Medial surface dynamics of an in vivo canine vocal fold during phonation

Michael Döllinger, David A. Berry, and Gerald S. Berke

Citation: The Journal of the Acoustical Society of America **117**, 3174 (2005); doi: 10.1121/1.1871772
View online: http://dx.doi.org/10.1121/1.1871772
View Table of Contents: http://asa.scitation.org/toc/jas/117/5
Published by the Acoustical Society of America

Medial surface dynamics of an *in vivo* canine vocal fold during phonation

Michael Döllinger,[a] David A. Berry, and Gerald S. Berke
The Laryngeal Dynamics Laboratory, UCLA Division of Head & Neck Surgery, 1000 Veteran Ave. Suite 31-24, Los Angeles, California, 90095-1794

(Received 6 May 2004; revised 25 January 2005; accepted 26 January 2005)

Quantitative measurement of the medial surface dynamics of the vocal folds is important for understanding how sound is generated within the larynx. Building upon previous excised hemilarynx studies, the present study extended the hemilarynx methodology to the *in vivo* canine larynx. Through use of an *in vivo* model, the medial surface dynamics of the vocal fold were examined as a function of active thyroarytenoid muscle contraction. Data were collected using high-speed digital imaging at a sampling frequency of 2000 Hz, and a spatial resolution of 1024 $\times$1024 pixels. Chest-like and fry-like vibrations were observed, but could not be distinguished based on the input stimulation current to the recurrent laryngeal nerve. The subglottal pressure did distinguish the registers, as did an estimate of the thyroarytenoid muscle activity. Upon quantification of the three-dimensional motion, the method of Empirical Eigenfunctions was used to extract the underlying modes of vibration, and to investigate mechanisms of sustained oscillation. Results were compared with previous findings from excised larynx experiments and theoretical models. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1871772]

PACS numbers: 43.70.Aj, 43.70−h [AL] Pages: 3174–3183

I. INTRODUCTION

Although voice production is a critical component of human communication, the process of voice generation is not well understood, particularly with regard to the voice source. Because of the small size and inaccessibility of the larynx, the possibilities for laryngeal imaging are limited. Most *in vivo* recordings of vocal fold vibration have been performed from a superior aspect using endoscopy.[1–3] From this aspect, it is impossible to quantify the medial surface dynamics of the vocal folds, which describe the opening and closing of the glottis that generates sound. From a superior aspect, it is also impossible to quantify the propagation of the mucosal wave along the medial surface of the folds. Because of these limitations, clinicians usually refer to mucosal wave propagation along the superior surface of the folds, where they see the wave more clearly. Unfortunately, the mucosal wave attenuates quickly upon reaching the superior surface of the folds.

In terms of its influence on voice production, the most critical region of mucosal wave propagation occurs on the medial surface of the vocal folds, where it originates.[4,5] Before reaching the superior surface, the wave travels a relatively long distance along the medial surface, it generates significant tissue vibrations that exhibit geometric and viscoelastic nonlinearities,[6,7] and it couples nonlinearly with other systems, including the opposite vocal fold (e.g., through collision during glottal closure) and sub- and supraglottal systems (e.g., through fluid–structure interactions with the glottal airflow).[8] Indeed, mucosal wave propagation along the medial surface of the folds is governed by a very complex set of dynamics that is only beginning to be understood.[5] The modeling of such complexity is still in its infancy, and requires further quantitative data for validation. Hence, quantification of the medial surface dynamics of the vocal folds holds promise for increasing our understanding of voice production and for supporting computational modeling efforts.[9–11]

Because of the difficulty in imaging the medial surface of the vocal folds, only a few quantitative studies have attempted such imaging.[12–16] However, of these investigations, Berry *et al.*[16] was the only study to image and quantify the medial surface of the vocal folds with sufficient spatial and temporal resolution to enable extraction of the underlying modes of vibration, sometimes referred to as Empirical Eigenfunctions.[17] Such modes have been shown to be helpful for investigating mechanisms of both normal and abnormal voice production.[16,17]

