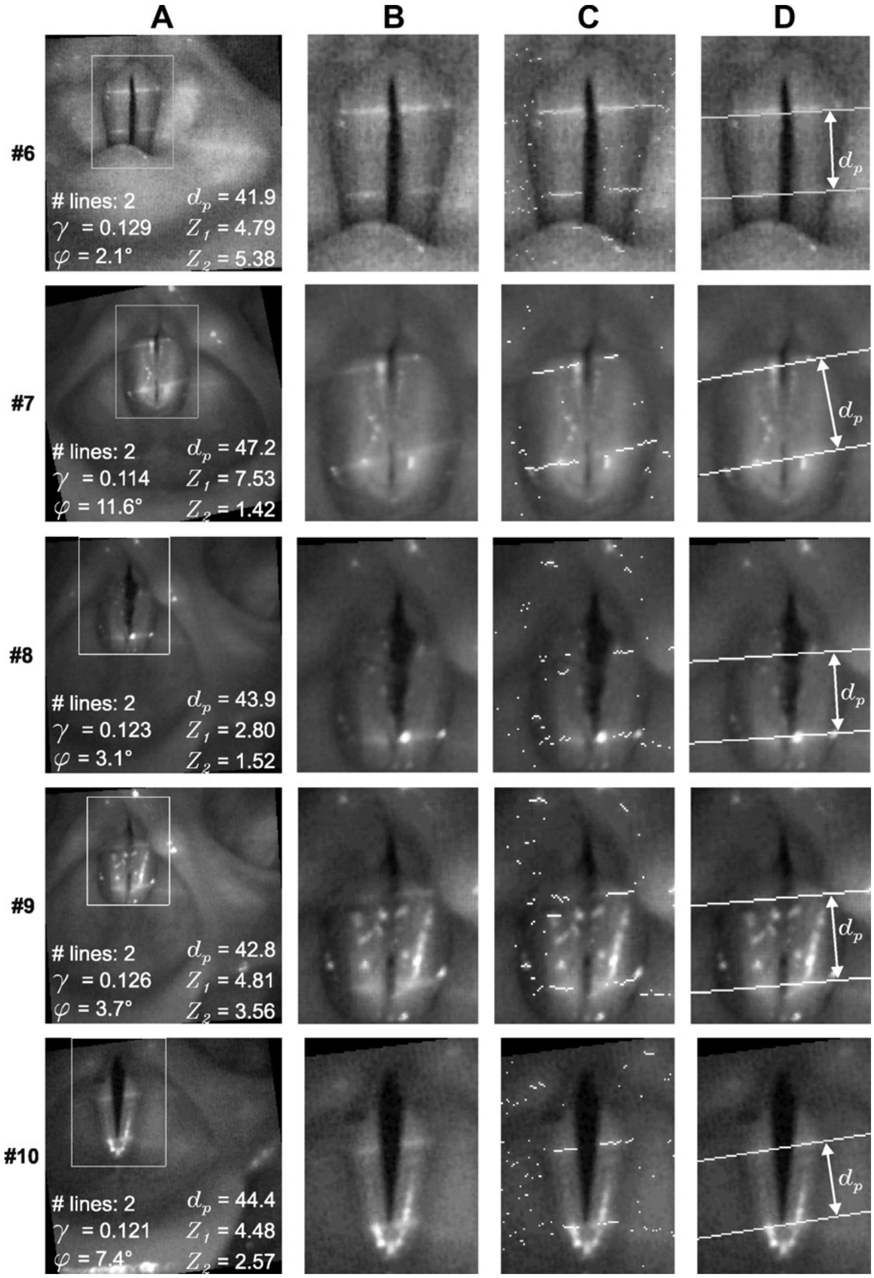

Fig. 18. The second five (#6–#10) line detection results. Explanations are given in the caption of Fig. 17.

of voice disorders, although they deliver only relative dimensions of the vocal fold dynamics. However, metrical calibrated recordings are essential for the clinical use in order to compare different recordings of normal and pathologic vibration dynamics among each other. Moreover, an automated calibration helps to clarify the genesis of vocal fold nodules or supports their prognosis. Hence, getting metrical data out of laryngoscopic recordings is a growing research field (Herzon and Zealear, 1997; Hertegård et al., 1998; Schuberth et al., 2002; Schade et al., 2004).

