Three-Dimensional Optical Reconstruction of Vocal Fold Kinematics Using High-Speed Video With a Laser Projection System

Georg Luegmair*, Daryush D. Mehta, James B. Kobler, and Michael Döllinger

***Abstract*—Vocal fold kinematics and its interaction with aerodynamic characteristics play a primary role in acoustic sound production of the human voice. Investigating the temporal details of these kinematics using high-speed videoendoscopic imaging techniques has proven challenging in part due to the limitations of quantifying complex vocal fold vibratory behavior using only two spatial dimensions. Thus, we propose an optical method of reconstructing the superior vocal fold surface in three spatial dimensions using a high-speed video camera and laser projection system. Using stereo-triangulation principles, we extend the camera-laser projector method and present an efficient image processing workflow to generate the three-dimensional vocal fold surfaces during phonation captured at 4000 frames per second. Initial results are provided for airflow-driven vibration of an ex vivo vocal fold model in which at least 75% of visible laser points contributed to the reconstructed surface. The method captures the vertical motion of the vocal folds at a high accuracy to allow for the computation of three-dimensional mucosal wave features such as vibratory amplitude, velocity, and asymmetry.**

***Index Terms*—Distance measurement, image segmentation, stereo vision, surface reconstruction, vocal folds.**

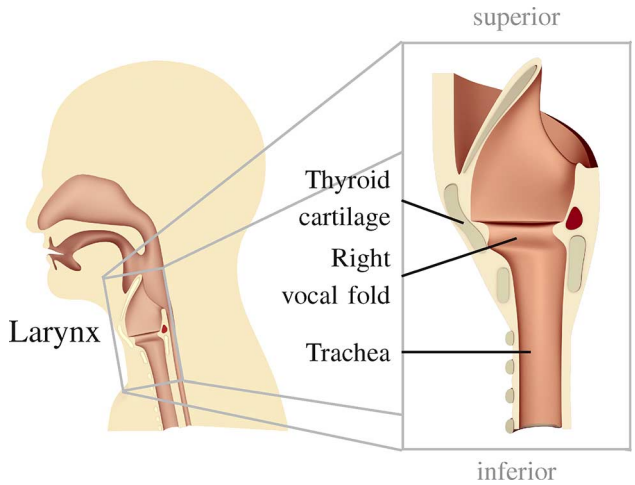

Fig. 1. Left: Cross-section through the human head and neck along the midsagittal plane. Right: Close up of the human larynx in the midsagittal plane showing the position of the right vocal fold.

I. Introduction

VERBAL communication is of ever-increasing importance in our society, and many voice disorders may significantly hamper the ability of workers to function in the workplace or even contribute to absences from work. It has been shown that successfully treating voice disorders might be more economical in the long term than leaving the condition untreated, allowing those affected to contribute to economic productivity [1]. The oscillation patterns of the vocal folds are key to the produced quality of the voice and are governed by tissue-acoustic interactions with the exhaled airflow from the lungs [2]. Of primary importance to acoustic voice production, the medial-lateral movements of the vocal folds open and close the laryngeal aperture (glottis) to modulate and interact with the airflow from the lungs to produce pressure fluctuations [3]. Therefore, characterizing vocal fold tissue properties and dynamic vibratory patterns is a major focus of voice science and clinical voice assessment.

Fig. 1 illustrates the position of the larynx—in particular, the vocal folds—in a cross-sectional region of the head and neck. Voice specialists make critical diagnostic, medical, therapeutic, and surgical decisions based on coupling visual observations of vocal fold tissue motion with auditory-perceptual judgments of voice quality [4]. Objective clinical assessment of vocal functions is typically performed using endoscopic imaging of vocal fold vibration and acoustic measurements of voice quality [5]. During office-based clinical procedures, videostroboscopy is typically used to provide the most direct observation of how pathology alters vocal fold vibration given a periodic and stable

Manuscript received January 09, 2015; revised June 09, 2015; accepted June 10, 2015. Date of publication June 16, 2015; date of current version November 25, 2015. M. Döllinger's contribution was supported by Deutsche Krebshilfe e.V. Grant 111332. Research supported by an American Speech-Language-Hearing Foundation Speech Science Grant, by the NIH National Institute on Deafness and Other Communication Disorders (R01 DC007640), and by the Voice Health Institute. The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. *Asterisk indicates corresponding author.*

*G. Luegmair is with the Speech Production Laboratory at University of California, Los Angeles, CA 90095 USA (e-mail: georg.luegmair@gmail.com).

D. D. Mehta is with the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114 USA, and with the Department of Surgery, Harvard Medical School, Boston, MA 02115 USA, and also with the Department of Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA 02129 USA (e-mail: mehta.daryush@mgh.harvard.edu).

J. B. Kobler is with the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114 USA, and also with the Department of Surgery, Harvard Medical School, Boston, MA 02115 USA (e-mail: james.kobler@mgh.harvard.edu).

M. Döllinger is with the Department of Phoniatrics and Pedaudiology, University Hospital Erlangen, 91054 Erlangen, Germany.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2015.2445921

vocal fold vibration [6]. Structural and vibratory information acquired from videostroboscopy is heavily relied upon for diagnosis, surgical planning, and assessment of surgical outcome, as well as other treatment procedures.

High-speed videoendoscopy (HSV) allows for enhanced temporal resolution to be able to visualize intra-cycle vocal fold tissue motion critical for the assessment for normal and pathological voice conditions in which vocal fold vibrations may not be periodic (thus, precluding the use of stroboscopic techniques that rely on periodicity) [7]. The fundamental frequency of vocal fold vibration during conversational speech ranges from 80–150 Hz in adult males to 150–300 Hz in adult females. The actual clinical utility of HSV, however, is limited by relatively low or moderate correlations between measures of irregularity in vocal fold vibration and acoustic parameters [8]–[10], providing motivation for continued efforts to account for more of the unexplained variance between acoustic and HSV-based measures. Our working hypothesis is that more salient relationships between HSV-based measures of vocal fold physiology and acoustic/aerodynamic measures of sound production will be obtained by quantifying the complex three-dimensional (3D) motion of the vocal fold tissue during phonation [11].

Several attempts have been made to capture the vertical (superior-inferior) spatial dimension during laryngeal videoendoscopy. Software-based processing of two-dimensional (2D) HSV images has attempted to estimate 3D information from pixel intensity [7]; although a potentially useful approach, this approach suffers severely from illumination inconsistencies that would be alleviated through more direct, hardware-based technology. Ultrasound imaging has been applied with insufficient spatial resolution [12]. A few setups have been devised as variants of the structured-light method. In general, these methods have relied on a camera-laser setup for stereo triangulation, e.g., with a laser line [13], [14]. The derived depth-kymography is limited, however, by spatial uncertainty (± 50 μm) and clinical interpretability of motion of a single coronal position [13]. Optical coherence tomography has shown promise in the clinical setting but current technology limits temporal resolution and spatial scanning range [15]. Lately, the method of stereo triangulation by feature matching has been proposed for this purpose [16]. However, the vocal folds lack distinct feature points, creating high uncertainty in the reconstruction process. We have chosen to apply optical reconstruction using structured illumination due to its applicability at fast video sampling rates (4000 frames per second).

Section II presents the steps for computation and interpolation of the vocal fold surface data. Sections II.D and III illustrate the application and results, respectively, of the three-dimensional reconstruction technique in an ex vivo larynx model. Section IV concludes with a discussion of the proposed method in relation to alternative visualization techniques for clinical voice assessment.

II. Methods

In this section, the developed method for 3D reconstruction of the superior surface of the vocal folds during phonation is described in a step-by-step manner. Empirical data illustrate the utility of the method in an ex vivo larynx experiment that allowed for the simultaneous capture of a laser projection system and high-speed videomicroscopy (HSVM) data. Parameters salient to voice production are then derived from the time-varying 3D vocal fold surfaces. Table IV defines the variables that will be used in this paper.

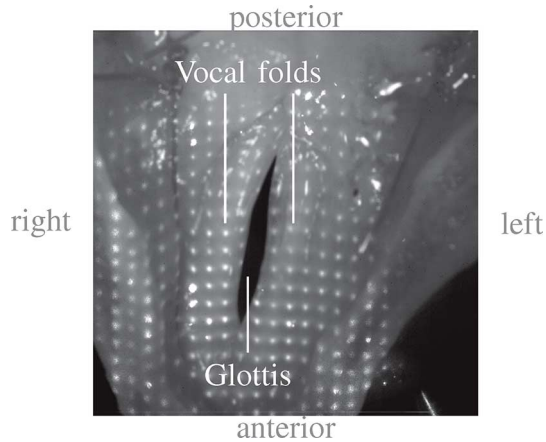

Fig. 2. Exemplary frame taken from a high-speed video recording of the oscillating vocal folds in an ex vivo experimental setup. The top-down perspective of the vocal folds and the space between them (glottis) mimics that of the endoscopic view where the anterior direction is toward the bottom of the frame.