One of the limitations of the previous study of medial surface dynamics was the use of the excised larynx,[16] which neglected the influence of the thyroarytenoid (TA) muscle on vocal fold vibration. With the excised larynx, laryngeal adjustments were implemented through means of external forces. For example, an external force that pulled downward on the thyroarytenoid notch caused the thyroarytenoid cartilage to rock forward on the cricoid cartilage, lengthening the vocal fold. Similarly, an external force that pulled anteriorly on the muscular process caused the arytenoid to rock anteriorly and medially, adducting the vocal process. However, while such external forces can simulate many laryngeal adjustments, they cannot simulate TA contraction. The simultaneous contraction, stiffening, and medial surface bulging of the TA can only occur through active muscular contraction.[18] Muscle shortening and medial surface bulging create a deeper medial surface over which the mucosal wave propagates, impacting the medial surface dynamics of the vocal

[a]Electronic mail: michael.doellinger@gmx.net

 0001-4966/2005/117(5)/3174/10/$22.50

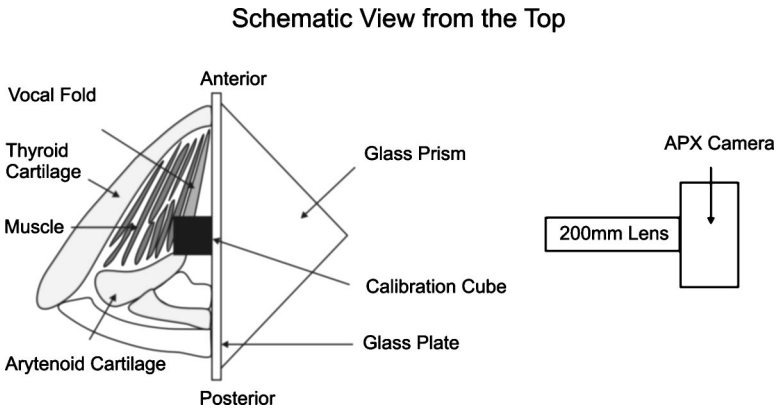

FIG. 1. A schematic representation of the experimental setup from a superior view: The glass plate, situated at the glottal midline, appears as a vertical line in the middle of the picture. The vocal fold and brass calibration cube are situated on the left side of the plate, and the glass prism and high-speed camera are situated on the right.

fold. Increased muscle stiffness may also increase the velocity of the mucosal wave. In addition, TA muscular adjustment is a critical factor influencing vocal registers, such as falsetto and chest.[18] For example, in the chest register, increased TA muscle contraction significantly increases the rate of closure.

On the other hand, the absence of TA muscle contraction simulates a condition of paralysis of the membranous vocal folds, in which the stiffness of the muscle is reduced, approaching the significantly lower stiffness of the cover.[19] In this situation, the vocal folds experience large vertical excursions, reduced mucosal wave velocity, and a relatively gentle closure of the membranous vocal folds. Interestingly, this is the only condition of the membranous glottis that can be simulated in the excised larynx, because the TA is denervated in this model. The inability to model active TA contraction limits the range of glottal conditions that can be investigated in the excised larynx model. The inclusion of the *in vivo* canine laryngeal model broadens the scope of an investigation of the medial surface dynamics of the vocal fold, allowing a more representative range of normal and pathological glottal conditions to be investigated.

Nevertheless, because of the invasive nature of the hemilarynx setup, many technical challenges were involved in adapting it for an *in vivo* study. Hence, the purpose of the present investigation was to demonstrate the viability of the hemilarynx model in quantifying the medial surface dynamics of an *in vivo* canine vocal fold as a function of active contraction of the thyroarytenoid muscle; to extract the underlying modes of vibration; and to investigate mechanisms of sustained oscillation.