The application of the LLPS enables metrical calibration of laryngoscopic image sequences by projecting parallel laser lines across the vocal folds. The line projection is a novelty compared to existing laser projection systems that project only one single or two laser spots onto the vocal folds (Larsson and Hertegård, 2004; Schade et al., 2004). As an advantage, the laser lines are more silhouetted against spot-like light reflections due to their line characteristic. Therefore, ambiguities between the laser marks and the specular reflections of the primary endoscope light source are systematically reduced. The algorithm delivers stable results even in the case of low contrast between the laser lines and the vocal folds. A further advantage of laser lines compared to laser spots is the simplified positioning of the calibration marks onto the vocal folds, which is especially sophisticated for the examiner in small larynges. In

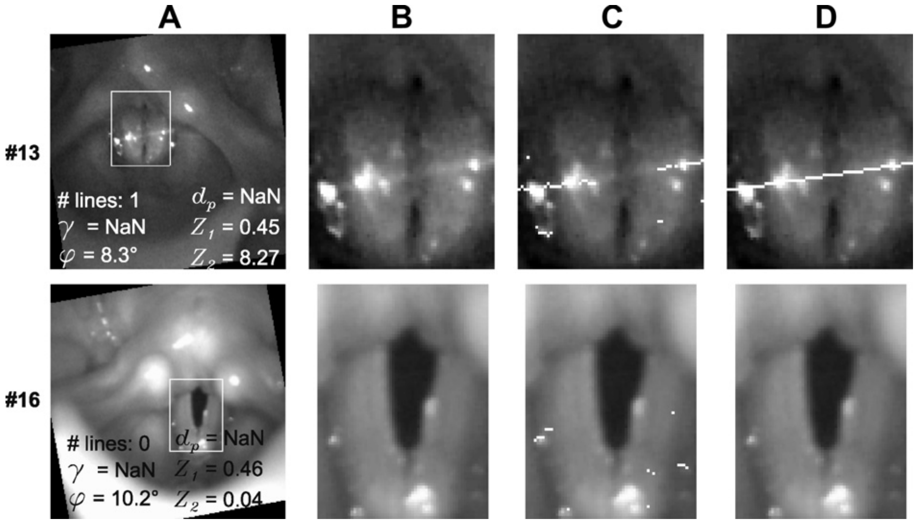

Fig. 19. The laser line detection result for recording #13 and #16. The vocal fold tissue reflects only a single and no laser line, respectively. The proposed line detection identifies these cases and marks the HS section as invalid for the calibration.

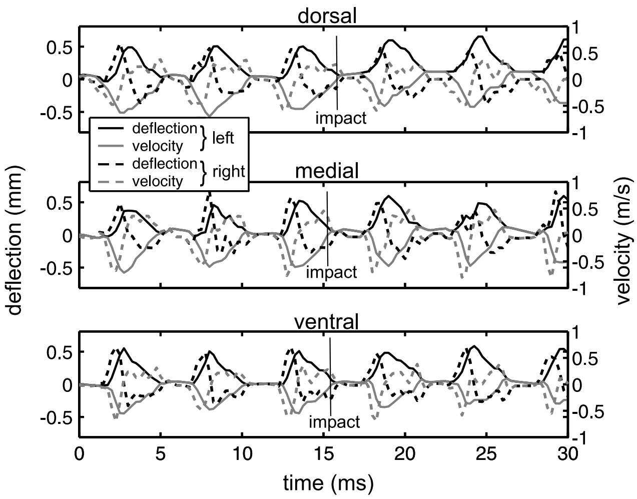

Fig. 20. Vocal fold deflections and velocities in metrical units for 30 ms of the recording number #5. One time instant is marked, which shows the beginning of collision between the left and right vocal fold side.

contrast using laser lines, the laser spots can easily drift away into the glottis, which means that no calibration information is visible within the recorded images.

The measurement precision of the LLPS hardware is in the μm region as shown with the accuracy measurements. The projected laser lines are parallel independent from the projection height. In HS recordings the spatial resolution of the HS camera sensor limits the attainable LLPS precision. Improving the spatial camera resolution of clinically available HS systems will yield in more precise measurements. There is a potential to be better by a factor of 119 as the accuracy measurements show. However, the measurement error due to the tilt angle in the HS recordings persists, which limits the accuracy in the LLPS measurement technique in general. Nevertheless, the LLPS coupled to the endoscopic HS camera enables high-precision measurements of laryngeal magnitudes in the submillimeter region.