A. Experimental Setup

Cadaver excised larynges acted as a physiological model of vocal fold tissue morphology. The larynges were stored in saline-soaked gauze in an 80°C freezer. After thawing, whole-mount preparations of the excised larynges were prepared by dissecting away supraglottal structures (hyoid, etc.), suturing the ventricular folds, and leveling the thyroid cartilage to provide for full exposure of the true vocal folds superiorly. Inferiorly, the trachea was cut to a length of approximately 5 cm and the specimen was mounted on a cylindrical pipe connected to an airflow supply.

Airflow was sent through a ConchaTherm-IV device (Hudson RCI, Research Triangle Park, NC) that warmed and humidified the air before directing the stream superiorly through the trachea to produce self-sustained vocal fold oscillations. A pneumatic pressure gauge regulated the force of the air stream.

Fig. 2 shows an exemplary HSVM snapshot with a structured pattern of laser light during phonation of the ex vivo human larynx model. A magnified view of the superior aspect of the larynx was provided by a Leica F40 surgical microscope with an integrated 300-W xenon light source that provided sufficient illumination for the HSVM recordings. Video recordings were acquired with a monochromatic high-speed video camera (Phantom v7.1; Vision Research, Inc., Wayne, NJ) at 4000 images per second. Spatial resolution was set to 608×600 pixels.

B. 3D Reconstruction of the Vocal Fold Surface

The laser projection system displayed a grid pattern on the superior laryngeal surface to track the three-dimensional deformation of the vocal folds across video frames, as described in a

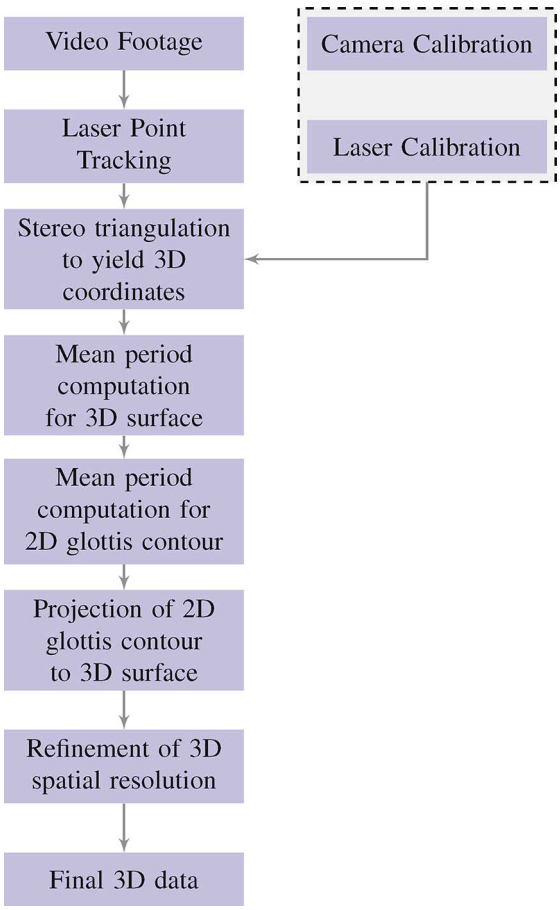

Fig. 3. Flowchart describing the 3D reconstruction process. The elements in the dashed box are part of the system calibration that are described in a prior publication [17].

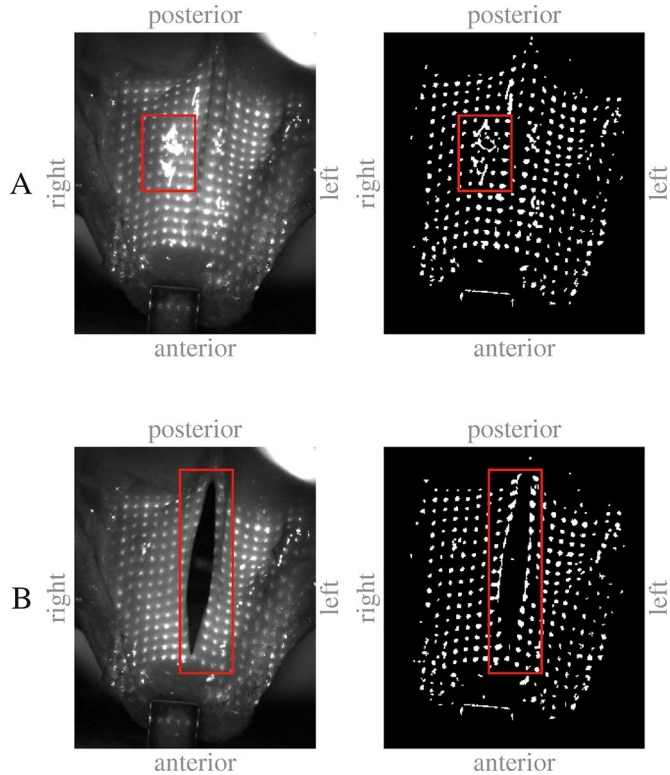

Fig. 4. Illustration of two challenges during the assignment of projected laser point indices: (A) specular reflections superposing multiple points and (B) interruption of the tracking of points within the glottal area. The areas of interest are highlighted by a bounding box on the original video frame before (left) and after (right) image processing.

previous publication [17]. The system is capable of a spatial resolution accurate to 15 μm. To further enhance this system, we implemented an improved computational process for the derivation of the vocal fold surface, which incorporated vocal fold edge information from the corresponding optical HSVM image sequence. This enhancement was important to segment the 3D surface according to the glottis contour to enable the derivation of salient features of phonatory function. Fig. 4 displays a flowchart of the 3D reconstruction steps detailed in the following sections.

1) Laser Point Tracking: Each laser ray intersected with an associated location on the surface of the vocal folds. The 3D reconstruction method necessitated the assignment of each laser point to its originating laser ray over time (i.e., from frame to frame). Assignments were maintained throughout the recording.

The time-varying trajectory of each laser point was derived using a combination of image processing techniques and user interaction. Laser points were extracted within each frame by first filtering gradual background lighting or diffuse scatter light with a homomorphic filter, where the laser points as brightest elements remained [18]. Then the image was top-hat filtered and binarized with a threshold calculated by Otsu's method [19], creating a binary image containing the laser points.

The movement of each laser point $\mathbf{m}$ was tracked, beginning at an initial frame n_{I}, where each point was assigned manually to its position in the laser grid. To track the movement of a point from frame $n-1$ to frame n, it proved to be sufficient to perform a constrained nearest-neighbor search in an area around $\mathbf{m}^{\mathrm{i,j}}(n-1)$ to obtain a point $\mathbf{m}^{\mathrm{i,j}}(n)$. The point assignment in subsequent frames was valid if

$$\|\Delta\mathbf{m}^{\mathrm{i,j}}(n)\|_2 < \gamma, \tag{1}$$

where

$$\Delta\mathbf{m}^{\mathrm{i,j}}(n) = \mathbf{m}^{\mathrm{i,j}}(n) - \mathbf{m}^{\mathrm{i,j}}(n-1); \tag{2}$$

i.e., the distance ($\Delta\mathbf{m}^{\mathrm{i,j}}$) traveled by a given laser point was constrained to lie below a threshold γ due to the fact that the velocity of a moving mass could not change arbitrarily due to inertia.

The frame-to-frame distance threshold γ was chosen to be $\gamma = 0.1 \cdot (\bar{d} - 3\sigma_d)$, where $\bar{d}$ and σ_d are the mean and standard deviation, respectively, of the gridpoint-to-gridpoint distance d.

Furthermore, the rate of change from frame $n-1$ to frame n (acceleration of the mass) was constrained through the following inequality:

$$\|\Delta\mathbf{m}^{\mathrm{i,j}}(n) - \Delta\mathbf{m}^{\mathrm{i,j}}(n-1)\|_2 < \varphi, \tag{3}$$

where $\varphi = 1.5 \cdot \Delta\mathbf{m}^{\mathrm{i,j}}(n)$.

In (1) and (3), γ and φ, respectively, act as regulation parameters that allow for the trajectory to vary in direction and velocity depending on the history of $\mathbf{m}^{\mathrm{i,j}}(n-1)$. Choosing the parameters based on gridpoint distance enabled flexibility in the choice of the nearest-neighbor search region.