II. MATERIAL AND METHODS

A. Experimental setup

The experimental setup in this study was an adaptation of recent work with an excised hemilarynx methodology.[16,20,21] The work was also inspired by previous quantitative imaging of a full excised larynx by Baer,[12] previous excised hemilarynx work by Jiang and Titze,[14] and previous investigations of the *in vivo* canine laryngeal model by Berke *et al.*,[22] and Bielamowicz *et al.*[23]

A male mongrel canine, weighing 25 kg, was selected for the study. The canine was premedicated intramuscularly with acepromazine and buprenex. Over the course of the experiment, halothane ventilation and intravenous pentobaritol was administered to prevent reflexive response to corneal stimulation. Following the experiment, the canine was humanely sacrificed with intravenous *Eutha-6*.

The canine was placed supine on an operating table. A hemilarynx was created through means of a neck dissection and a hemilaryngectomy, which included the removal of half of the hyoid bone. To provide a controlled air supply for phonation, air was insufflated rostrally through a tracheostomy. For the vibration patterns reported in this study, the glottal airflow was held relatively constant at 660 ± 15 ml/s, and the subglottal pressure was varied between 1.2 and 3.7 kPa. The air was heated (37 °C) and humidified using a humidifier (Concha Therm III, Servo-Controlled Heater). The canine was assist-ventilated (Monaghan 300 D/M Ventilator) with 4 L/min O_2 via an endotracheal tube. A schematic representation of the experimental setup is shown in Fig. 1.

So that specific flesh points could be tracked during the experiments, small black monofilament nylon microsutures were placed on the medial surface of the vocal fold (size: 9-0, diameter: 0.034 mm, Ethicon, Inc.). To minimize any disturbance of the natural dynamics, the microsutures were positioned to penetrate only the mucosal epithelium, with a reported thickness of 0.05–0.1 mm.[24] Five columns of microsutures with five microsutures per column were mounted along the medial–superior surface of the vocal fold, where the greatest amplitudes of vibration were expected to occur; see Fig. 2. The distance between the flesh points were 3.1 ± 0.2 mm in the anterior–posterior direction and 2.1 ± 0.2 mm in the vertical direction.

Next, the trachea was mounted over a stainless steel cylindrical tube with an inner diameter of 8 mm. A wedge was mounted within the tube to smoothly channel the airflow beneath the vocal fold. At the top of the tube, a glass plate (width: 61.2 mm, height: 61.1 mm, thickness: 3.2 mm was mounted at the glottal midline; Fig. 1. To eliminate airflow leaks between the vocal fold and the glass plate, strips of gauze, coated with vacuum grease, were inserted between them. The preparation of the experiment took approximately four hours. Following preparation, the experiment itself took less than two hours.

To simultaneously achieve muscle contraction, muscle stiffening, medial surface bulging, and vocal fold adduction, the recurrent laryngeal nerve (RLN) was stimulated using a constant-current nerve stimulator (WR Medical Electronics

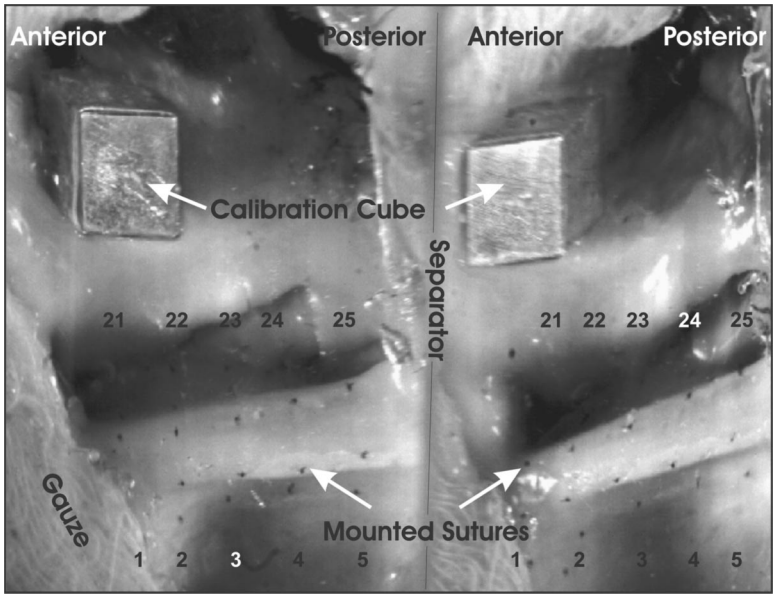

FIG. 2. Split view provided by the prism, as seen through the high-speed camera. Here 25 sutures are enumerated in both views. Above the sutures, the calibration cube is also visible in both views. In both views, the anterior sutures are to the left, and the posterior sutures are to the right. On the lower left side, gauze is visible concealing suture number one.