The automated laser line detection is a two-stage process, which uses temporal and spatial characteristics of

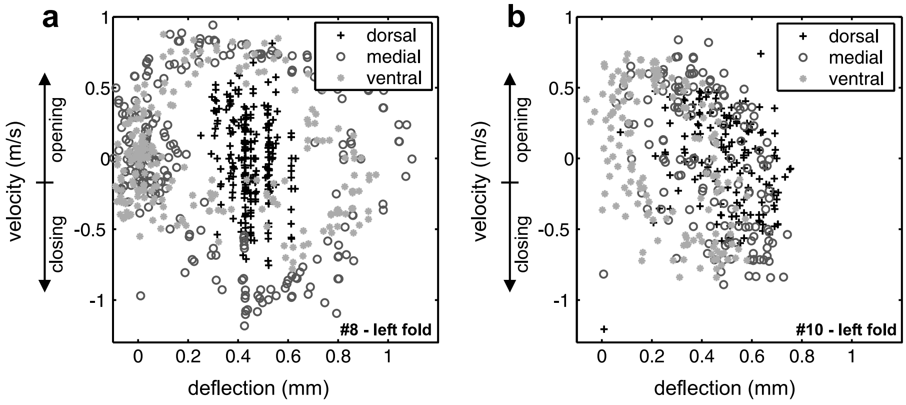

Fig. 21. Deflection–velocity profile of the left vocal fold for all time samples of one HS recording. (a) HS recording number #8. (b) HS recording number #10.

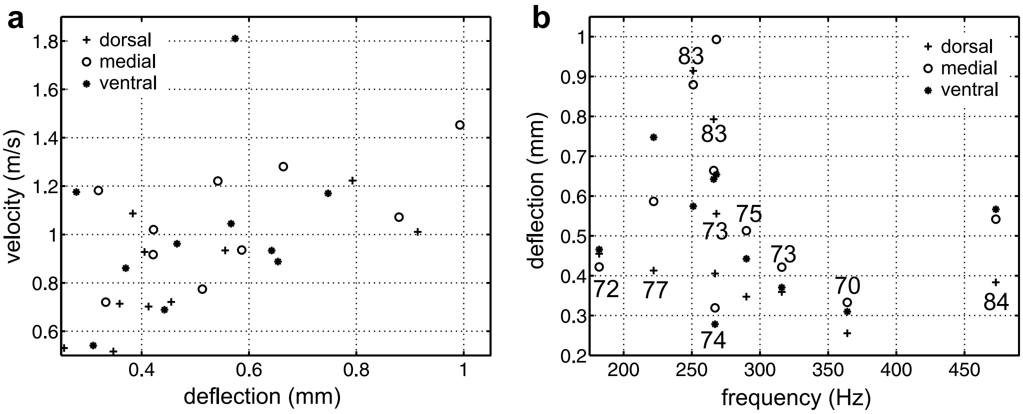

Fig. 22. Calibrated measures of vocal fold dynamics extracted at a dorsal, medial, and ventral position for the HS sections #1–#10. (a) Peak velocity over maximum deflection of the vocal fold edges from the glottal axis. (b) Maximum deflection over vibration frequency. The given numbers denote the sound pressure level in dB.

the projected laser lines one after another. The detection method involves basic image transforms and morphological operations, which are parameter controlled. The robustness of the line detection depends on the selection of these parameters. One of the parameters is regulated in dependence of the vocal folds vibration frequency, since periodically returning light spots arise due to their cyclic opening and closing. The other parameters could be suitable adjusted, since a priori knowledge about the recording specifications and the LLPS is available. For example, the laser line width w within the recordings is approximately known. If the spatial resolution of the camera or the dispersion of the projected laser lines improves, the corresponding parameters can easily be tuned to the new specifications.

From a computational point of view, the detection of the laser lines in each kymogram is advantageously compared to the detection in every frame of the raw image series. Not only one frame/kymogram is processed but a sequence/set is used to detect the laser lines. There is a gain factor as long as the number of processed frames N is greater than the number of kymograms $2x_w$ in the kymogram set. In the average the factor is $N/2x_w \approx 128/70 = 1.8$ within the processed data sets. Further on, the parallel laser lines are not strictly horizontal in the raw data. For a sufficient accurate Radon transformation of a raw image the discrete spacing has to be $\Delta\theta = 0.33°$ and $\Delta\rho = 0.7$ pixel after Toft (1996a). Within a kymogram, however, the laser lines are known to be horizontal. This a priori knowledge is used to limit the transformation range to a single discrete angle of $\theta = 90°$ and a discrete spacing of $\Delta\rho = 1$ pixel for the distances to the origin. Hence, the Radon transform of $2x_w$ kymogram images with a priori knowledge is by a factor of $\frac{N}{2x_w}\cdot\frac{1}{\Delta\rho}\cdot\frac{1}{\Delta\theta}\cdot 180 \approx \frac{128}{70}\cdot\frac{1}{0.7}\cdot\frac{1}{0.33}\cdot 180 = 1403$ less than a Radon transform of the whole 180° angle range on $N = 128$ single image frames.