As a result, the parameters are dependent on the frame rate of the camera. The γ and φ parameters are chosen empirically for a frame rate of 4000 frames per second to achieve stable tracking. This frame rate is typical for clinical HSV and represents a balance among frame rate, spatial resolution, and image quality.

Fig. 4 illustrates two issues in tracking a laser point from frame to frame. First, highlights or specular reflections may superpose multiple laser points—i.e., the laser points would be indistinguishable from each other (see Fig. 4(A)). Second, the oscillation of the vocal folds (opening and closing of the glottis) removes and adds points in the image (see Fig. 4(B)). Both issues prevent a continuous tracking of particular points. A previously undetected point in the image must be reassigned to a trajectory for the tracking to function properly.

To this end, point $\mathbf{m}$ was integrated into the grid of laser points, i.e., the assignment to its corresponding laser ray. This was achieved by evaluating the following criteria sequentially:

a) Deviation $\Delta\mathbf{m}^{\mathrm{i,j}}$

$$\Delta\mathbf{m}^{\mathrm{i,j}} = \mathbf{m}(n) - \mathbf{m}^{\mathrm{i,j}}(n-T), \tag{4}$$

with T as the period length. If

$$\|\Delta\mathbf{m}^{\mathrm{i,j}}\|_2 < \gamma, \tag{5}$$

assign $\mathbf{m}(n)$ to the trajectory $t^{i,j}$.

b) Probability $p(\mathbf{m})$, determined by the distance to neighboring grid points, with

$$p(\mathbf{m}) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}(\|\mathbf{m}-\mathbf{m}_{\mathrm{m}}\|_2)^2} \tag{6}$$

with $\mathbf{m}_{\mathrm{m}}$ as the center-of-mass between the four neighbors to the left, to the right, to the top and to the bottom, with respect to image orientation. Thus, the probability of the location $\mathbf{m}$ belonging to a grid position i, j is evaluated by a 2D Gauss curve centered around $\mathbf{m}_{\mathrm{m}}$. The trajectory $t^{i,j}$ of maximum probability is chosen.

Ideally, the trajectories of the laser points were lines, which were determined by the orientation of the laser ray. However, the identification of a point's center-of-mass and, subsequently, the determination of its position $\mathbf{m}$, were dependent on the shape of the thresholded element. Thus, several parameters of the image processing, e.g., the top-hat filtering or the threshold value in the segmentation process, added to the uncertainty in determining the point's position. As a consequence, noise was added to each trajectory $t^{i,j}$.

Then, a principle component analysis (PCA) was applied to minimize the influence of the added uncertainty. Points in the image moved along a straight line due to the projection of the laser rays. Thus it was a valid assumption that the majority of the points' movements in the 2D domain of the image was captured by a single eigenvector. The movement was decomposed into two eigenmodes, where the first eigenmode described the actual movement of the point, and the second eigenmode described the noise in an out-of-trajectory direction. By eliminating the energy of the second eigenmode, noise in the trajectory was significantly reduced through directional selectivity.

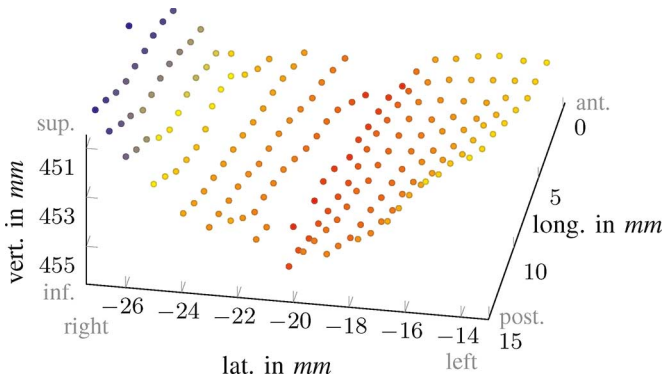

Fig. 5. Optical reconstruction of 150 out of 328 laser points on the laryngeal surface in a 3D coordinate system. At this stage of the reconstruction, the points form a continuous surface covering the vocal folds and glottis (vocal fold edges are not modeled). The points are color coded such that points that are more superior are coded in colder colors.

2) Stereo Triangulation to Yield 3D Coordinates: Optical reconstruction in 3D was based on the principle of stereo triangulation. The calibration process for the camera and laser projection system is detailed in a prior publication [17]. Briefly, a point $\mathbf{m}_{\mathrm{w}} \in \mathbb{R}^3$ was reconstructed, i.e., its 3D coordinates were computed from two independent central projections: $\mathbf{m}_1 \in \mathbb{R}^2$ onto the image I_1 and $\mathbf{m}_2 \in \mathbb{R}^2$ onto the image I_2 with known transformation between the two images:

$$\mathbf{m}'_1 = \mathbf{R}\,\mathbf{m}'_2 + \mathbf{t}. \tag{7}$$

This equation was solved for z_i of $\mathbf{m}'_{\mathrm{i}}$, with $\mathbf{m}'_{\mathrm{i}} = [x_i\ y_i\ z_i]$.

Fig. 5 shows the laser points of an exemplary video frame in a 3D coordinate system. For this system an array of 20×20 points was projected with a 532 nm laser. The laser points toward the left and right posterior end were higher than other points of the surface due to the underlying arytenoid cartilages at the posterior end of the vocal folds. The "valley" along the anterior-posterior medial line indicated the location of the glottis.

3) Mean Period Computation for 3D Surface: Even though the trajectories of the segmented laser points were filtered by PCA, noise persisted in the first eigenvector of the trajectories $t^{i,j}$ and consequently in the reconstructed 3D points $\mathbf{m}^{\mathrm{i,j}}_{\mathrm{w}}$. This noise had the potential to yield physiologically implausible values when extracting parameters such as vocal fold tissue amplitude and velocity. To reduce the effect of these errors, a mean period was calculated.

Fig. 6 illustrates the derivation of period information from a basic analysis of the glottal area waveform (GAW) from the 2D HSVM frames. Periods were defined as the time span between maxima of the GAW. The image points $\mathbf{m}$ were combined to a set C for each frame. For this purpose, the glottis is segmented from the video footage with a region growing algorithm [20]. The time-dependent area signal derived from C yielded the GAW and therefrom the individual period lengths T_{p}, the

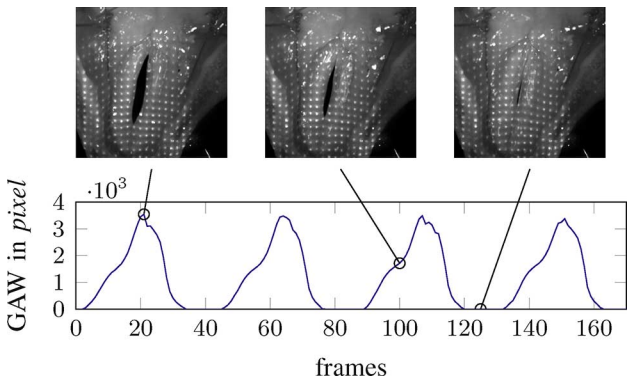

Fig. 6. Exemplary glottal area waveform (GAW) indicating oscillatory behavior due to pseudo-periodic vibration of the vocal fold tissue. Video frames are displayed for three different phases of the GAW.

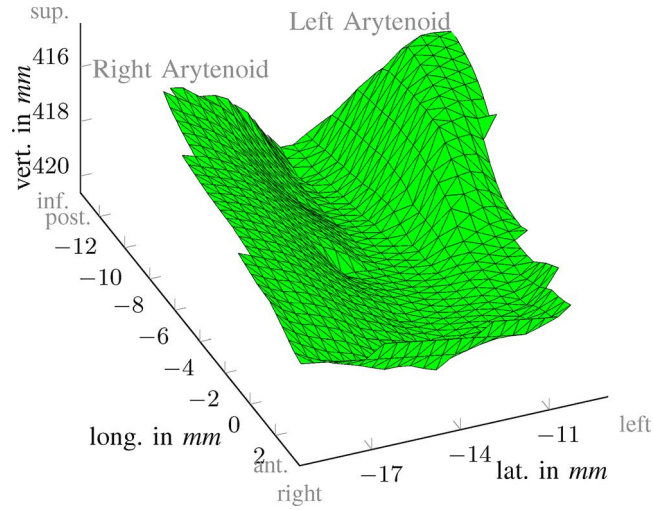

Fig. 7. Coarse interpolation across the vocal fold surface and glottis. Steep inclination toward the left and right arytenoids is visible.

number of frames $N_{\mathrm{f,p}}$ in each period n_{p}, and the number of periods N_{p} [21], [22].