Co. Model 2SLH, St. Paul, MN). The RLN was stimulated with currents ranging from 0.04–0.08 mA at 80 Hz, with a 1.5 ms pulse duration.[22,23] The superior laryngeal nerve (SLN) was not stimulated. The RLN was isolated 5 cm inferior to the larynx in the tracheoesophageal groove. Custom designed rubber electrodes (monopolar, flexible, conductive neopreme with silicone, and silicone insulation KE45) were applied to the isolated nerve at the most proximal point dissected. Electrical isolation of the nerve was confirmed by direct visualization of the vocal folds during stimulation. The animal required dissection and division of the posterior branch of the RLN to the posterior cricoarytenoid muscle to achieve adduction and modal phonation.

For calibration purposes, a brass cube (5.04 mm $\times$5.04 mm$\times$5.06 mm) was glued to the glass plate superior to the vocal fold. The cube was large enough that one could identify its corners and edges, but small enough to avoid significant disturbance of the glottal airflow (Fig. 1). A right-angle prism was placed against the glass plate, on the side opposite to the vocal fold. The prism yielded two different camera views,[16] which was necessary for the computation of three-dimensional movement.[25] The prism was large enough to yield a view of the entire medial surface of the vocal fold.

The vibrations of the vocal fold were imaged with a high-speed digital camera (Fastcam-Ultima APX, Photron Unlimited, Inc.) with a 200 mm lens (AF MICRO NIKKOR) at a frame rate of 2000 Hz and spatial resolution of 1024 $\times$1024 pixels. Three 150 W lamps (MI-150 Illuminator, Dolan-Jenner, Inc.) served as light sources. The light sources were only applied for a few seconds at a time during imaging to prevent tissue dehydration. Appropriate tissue hydration was further facilitated using a humidified air supply, as already described. The acoustical signal was simultaneously recorded at a distance of 10 cm to the acoustical source with a condenser microphone (Model No. 4193, Brüel & Kjaer Inc.) and digitized at 44.1 kHz. The high-speed recording and the acoustical signal were synchronized through means of an external trigger.

B. Extraction of flesh points

Two primary issues complicated automatic detection of the mounted sutures (i.e., automatic flesh point extraction). First, the sutures on the superior surface of the folds (the top two anterior–posterior rows of sutures) were concealed during portions of the glottal cycle (Fig. 2). Second, the close proximity of the sutures sometimes resulted in overlapping trajectories. Hence, a semi-automatic algorithm was developed, which allowed user intervention to correct possible errors. The algorithm was implemented as follows: First, to increase visibility of the flesh points, all frames were contrast enhanced.[26] Next, for the first five frames, the flesh point positions were selected manually. From these data, the mean values and the variances of each flesh point were computed. In succeeding frames, flesh point positions were determined automatically: a quadratic extrapolation estimated the new flesh point position, a search was performed around this position, with a radius corresponding to the predetermined variance, to locate the cluster of points with the lowest gray values (darkest points). The center of mass of these points was selected as the new flesh point position, and the mean values and variances were appropriately updated. As previously indicated, manual intervention was allowed to correct possible errors.