The automatically measured values of vocal fold length, oscillation amplitude, and velocity agree with the values known from the literature (Schuberth et al., 2002; Schuster et al., 2005; Larsson and Hertegård, 2004; Schade et al., 2005). We measured a RMS acceleration of 1000 m/s^2. Titze et al. (2003) derived a RMS acceleration of 2256 m/s^2 from empirical rules for viscosity and vocal fold deformation, which is even higher than our measurements. Döllinger et al. (2005) measured a peak acceleration of an in vivo canine vocal fold to 620 m/s^2. The comparison with the literature values confirms the LLPS measured accelerations of the vocal fold edges. In a further study, a database with the metrical measurements of vocal fold vibrations will be built up for the quantitative comparison of normal and pathological vibration behavior.

9. Conclusion

The metrical measurement of vocal fold magnitudes in endoscopic high-speed recordings is enabled by a device that projects parallel laser line calibration marks onto vocal folds. The line characteristics of the calibration marks help to distinguish from common light disturbances in laryngoscopic imaging. The proposed line detection algorithm successively uses the temporal and spatial calibration information of the laser lines in the recorded image sequences. The algorithm is robust and automatically decides if the laser lines are on the vocal folds, as shown with real clinical endoscopic recordings. The laser line projector combined with the detection algorithm allows to derive metrical measurements of vocal fold dimensions and dynamics within the glottal plane of laryngoscopic recordings.

Acknowledgements

This work was supported by Grant from the Deutsche Forschungsgemeinschaft (DFG) Project DFG HO 2177/2-3 and from the Sonderforschungsbereich (SFB) 603 subproject B5.

References

Arndt, H.J., Schäfer, A., 1994. The width-length quotient of the glottis as a measure of amplitude values. Folia Phoniatr Logop 46, 265–270.

Beyerer, J., León, F., 2002. Die Radontransformation in der digitalen Bildverarbeitung. at – Automatisierungstechnik 50 (10), 472–480 <http://www.mrt.uni-karlsruhe.de/download/at0210_472.pdf>.

Bracewell, R.N., 1995. Two-dimensional Imaging. Prentice Hall, Englewood Cliffs, NY.

Chrastek, R., Wolf, M., Donath, K., Niemann, H., Paulus, D., Hothorn, T., Lausen, B., Lmmer, R., Mardin, C.Y., Michelson, G., 2005. Automated segmentation of the optic nerve head for diagnosis of glaucoma. Med. Image Anal. 9 (4), 297–314 <http://dx.doi.org/10.1016/j.media.2004.12.004>.

Deans, S.R., 1983. The Radon Transform and Some of Its Applications. John Wiley & Sons, New York.

Döllinger, M., Berry, D.A., Berke, G.S., 2005. A quantitative study of the medial surface dynamics of an in vivo canine vocal fold during phonation. Laryngoscope 115, 1233–1268.

Eysholdt, U., Rosanowski, F., Hoppe, U., 2003. Irregular vocal fold vibrations caused by different types of laryngeal asymmetry. Eur. Arch. Otorhinolaryngol. 260, 412–417.

Hand, D., Mannila, H., Smyth, P., 2001. Principles of Data Mining. Adaptive Computation and Machine Learning. MIT Press, Cambridge.

Hertegård, S., Björck, G., Manneberg, G., 1998. Using laser triangulation to measure vertical distance and displacement of laryngeal mucosa. Phonoscope 1, 179–185.

Herzon, G.D., Zealear, D.L., 1997. New laser ruler instrument for making measurements through an endoscope. Otolaryngol. Head Neck Surg. 116, 689–692.

Hoppe, U., Rosanowski, F., Döllinger, M., Lohscheller, J., Eysholdt, U., 2003. Visualization of the laryngeal motorics during a glissando. J. Voice 17, 370–376.

Illingworth, J., Kittler, J., 1988. A survey of the Hough transform. CVGIP 44, 87–116.

Jiang, J.J., Shah, A.G., Hess, M.M., Verdolini, K., Banzali Jr., F.M., Hanson, D.G., 2006. Vocal fold impact stress analysis. J. Voice 15, 4–14.