The mean period was computed by normalizing the 3D coordinates within each period of vocal fold vibration to a common time base, i.e., to a common target period length T consisting of $N_{\mathrm{f,T}}$ frames. The 3D points periods for which $N_{\mathrm{f,p}} \neq N_{\mathrm{f,T}}$ or $T_{\mathrm{p}} \neq T$ were interpolated to the common time base. For every frame of the mean period, mean values $\bar{\mathbf{m}}_{\mathrm{w}}^{\mathrm{i,j}}$ were calculated over all periods.

The period lengths were normalized to $T = 1$ and the target time points of frames in the target time frame were designated in the interval

$$t_{\mathrm{T,j}} \in [0 \ldots 1], \tag{8}$$

with $j = 1 \ldots N_{\mathrm{f,T}}$ compiled to the target vector $\mathbf{t}_{\mathrm{T}}$. Likewise, the time points in the source were designated

$$t_{\mathrm{S,k}} \in [0 \ldots 1], \tag{9}$$

with $k = 1 \ldots N_{\mathrm{f,p}}$ compiled to the source vector $\mathbf{t}_{\mathrm{S}}$.

The coordinates of $\mathbf{m}_{\mathrm{w}} \in \mathbb{R}^3$ or $\mathbf{m} \in \mathbb{R}^2$ at the target time points of $\mathbf{t}_{\mathrm{T}}$ were computed by interpolating the separate dimensions linearly. Mean data $\overline{\mathbf{m}}_{\mathrm{w}}^{i,j}$ were then created by averaging over the data available for element of $\mathbf{t}_{\mathrm{T}}$.

Fig. 7 illustrates the "closed surface" generated by linear interpolation of the reconstructed points $\bar{\mathbf{m}}_{\mathrm{w}}^{i,j}$ across the glottis and vocal fold surfaces in one frame.

4) Mean Period Computation for 2D Glottis Contour: The mean period computation for the 2D glottis contour was more complex, as not a defined point but a shape was interpolated over time. For periods with $T_{\mathrm{p}} \neq T$, the glottis contour $C \subset \mathbb{R}^2$ had to be re-interpolated. The new contour of $C_{\mathrm{T}} \subset \mathbb{R}^2$ was found by interpolating between the closest neighbors in time: $t_{\mathrm{S,p}}$ previous to and $t_{\mathrm{S,a}}$ after each element $t_{\mathrm{T,i}} \in \mathbf{t}_{\mathrm{T}}$. It was assumed that the contour scales linearly between these two time points using temporal scaling factor s_{GC}:

$$s_{\mathrm{GC}} = \frac{t_{\mathrm{T,i}} - t_{\mathrm{S,p}}}{t_{\mathrm{S,a}} - t_{\mathrm{S,p}}}. \tag{10}$$

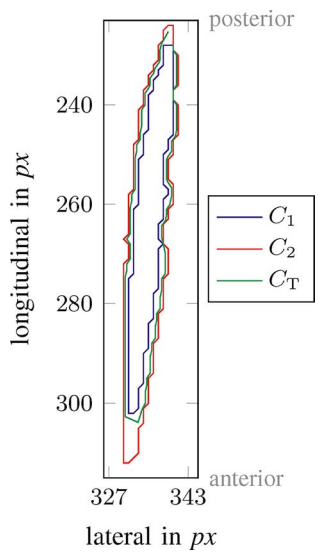

Fig. 8. Glottis contour interpolation C_{T} (green) using scaling factor $s_{\mathrm{GC}} = 0.68$ between two glottis contours C_1 (blue) and C_2 (red). For clarity, the normals for the interpolation are not depicted.

Scaling factor s_{GC} was applied by projecting the contour C_1 containing less image points along its normals toward the contour C_2 with more image points. In case that $|C(t_{\mathrm{s,p}})| \neq |C_1|$, the scaling had to be reversed. Thus

$$s_{\mathrm{GC}} = 1 - s_{\mathrm{GC}}. \tag{11}$$

Fig. 8 shows an example interpolated glottis contour. A contour C_{T} (green) was interpolated for a value $s_{\mathrm{GC}} = 0.68$ between the smaller contour C_1 (blue) and the larger contour C_2 (red) (i.e., $|C_1| < |C_2|$). The glottis contours were not smooth due to the resolution of the video frame. Subsequently, the scaling in areas with great curvature (e.g., at the lower end of the contour) was less accurate. This step thus created a set of $N_{\mathrm{f,T}} \cdot N_{\mathrm{p}}$ contours, with N_{p} contours for every frame $n_{\mathrm{f,T}}$.

Fig. 9 shows an example mean glottis contour with further smoothing. The smoothed glottis contour $\overline{C}_{\mathrm{f,T}}$ was obtained by

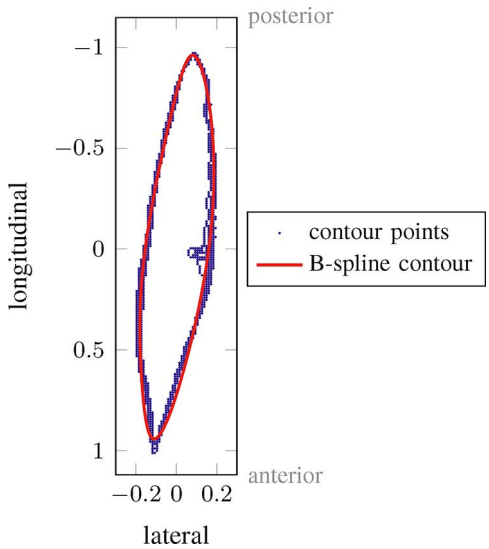

Fig. 9. Approximation of a mean glottis contour for frame k of the mean period. For this purpose, a closed B-spline curve (red) is fitted to the point cloud (blue points) formed by the set of N_p contours belonging to the time step $n_{\mathrm{f,T}}$. Values for x and y directions are normalized to zero mean.

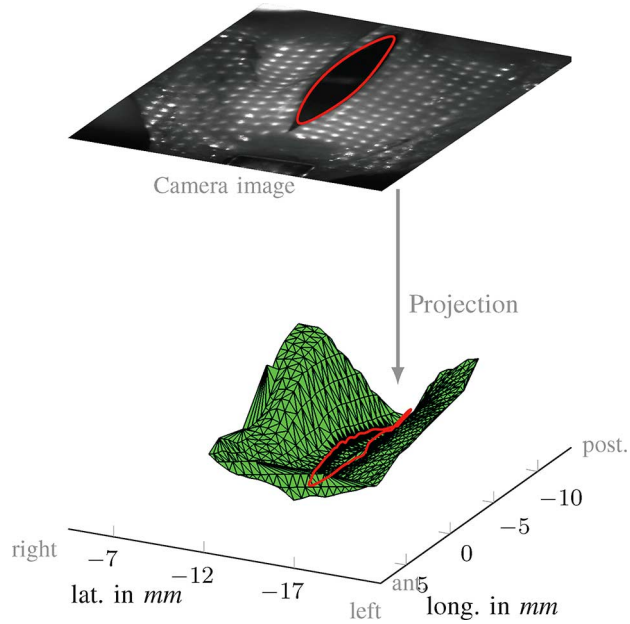

Fig. 10. Projection of the 2D glottis contour extracted from the high-speed video frame onto the 3D surface with coarse interpolation.

optimizing a closed B-spline curve over the N_p single contours for frame $n_{\mathrm{f,T}}$. A B-spline curve of the degree n with m knots was defined as

$$f(s) = \sum_{i=0}^{m-n-2} \mathbf{m}_i b_{i,n}(s), \quad s \in [s_n, s_{m-n-1}], \tag{12}$$

with parameter s, basic B-spline functions $b_{i,n}$ and control points $\mathbf{m}_i$. A uniform B-spline function of order 4 with eight control points proved to be sufficient. The curve was optimized with the following target function:

$$\min_{\mathbf{m}_i \in \mathbb{R}^2} \sum_{i=1}^{N} \|d(f, \mathbf{m}_i)\|_2^2, \tag{13}$$

where d is the Euclidean distance between the curve f and a point $\mathbf{m}_i$, and N is the cumulative number of contour points from N_p contours at frame $n_{\mathrm{f,T}}$. A benefit of this step was the smoothing of the previously coarse glottis contour.