C. Computation of physical coordinates

To compute the three-dimensional coordinates of the flesh points, at least two two-dimensional camera views were necessary.[25] As in Berry *et al.*,[16] the split view through the prism provided the two camera views (Fig. 2). In the previous hemilarynx study,[16] calibration was performed at the end of the experiment using a Direct Linear Transform (DLT). This method would be inappropriate for the *in vivo* canine, because it would be impossible to keep camera parameters constant across experimental conditions. For example, canine breathing caused lifting and lowering of the chest and the hemilarynx setup (i.e., the prism and the glass plate with

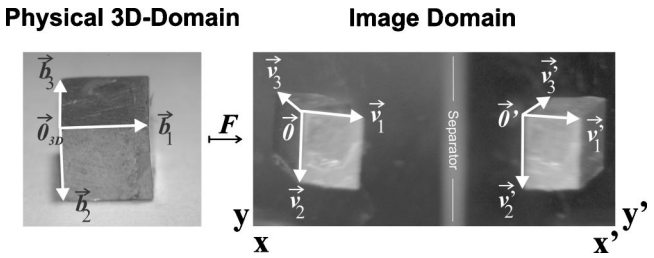

FIG. 3. Schematic representation of the LA method through means of the transformation matrix *F*. The left picture shows the brass cube in the physical domain, while the right picture shows the two camera views through the prism. Both camera views were assumed to preserve parallel projections.

a cube). Stimulating the RLN repositioned the vocal fold at the glass plate. The vocal fold vibrations themselves could potentially induce slight translations of the experimental setup. Also, because of the high sampling rates required to capture laryngeal vibrations, along with the high light intensity, and low *f* stop (i.e., small depth of field), it was often necessary to refocus the camera to maintain sharp focus across experimental conditions.

To deal with the possibility of changing camera parameters, a calibration object (i.e., a brass cube) was present in each imaging frame; see Fig. 2. Because the cube was situated above the vocal fold, an extrapolation had to be performed to compute the physical coordinates of the sutures. This disqualified the use of the DLT,[27–30] which works well under interpolation, but not under extrapolation.[31] Consequently, a linear approximation method (LA method) was implemented. In a recent excised hemilarynx investigation, the LA method yielded a more accurate calibration than the DLT,[20,21] presumably because of the DLT's poor extrapolation performance.

The primary assumption of the LA method was that parallel projections were preserved in the two different camera views. In other words, the parallel edges of the brass cube would remain parallel in both camera views; see Fig. 3. For this assumption to be valid, the camera lens had to be positioned approximately perpendicular to the glass plate (Fig. 1). This assumption was not difficult to satisfy given the relatively small region of interest (e.g., approximately a 1 cm width, 1 cm height, and 0.5 cm depth); Fig. 2. Therefore, any distortions caused by nonparallelism between the camera lens and the glass plate induced only small errors in the mapping to three-dimensional coordinates. Indeed, when parallelism was compromised because of camera orientation, the distortions of the brass cube were immediately apparent in the camera images. As a result, the camera position was refined until the distortions were no longer apparent by mere inspection. The final distortions were quantified using the root-mean-square error.[32] In the recordings to be presented later, the computed linearization error was 9.1%, which was presumably introduced by camera orientation.

Mathematically the LA method may be explained as follows. First, there is a mapping *f* of the three-dimensional physical domain $\mathbb{R}^3$ in a two-dimensional image $\mathbb{R}^2$ (camera view); see Fig. 3. By recording images through a prism, the camera view is split and the image domain can be regarded as two two-dimensional spaces $\mathbb{R}^2\times\mathbb{R}^2$, or as a subspace of $\mathbb{R}^4$. For computational reasons (e.g., matrix inversion), the physical domain was expanded from $\mathbb{R}^3$ to $\mathbb{R}^4$ to yield a mapping: $f:\mathbb{R}^4\mapsto\mathbb{R}^4$. The added fourth dimension in the physical domain served as a dummy dimension. Linear mappings are uniquely determined by the mapping of the basis vectors of the domain. As basis vectors, the three orthogonal directions of the brass cube were chosen, with the origin of the coordinate system at the left upper corner of the cube $\vec{0}_{3\text{-D}}$; see Fig. 3: $\vec{b_1}$, $\vec{b_2}$, $\vec{b_3}$. A fourth orthogonal basis vector, $\vec{b_4}$ was chosen, which spanned the introduced fourth dimension. The basis vectors were mapped by the split camera view (Fig. 3) into

$$\begin{aligned} f(\vec{b_1}) &= (\vec{v}_1,\vec{v}_1') = (x_1,y_1,x_1',y_1'),\\ f(\vec{b_2}) &= (\vec{v}_2,\vec{v}_2') = (x_2,y_2,x_2',y_2'),\\ f(\vec{b_3}) &= (\vec{v}_3,\vec{v}_3') = (x_3,y_3,x_3',y_3')\\ f(\vec{b_4}) &= (0,0,0,y_4'). \end{aligned} \tag{1}$$