Larsson, H., Hertegård, S., 2004. Calibration of high-speed imaging by laser triangulation. Logoped Phoniatr. Vocol. 29, 154–161.

Leavers, V.F., 1992. Use of the Radon transform as a means of extracting symbolic representations of shape in two dimensions. Image Vision Comput. 10, 99–107.

Leavers, V.F., 1993. Which Hough transform? CVGIP 58, 250–264.

Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., Döllinger, M., 2007. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. Med. Image Anal. 11, 400–413.

Manneberg, G., Hertegård, S., Liljencrantz, J., 2001. Measurement of human vocal fold vibrations with laser triangulation. Opt. Eng. 40, 2041–2044.

Mergell, P., Titze, I.R., Herzel, H., 2000. Irregular vocal-fold vibration – high-speed observation and modeling. J. Acoust. Soc. Am. 108, 2996–3002.

Mierswa, I., Wurst, M., Klinkenberg, R., Schulz, M., Euler, T., 2006. YALE: rapid prototyping for complex data mining tasks. In: Proceedings of the International Conference Knowledge Discovery and Data Mining. ACM Press, New York, USA.

Neubauer, J., Mergell, P., Eysholdt, U., Herzel, H., 2001. Spatio-temporal analysis of irregular vocal fold oscillations: biphonation due to desynchronization of spatial modes. J. Acoust. Soc. Am. 110, 3179–3192.

Okazawa, S.H., Ebrahimi, R., Chuang, J., Rohling, R.N., Salcudean, S.E., 2006. Methods for segmenting curved needles in ultrasound images. Med. Image Anal. 10, 330–342.

Perlman, A.L., Alipour-Haghighi, F., 1988. Comparative study of the physiological properties of the vocalis and cricothyroid muscles. Acta Otolaryngol. 105, 372–378.

Rousseau, F., Hellier, P., Barillot, C., 2005. Confhusius: a robust and fully automatic calibration method for 3D freehand ultrasound. Med. Image Anal. 9, 25–38.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. John Wiley & Sons Inc.

Ruben, R.J., 2000. Redefining the survival of the fittest: communication disorders in the 21st century. Laryngoscope 110, 241–245.

Saadah, A.K., Galatsanos, N.P., Bless, D., Ramos, C.A., 1998. Deformation analysis of the vocal folds from videostroboscopic image sequences of the larynx. J. Acoust. Soc. Am. 103, 3627–3641.

Schade, G., Kirchhoff, T., Hess, M., 2005. Geschwindigkeitsmessung der Stimmlippenbewegung. Folia Phoniatr. Logop. 57, 202–215.

Schade, G., Leuwer, R., Kraas, M., Rassow, B., Hess, M., 2004. Laryngeal morphometry with a new laser 'clip on' device. Lasers Surg. Med. 34, 363–367.

Schuberth, S., Hoppe, U., Döllinger, M., Lohscheller, J., Eysholdt, U., 2002. High-precision measurement of the vocal fold length and vibratory amplitude. Laryngoscope 112, 1043–1049.

Schuster, M., Lohscheller, J., Kummer, P., Eysholdt, U., Hoppe, U., 2005. Laser projection in high-speed glottography for high-precision measurements of laryngeal dimensions and dynamics. Eur. Arch. Otorhinolaryngol. 262, 477–481.

Titze, I.R., 1994. Mechanical stress in phonation. J. Voice 8, 99–105.

Titze, I.R., Svec, J.G., Popolo, P.S., 2003. Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. J. Speech Lang. Hear Res. 46, 919–932.

Toft, P., 1996a. The Radon transform – theory and implementation. Ph.D. Thesis, Department of Mathematical Modelling, Section for Digital Signal Processing, Technical University of Denmark, Lyngby, Denmark, <http://pto.linux.dk/PhD/>.

Toft, P.A., 1996b. Detection of lines with wiggles using the Radon transform. In: Proceedings of the NORSIG, pp. 267–270.

Whitaker, R.T., Elangovan, V., 2002. A direct approach to estimating surfaces in tomographic data. Med. Image Anal. 6, 235–249.

Wittenberg, T., Moser, M., Tigges, M., Eysholdt, U., 1995. Recording, processing, and analysis of digital high-speed sequences in glottography. Mach. Vis. Appl. 8 (6), 399–404.

Yan, Y., Chen, X., Bless, D., 2006. Automatic tracing of vocal-fold motion from high-speed digital images. IEEE Trans. Biomed. Eng. 53, 1394–1400.