5) Projection of 2D Glottis Contour to 3D Surface: Fig. 10 illustrates the projection of the mean glottis contour $\bar{C}_k$ onto the 3D closed surface to identify the glottal area between the vocal folds. The captured image using the high-speed video camera is shown with the segmented glottis contour. The contour is projected by the pinhole model of the camera onto the 3D closed surface. The projection of the image point $\mathbf{m} \in \bar{C}_k$ of the glottis contour was computed by finding the intersection of the projection ray with the surface.

6) Refinement of 3D Spatial Resolution: A high spatial resolution is required to quantify vocal fold kinematics, especially the region medial to the last data point $\mathbf{m}_\mathrm{w}^{\mathrm{i,j}}$ where vocal fold surface curves strongly to form the vocal fold edge. A finer, regular base grid allowed for the approximation of a curve with defined end conditions using bi-cubic spline function

$$f(u,v) = \sum_{i=0}^{m} \sum_{j=0}^{n} c_{i,3}(u) c_{j,3}(v) \mathbf{m}_\mathrm{w}^{i,j}, \tag{14}$$

where $f(u,v)$ is an arbitrary point on the vocal fold surface, $c_{i,3}(u)$ and $c_{j,3}(v)$ are cubic spline basis functions, and $\mathbf{m}_\mathrm{w}^{i,j}$ builds the grid of $m \times n$ control points. Cubic splines were chosen following other studies that interpolated tissue surfaces [23]–[25]. The spline fit is achieved by an optimization procedure that varies the control points with a least-squares distance of every point as target function.

Then, the glottis contour was added to the finely interpolated 3D surface. The base grid was sliced in longitudinal direction, in which each slice indicated a cross-section of the surface with a plane of constant y-values and expanding in z-direction, as seen in Fig. 11. Each slice was checked for an intersection with the glottis contour. If an intersection were found, the surface was "opened" in the area of the glottis and curved to create a vocal fold edge. The intersections of the slice with the glottis contour $\mathbf{m}_{\mathrm{I,l}}$ and $\mathbf{m}_{\mathrm{I,r}}$ and the normal directions of the glottis contour $\mathbf{n}_{\mathrm{I,l}}$ and $\mathbf{n}_{\mathrm{I,r}}$ were calculated. Left or right laterality is omitted.

As the curvature at the vocal fold edge cannot be directly measured from the video frame, the surface was assumed to be approximately tangential to the projection line of the glottis contour based on experimental results [26], [27]. A similar premise has been used for glottal area segmentation from 2D high-speed video [20].

The following computations were limited to the intersecting plane with a constant y-value. Thus, the number of dimensions that needed to be taken into account was reduced to two. With knowledge of the tangent for the bending, the vocal fold edge was assumed to behave as a quadratic Bézier curve

$$f(s) = \sum_{i=0}^{2} \binom{2}{i} s^i (1-s)^{2-i} \mathbf{m}_i, \tag{15}$$

where $\mathbf{m}_i$ indicates the ith control point of the curve and parameter $s \in [0,1]$ for the interval between the last data point

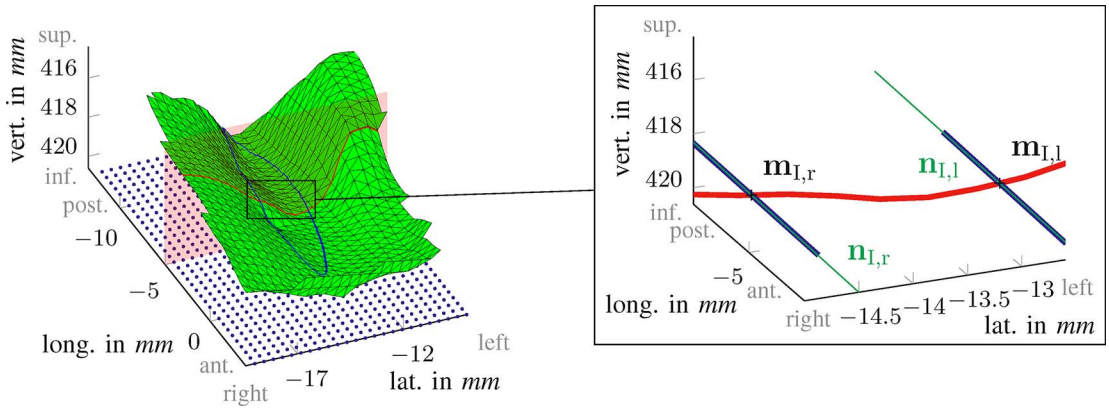

Fig. 11. Refinement of the vocal fold surface (green). Blue points indicate the base grid for interpolation. The red plane indicates an exemplary slice, the red line the intersection of the slice with the surface. The blue line is the projected glottis contour. The image on the right is the zoomed in area where slice and glottis contour intersect.

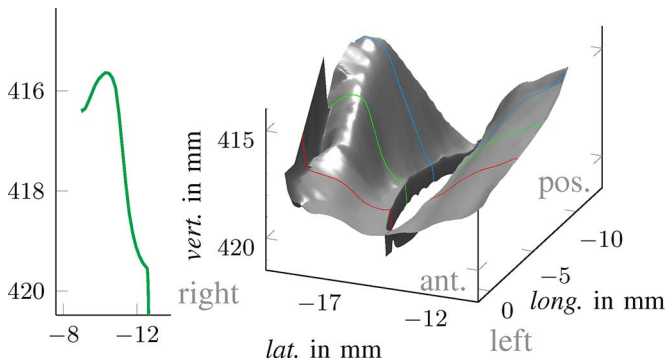

Fig. 12. Left: An exemplary section of the reconstructed surface of the right vocal fold. Right: Sections are positioned at 25% (red), 50% (green) and 75% (blue) of the glottal midline from the anterior glottis end for the analysis.

and the vocal fold edge. This allowed for a control of the first derivative of the curve in the transition points.

The curvature toward the vocal fold edge was governed by the position of three control points: (1) the point closest to $\mathbf{m}_{\mathrm{w}}$; (2) the intersection of the surface tangent in $\mathbf{m}_{\mathrm{w}}$ with the projection of the glottis contour, thus creating a G^1-continuous transition from the cubic spline of the superior surface to the Bézier curve; and (3) the point along the projection line $\mathbf{l}$ with $\|l\|_2 = 1$ (currently, this endpoint is fixed at $\mathbf{m}_i + \mathbf{l}$). To create a vocal fold edge that was not only laterally (i.e., along the slice) but also longitudinally continuous, the three control points of every Bézier curve were combined to create a Bézier surface.

Finally, the bi-cubic interpolated 3D vocal fold surface and the Bézier surface of the vocal fold edges were merged. An example of the merged result is shown in Fig. 12. Compared to Figs. 5 and 7, the 3D surface was interpolated and included the glottal contour and vocal fold edges.

C. Parameterization of 3D Vocal Fold Vibration

Vocal fold kinematic parameters were defined using the 3D vocal fold surface, including amplitudes and velocities of the vocal fold edge. The vocal fold edge separates the superior from the medial vocal fold surface. Common 2D video—based analysis methods define the vocal fold edge as the most medial visible point [20]. Measures derived from these points have proven to yield statistically significant correlations between vocal fold kinematics and voice pathology [21], [28]. For the 3D case, the chosen endpoint of the Bézier surface was defined as the vocal fold edge.

Fig. 12 depicts the orientation of three cross-sections through the vocal folds perpendicular to the glottal midline at 25%, 50%, and 75% of the midline length from the anterior glottis end. Amplitudes of oscillation are derived from the trajectories of vocal fold edges at these three cross-sections. Velocities are then computed by differentiating the amplitudes with respect to time.

A set of 3D-derived parameters was computed and correlated to the degree of asymmetry. The set was chosen to include mean and standard deviations of the amplitude A and velocity v for each of the three cross-sections on the left and right vocal fold. Maximum and mean velocity were computed from the entire period, from the opening phase, and from the closing phase. Additionally, the mean deviation of the surface and glottis contour of each period to the averaged period was computed.

D. Application to Ex Vivo Larynx Experiments

Two cadaver larynges (L1 and L2) were investigated in the ex vivo experimental setup. Two manipulators were used to induce independently varying levels of torque to the left and right arytenoids [29]. Control over torque loads allowed for different tension and vocal fold adduction scenarios.

Tables V and VI list the load scenarios and subglottal pressures applied to both larynges. In the case of L1, a one-sided asymmetry was simulated. For L2, the asymmetry was alternated between the left and right side. The different load cases were quantified by the asymmetry quotient

$$A = \frac{D_{\mathrm{R}} - D_{\mathrm{L}}}{D_{\mathrm{L}} + D_{\mathrm{R}}} \cdot 100\%. \tag{16}$$

where D_{R} and D_{L} are the torques applied to the right and left arytenoids, respectively. Symmetric conditions yielded a value of 0%. A negative (positive) asymmetry value indicated a higher (lower) torque on the left arytenoid. Values for vocal fold vibratory asymmetry ranged from −66% to 66%. L1 displayed complete glottal closure for all load scenarios, whereas L2 displayed incomplete glottal closure for load scenarios of higher

TABLE I
DETECTION SUCCESS RATE (PERCENTAGE OF DETECTED POINTS) AND TRACKING ERRORS FOR 10 HIGH-SPEED RECORDINGS OF LARYNX L1. σ_D (IN PERCENTAGE POINTS, PP) AND σ_T (IN PIXELS, PX) INDICATE THE STANDARD DEVIATIONS IN RELATION TO THE GOLD STANDARD AND ACROSS THE FRAMES

No.	mean det. (in %)	σ_D (in pp)	σ_T (in px)
1	76.22	5.32	1.31
2	68.35	3.30	1.48
3	76.39	3.16	1.14
4	78.79	3.62	1.06
5	66.12	5.01	1.15
6	77.16	4.07	1.19
7	74.38	3.59	1.24
8	75.33	4.46	1.17
9	73.60	4.21	1.16
10	75.53	3.49	1.19
mean	74.19	4.02	1.21

TABLE II
DETECTION SUCCESS RATE (PERCENTAGE OF DETECTED POINTS) AND TRACKING ERRORS FOR 10 HIGH-SPEED RECORDINGS OF LARYNX L2

No.	mean det. (in %)	σ_D (in pp)	σ_T (in px)
1	87.45	2.71	0.73
2	86.83	3.95	0.83
3	85.70	3.70	0.84
4	87.28	4.25	0.88
5	86.68	4.91	0.84
6	85.54	4.54	0.82
7	84.98	4.31	1.00
8	80.62	4.98	0.96
9	73.16	4.31	0.98
10	82.36	4.01	1.02
mean	84.06	4.17	0.89

asymmetry. Glottal closure during phonation is associated with healthy voices, whereas incomplete glottal closure is often related to inefficient voice production.

The accuracy and efficiency of the laser point tracking algorithm was determined by the ability to identify a laser point in the image and add it to a trajectory. The success rate of the detection and the deviation of the tracked trajectory were related to a gold standard set by human includegraphics for a selection of 20 high-speed recordings (10 from L1 and 10 from L2) of 150 frames each.

III. RESULTS

Tables I and II list the laser point tracking detection rate for L1 and L2, respectively. The tables contain values for the mean and standard deviation σ_D over all frames of a recording, and the standard deviation σ_T of the tracked point trajectories to the reference. The image segmentation detected approximately 75% and 84% of the points for L1 and L2, respectively. For at least one trial for each of the larynges, the detection rate was around

TABLE III
COMPUTATIONAL COSTS FOR 3 EXEMPLARY RUNS OF L2. SEE TABLE IV FOR SYMBOL DEFINITIONS

No.	$N_{f,p}$	N_{LP}	$t_{3D,500}$ (in s)	$\bar{t}_{Ip}$ (in s)	t_{GLC} (in s)
2	30	288	8.13	41.08	7740,47
3	30	240	7.73	49.10	7562,50
10	41	180	7.49	47.56	11827,84

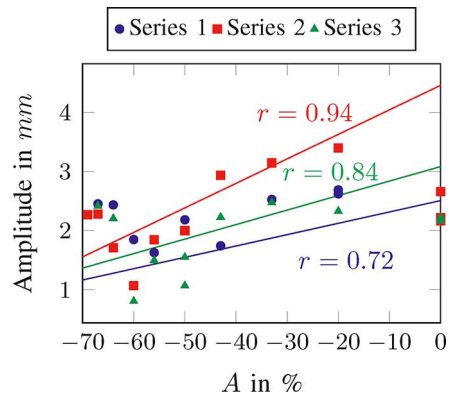

Fig. 13. Amplitude of oscillation for the medial point of the left vocal fold edge for L1. The amplitude generally increases with decreasing asymmetry (correlation coefficients reported for each series).

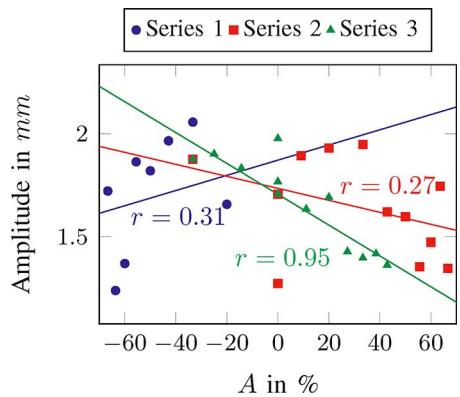

Fig. 14. Amplitude of oscillation for the medial point of the left vocal fold edge for L2. The amplitude generally decreases with increasing positive values of asymmetry.

10 percentage points (pp) lower than the average. Nevertheless, σ_D averaged approximately 4 pp across both larynges.

Accuracy of the tracking showed that a standard deviation of about 1.2 pixels and 0.9 pixels for L1 and L2, respectively, with respect to the defined gold standard was maintained by the tracking. The values for the success rate and accuracy were consistent for each larynx and equal for σ_D. However, the detection rate for L2 was approximately 10 pp higher than that of L1.

The computational costs are depicted in Table III for 3 runs of L2. The table lists the values of $N_{f,p}$, N_{LP}, $t_{3D,500}$, $\bar{t}_{Ip}$, and t_{GLC}. The runs vary in the number of frames per period and the number of reconstructed laser dots. The time for reconstruction of the actual 3D data of 500 frames is in the range of 7.49 s to 8.13 s. For the analysis, a range between 41.08 s and 49.10 s per

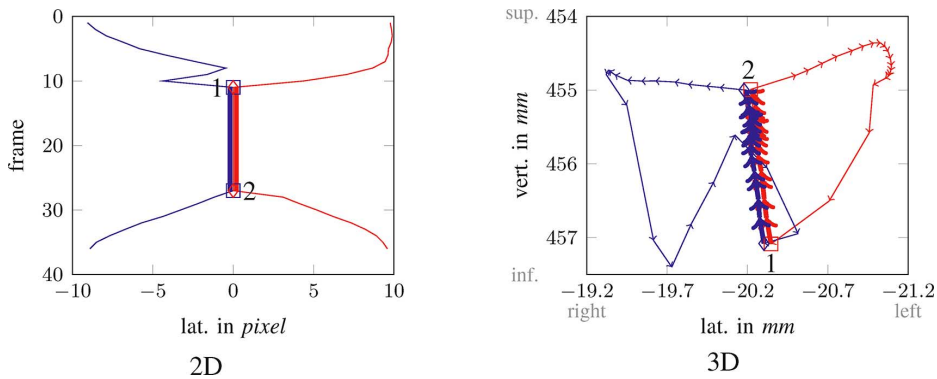

Fig. 15. Comparison of the captured trajectories of the medial points along the left (blue) and right (red) vocal fold edges from 2D (left) and 3D (right) information for one oscillation period. The graphs start at an open position of the vocal folds. The vocal fold collide at point 1 and separate at point 2. The 2D graph plots time in frames over position. The 3D graph plots the vertical over the lateral axis, with arrows indicating the direction of vocal fold edge motion. Whereas there is no information about vocal fold kinematics in 2D during the closed phase, the 3D method captures the bulk motion of the vocal folds in the superior direction.

frame of the normalized period was necessary. The most computational time is spent on computing the mean glottis contour, reaching as high as between 7740 s and 11 827 s.

In Figs. 13 and 14, the amplitudes of the left vocal fold edge at a medial position in the longitudinal direction are depicted for L1 and L2. Measurements were taken in 3 repeated series for L1 and with 3 different asymmetry values for L2. The amplitude increases and decreases visibly with an increase or decrease in asymmetry. This amplitude took into account both vertical and lateral motion, as opposed to only the lateral motion that would be captured in 2D. The amplitude variation generally increased with the degree of asymmetry.

Through the reconstruction of the vertical component, the method also obtains the amplitude of the vocal folds during the closed phase of the oscillation period, when the glottal area is at its minimum. In Fig. 15, left, the medial points of the left and right vocal fold edge are extracted from the high-speed recordings in a purely 2D manner. The y-axis indicates the time in frames covering one period. The x-axis indicates the vocal fold edge position relative to the glottal midline ($x = 0$, in pixels). A negative (positive) value indicates motion toward the left (right) of the glottal midline. Initially, the vocal folds move toward each other during the closing phase. The vocal folds collide at point 1 (frame 11) and remain in contact until point 2 (frame 27). The straight line between points 1 and 2 indicate that apparently no movement is detectable. After that point, the vocal folds open, indicated by the separation of the left and right vocal fold trajectories.

Fig. 15, right, shows a medial slice of the 3D reconstructed vocal fold surface. The x-axis and y-axis represent calibrated coordinates in the lateral and vertical direction, respectively. The vocal fold edge was tracked during one oscillation period, and the motion of the vocal fold edge is indicated by the arrows. Similar to the 2D case, the points of vocal fold closure (1) and opening (2) are indicated. As an advantage over the 2D-only method, the 3D reconstruction captures the vertical movement of the superior vocal fold surface. The vocal folds bulge upward in a vertical direction while closed, which is challenging to extract from 2D information. In addition, the technique is designed to describe laryngeal motion exhibiting a mucosal wave (crest) on the superior surface of the vocal folds.

IV. Discussion

The presented method allows for the 3D reconstruction of the superior vocal fold surface. The incorporation of the glottis contour into the 3D data enables tracking of the lateral and vertical motion of the vocal fold edge. The presented results show the possibilities of measuring vocal fold kinematics during phonation. The system combines high-resolution 3D spatial information with the high temporal resolution afforded by high-speed video imaging.

An error analysis of laser point tracking showed consistency in the detection success rate within each larynx, although a difference between the two larynges was observed. Visual inspection of the recordings indicated that the image data for L2 was of higher contrast and laser points in the image were slightly larger. These observations point to the potential sensitivity of the method toward the quality of the recording in terms of contrast and size of the region of interest in the camera's field-of-view. Nonetheless, the method was able to obtain up to 84% of the points with respect to the manual reference. Also, the standard deviation of the tracked trajectories in comparison to the manual reference was approximately 1 pixel.

The method was applied to two cadaver larynges in an ex vivo setup. For each larynx, tension was introduced to the laryngeal configuration through arytenoid torque variation. The measured values of amplitude and velocity were in the range of other findings, which reported amplitudes of 1 mm to 2 mm for adults and velocities between 0.4 m/s and 1 m/s [13], [14], [30], [31]. Contributions of the current work include relating vibratory parameters to values of induced torque. 3D information of vocal fold oscillation for the entire vocal fold surface and the vocal fold edge has the potential to aid in investigating previously challenging relationships between laryngeal physiology and voice production.

One limitation of the presented method is its limited perspective of vocal fold motion from the superior angle. Additionally, the presented analysis process builds upon a smoothened period of the vocal fold movement. However, by improving on especially the image processing and tracking, a continous and thus, temporally high-resolved analysis

TABLE IV
LIST OF GLOBALLY USED SYMBOLS, CORRESPONDING UNITS, AND DESCRIPTIONS

Symbol	Unit	Description
A		Asymmetry quotient
B		Bézier basis function
C		set of points forming the glottis contour
D_{L}, D_{R}	mNm	Torque to the left, right arytenoid
N_{f}		number of frames in an high-speed video film
$N_{\mathrm{f,p}}$		number of frames in a period
N_p		number of oscillation periods in an high-speed video film
N_{LP}		number of laser points that were detected
T_i	s	period length of the i-th period
b		B-spline basis function
c		cubic spline basis function
d		distance function between a curve and a point
$\mathbf{l}$		orientation of a ray
$\mathbf{n}$		normal vector onto a plane
n_{f}		index into the frames
n_{I}		index of the initial tracking frame
n_{p}		index into the oscillation periods
$\mathbf{m}$	pixel	laser point in an image
$\mathbf{m}_{\mathrm{m}}$	pixel	averaged point position
$\mathbf{m}^{\mathrm{i,j}}$	pixel	laser point in an image, assigned to a grid coordinate i, j
$\mathbf{m}_{\mathrm{w}}$	mm	reconstructed 3D point
s_{GC}		scaling factor of the glottis contour
t		trajectory of point
$t^{i,j}$		trajectory of an image point, assigned to a grid place i, j
$t^{i,j}_{\mathrm{T,j}}$		target time vector, assigned to a grid place i, j
$t^{i,j}_{\mathrm{S,j}}$		source time vector, assigned to a grid place i, j
$t_{\mathrm{3D,500}}$	s	time necessary to reconstruct the 3D coordinates of points for 500 frames
$\bar{t}_{\mathrm{Ip}}$	s	mean time to compute the interpolated surface for one frame
t_{GLC}	s	time to compute the average glottis contours for all frames of a normalized period
u		normalized coordinate on a surface/curve
v		normalized coordinate on a surface/curve
δ	pixel	deviation threshold from period to period
γ		velocity threshold for tracking
φ		acceleration threshold for tracking

TABLE V
EXPERIMENTAL PARAMETERS FOR LARYNX L1. GIVEN ARE SUBGLOTTAL PRESSURE P_{sub}, APPLIED TORQUES D_{L} AND D_{R} TO LEFT AND RIGHT ARYTENOIDS, RESPECTIVELY, AND THE ASYMMETRY QUOTIENT A

	Run #	P_{sub} (cm H_2O)	D_{L} (mN m)	D_{R} (mN m)	A (%)
Series 1	1	8.4	5	5	0
	2	10.4	7.5	5	-20
	3	9.9	7.5	5	-20
	4	13.3	10	5	-33
	5	10.7	12.5	5	-43
	6	12.2	15	5	-50
	7	13.7	17.5	5	-56
	8	14.2	20	5	-60
	9	15.1	22.5	5	-64
	10	15.7	25	5	-67
Series 2	11	10.1	5	5	0
	12	11.0	7.5	5	-20
	13	11.8	10	5	-33
	14	14.2	12.5	5	-43
	15	14.2	15	5	-50
	16	15.6	17.5	5	-56
	17	16.0	20	5	-60
	18	16.0	22.5	5	-64
	19	16.3	25	5	-67
	20	16.4	27.5	5	-69
Series 3	21	14.9	5	5	0
	22	15.1	7.5	5	-20
	23	15.3	10	5	-33
	24	15.2	12.5	5	-43
	25	15.0	15	5	-50
	26	15.1	17.5	5	-56
	27	15.3	20	5	-60
	28	15.3	22.5	5	-64
	29	15.4	25	5	-67
	30	14.8	15	5	-50

is possible (e.g., for period-to-period analysis). Several alternative methods such as optical coherence tomography [32] or magnetic resonance imaging [33] offer the potential for imaging beneath the vocal fold surface; however, these methods have not proven to provide adequate temporal and spatial resolution and thus are only applicable in a stroboscopic manner (i.e., not at a high-speed frame rate). Vocal fold vibration associated with severe vocal pathology often require high-speed imaging for laryngeal imaging.

The assumption regarding the position of the vocal fold edge has to be confirmed in the future, as it governs the shaping of the vocal fold surface. The validity could be investigated by further experiments that visualize the medial and superior vocal fold surface simultaneously. It would be especially interesting to combine methods such as optical coherence tomography with high-speed video to obtain 3D information of the superior, medial, and inferior vocal fold edges.

The goals of the current work are motivated by the clinical need for systematic studies to describe and develop acoustic correlates of irregularities in vocal fold vibration to aid clinicians in the effective management of voice disorders. The clinical utility of results in the literature are limited by the relatively low or nonexistent correlations between measures of irregularity in vocal fold vibration and acoustic parameters, providing motivation for continued efforts to account for more of the unexplained variance in the acoustic and HSV-based measures. We suggest that such efforts include the addition of aerodynamic measures and methods for capturing the three-dimensional motion of the vocal folds to more comprehensively describe the complex fluid-structure-acoustic interaction that takes place during phonation.

With regards to clinical application, obtaining 3D vocal fold kinematic information in vivo may aid in the diagnosis and treatment of voice disorders, with the creation of an endoscopic system a necessary step. Fundamental research into human voice production and, thus, the understanding of the phonation process has the potential to provide critical insight into relationships such as between mechanical properties of the vocal folds and vibratory characteristics. Enhanced imaging can provide the clinician with insight into vibratory properties not visible with traditional imaging modalities.

TABLE VI
EXPERIMENTAL PARAMETERS FOR LARYNX L2. GIVEN ARE SUBGLOTTAL PRESSURE P_{sub}, APPLIED TORQUES D_{L} AND D_{R} TO THE LEFT AND RIGHT ARYTENOIDS, RESPECTIVELY, AND THE ASYMMETRY QUOTIENT A

	Run #	P_{sub} (cm H_2O)	D_{L} (m Nm)	D_{R} (m Nm)	A (%)
Series 1	1	10.2	5	5	0
	2	10.3	7.5	5	-20
	3	12.1	10	5	-33
	4	10.0	12.5	5	-43
	5	9.7	15	5	-50
	6	10.7	17.5	5	-56
	7	11.5	20	5	-60
	8	10.5	22.5	5	-64
	9	10.8	25	5	-67
Series 2	10	15.2	5	5	0
	11	12.1	5	6	9
	12	10.4	5	7.5	20
	13	11.7	5	10	33
	14	10.7	5	12.5	43
	15	9.8	5	15	50
	16	9.8	5	17.5	56
	17	10.8	5	20	60
	18	12.4	5	22.5	64
	19	11.3	5	25	67
Series 3	20	10.6	10	5	-33
	21	10.0	10	6	-25
	22	9.8	10	7.5	-14
	23	10.0	10	10	0
	24	15.2	10	12.5	11
	25	10.6	10	15	20
	26	9.3	10	17.5	27
	27	9.3	10	20	33
	28	10.9	10	22.5	38
	29	10.6	10	25	43

APPENDIX

See Table IV–VI.

REFERENCES

[1] R. J. Ruben, "Redefining the survival of the fittest: Communication disorders in the 21st century," *Laryngoscope* vol. 110, no. 2, pt. 1, pp. 241–245, Feb. 2000.

[2] I. R. Titze, Myoelastic aerodynamic theory of phonation Nat. Center Voice Speech, 2006.

[3] K. Stevens, *Acoustic Phonetics, 1998*. Cambridge, MA: MIT Press, 1999.

[4] S. M. Zeitels, A. Blitzer, R. E. Hillman, and R. R. Anderson, "Foresight in laryngology and laryngeal surgery: A 2020 vision," *Ann. Otol. Rhinol. Laryngol Suppl.*, vol. 198, pp. 2–16, Sept. 2007.

[5] R. E. Hillman, W. W. Montgomery, and S. M. Zeitels, "Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders," *Curr. Opin, Otolaryngol, Head Neck Surg.*, vol. 5, pp. 172–175, 1997.

[6] D. D. Mehta and R. E. Hillman, "Current role of stroboscopy in laryngeal imaging," *Curr. Opin. Otolaryngol. Head Neck Surg.* vol. 20, no. 6, pp. 429–436, Dec. 2012.

[7] D. Deliyski *et al.*, "Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution," *Folia Phoniatrica et Logopaedica*, vol. 60, pp. 33–44, 2008.

[8] D. D. Mehta, M. Zañartu, T. F. Quatieri, D. D. Deliyski, and R. E. Hillman, "Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy," *J. Acoust. Soc. Am.* vol. 130, no. 6, pp. 3999–4009, 2011.

[9] D. D. Mehta *et al.*, "High-speed videoendoscopic analysis of relationships between cepstral-based acoustic measures and voice production mechanisms in patients undergoing phonomicrosurgery," *Ann. Otol. Rhinol. Laryngol.*, vol. 121, no. 5, pp. 341–347, 2012, 5.

[10] D. D. Mehta, D. D. Deliyski, S. M. Zeitels, T. F. Quatieri, and R. E. Hillman, "Voice production mechanisms following phonosurgical treatment of early glottic cancer," *Ann. Otol. Rhinol. Laryngol.*, vol. 119, no. 1, pp. 1–9, Jan. 2010.

[11] M. Döllinger, D. A. Berry, and G. S. Berke, "Medial surface dynamics of an in vivo canine vocal fold during phonation.," *J. Acoust. Soc. Am.*, vol. 117, no. 5, pp. 3174–3183, May 2005.

[12] C.-G. Tsai, J.-H. Chen, Y.-W. Shau, and T.-Y. Hsiao, "Dynamic b-mode ultrasound imaging of vocal fold vibration during phonation," *Ultrasound Med. Biol.* vol. 35, no. 11, pp. 1812–1818, Nov. 2009.

[13] N. A. George, F. F. M. de Mul, Q. Qiu, G. Rakhorst, and H. K. Schutte, "Depth-kymography: High-speed calibrated 3D imaging of human vocal fold vibration dynamics," *Phys. Med. Biol.*, vol. 53, no. 10, pp. 2667–2675, 2008.

[14] R. R. Patel, K. D. Donohue, D. Lau, and H. Unnikrishnan, "In vivo measurement of pediatric vocal fold motion using structured light laser projection," *J. Voice* vol. 27, no. 4, pp. 463–472, July 2013.

[15] L. Yu *et al.*, "Office-based dynamic imaging of vocal cords in awake patients with swept-source optical coherence tomography," *J. Biomed. Opt.* vol. 14, no. 6, p. 064020, 2009.

[16] D. E. Sommer and I. T. Tokuda *et al.*, "Estimation of inferior-superior vocal fold kinematics from high-speed stereo endoscopic data in vivo," *J. Acoust. Soc. Am.* vol. 136, no. 6, pp. 3290–3300, 2014.

[17] G. Luegmair *et al.*, "Optical reconstruction of high-speed surface dynamics in an uncontrollable environment," *IEEE Trans. Med. Imag.* vol. 29, no. 12, pp. 1979–1991, Dec. 2010.

[18] I. Pitas and A. Venetsanopoulos, "Homomorphic filters," in *Nonlinear Digital Filters*, ser. Int. Ser. Eng. Comput. Sci.. New York: Springer, 1990, vol. 84, pp. 217–243.

[19] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man., Cybern., Syst.* vol. 9, no. 1, pp. 62–66, Jan. 1979.

[20] J. Lohscheller, U. Eysholdt, H. Toy, and M. Döllinger, "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics," *IEEE Trans. Med. Imag.*, vol. 27, no. 3, pp. 300–309, Mar. 2008.

[21] E. C. Inwald, M. Döllinger, M. Schuster, U. Eysholdt, and C. Bohr, "Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging," *J. Voice* vol. 25, no. 5, pp. 576–590, 2011.

[22] J. Kreiman *et al.*, "Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation," *J. Acoust. Soc. Am.* vol. 132, no. 4, pp. 2625–2632, 2012.

[23] J. Dauguet, A.-S. Hérard, J. Declerck, and T. Delzescaux, "Locally constrained cubic b-spline deformations to control volume variations," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano to Macro*, 2009, pp. 983–986.

[24] S. Sandor and R. Leahy, "Surface-based labeling of cortical anatomy using a deformable atlas," *IEEE Trans. Med. Imag.*, vol. 16, no. 1, pp. 41–54, Feb. 1997.

[25] C. Sinthanayothin and W. Bholsithi, "3D facial deformable using cubic spline and thin plate spline," in *Proc. 6th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol.*, 2009, vol. 02, pp. 668–671.

[26] M. Döllinger and D. A. Berry, "Computation of the three-dimensional medial surfce dynamics of the vocal folds," *J. Biomech.*, vol. 39, no. 2, pp. 369–374, 2006.

[27] M. Döllinger and D. A. Berry, "Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation," *J. Voice* vol. 20, no. 3, pp. 401–413, Sep. 2006.

[28] C. Bohr, A. Kraeck, U. Eysholdt, A. Ziethe, and M. Döllinger, "Quantitative analysis of organic vocal fold pathologies in females by high-speed endoscopy," *Laryngoscope* vol. 123, no. 7, pp. 1686–1693, 2013.

[29] in *Proc. 13th Mechatron. Forum Int. Conf.*, Sep. 17–19, 2012, vol. 3/3, pp. 135–141, 2012.

[30] T. Wurzbacher *et al.*, "Calibration of laryngeal endoscopic high-speed image sequences by an automated detection of parallel laser line projections," *Med. Image Anal.*, vol. 12, no. 3, pp. 300–317, Jun. 2008.

[31] G. Schade, T. Kirchhoff, and M. Hess, "Laser measuring device for phonation," *Folia Phoniatr Logop*, vol. 57, no. 4, pp. 202–215, 2005.

[32] J. B. Kobler, E. W. Chang, S. M. Zeitels, and S.-H. Yun, "Dynamic imaging of vocal fold oscillation with four-dimensional optical coherence tomography," *Laryngoscope* vol. 120, no. 7, pp. 1354–1362, Jul. 2010.

[33] A. Majumdar, R. K. Ward, and T. Aboulnasr, "Compressed sensing based real-time dynamic MRI reconstruction," *IEEE Trans. Med. Imag.* vol. 31, no. 12, pp. 2253–2266, Dec. 2012.