The linear mapping *f* from the three-dimensional physical world to the split camera view could now be described by a matrix *F*. Finally, the three-dimensional physical coordinates $\vec{v}_{3\text{-D}}$ of the sutures were computed by multiplying the images coordinates $\vec{v}_{\text{rec}}$ by the inverse matrix F^{-1}:

$$F=\begin{pmatrix} x_1 & x_2 & x_3 & 0\\ y_1 & y_2 & y_3 & 0\\ x_1' & x_2' & x_3' & 0\\ y_1' & y_2' & y_3' & y_4' \end{pmatrix}, \quad \text{with} \quad \vec{v}_{3\text{-D}}=F^{-1}\cdot\vec{v}_{\text{rec}}. \tag{2}$$

For better computational results, the entries of *F* were optimized using the Nelder–Mead Algorithm,[33,34] which has successfully optimized parameters of theoretical models to fit measured vocal fold vibration patterns.[3,35]

D. Data analysis

Because some sutures were not visible during portions of the glottal cycle, not all sutures could be extracted for an analysis. In particular, the two most superior rows (sutures 16–25, Fig. 2), which were placed on the superior surface of the vocal fold, could not be extracted over the entire glottal cycle. Hence, the quantitative investigations for the later reported recording R1 were restricted to sutures No. 1–15 (e.g., the lower three rows), which were placed on the medial surface of the vocal fold. For the other analyzed recording (R2), additionally, sutures no. 1,6,11 were not considered, since the extracted displacements were very small (i.e., 1–2 pixels) and therefore not appropriate for further computation. The analysis was performed over a time period of 100 ms, or 200 frames. The acoustically measured fundamental frequency ranged from 180–220 Hz for chest-like phonation and from 30–80 Hz for fry-like phonation.

Displacements, velocities, and accelerations were computed over the entire medial surface of the vocal fold. To further probe mechanisms of sustained oscillation, Empirical Eigenfunctions (EEF) were also computed. Because EEFs have been described in earlier work,[16,17,36] they will not be explained here in detail. Conceptually, they may be viewed

as the basic building blocks of many simple and complex vibration patterns. In other words, they may be viewed as a decomposition of the vocal fold oscillations into their basic vibratory degrees of freedom. In previous computational studies, essentially constant EEFs explained a variety of periodic and aperiodic vocal fold vibration patterns, which were distinguished primarily by the entrainment (i.e., the same fundamental frequency) or lack of entrainment of the EEFs.[10,17]

III. RESULTS AND DISCUSSION

The hemilarynx model was adapted for the *in vivo* canine, yielding a clear view of the medial surface of the vocal fold, and opening the door for future detailed examination of the influence of the thyroarytenoid muscle on phonation. Improvements to the experimental setup and calibration procedure yielded high quality recordings and quantitative results. In comparison to previous hemilarynx studies, the entire medial surface of the vocal fold was imaged, instead of just one coronal cross section.[16] Furthermore, three-dimensional vibrations were reported in the present investigation, in comparison to previous computational and experimental studies in which only two-dimensional vibrations were reported (e.g., anterior–posterior vibrations were ignored).[10,16,17]

For this *in vivo* hemilarynx, phonation threshold pressure occurred at 1.2 kPa, and phonation was observed across a subglottal pressure range of 1.2–3.7 kPa. Phonation was also observed across a range of RLN stimulation from 0.04–0.08 mA. The original design of this experiment was to report a variety of vibrations as a function of the stimulation current to the RLN. However, this was not a stable input measure with which to differentiate the different vibration patterns. That is, both chest-like and fry-like vibrations were observed across the entire range of current stimulations.

As shown in Nasri *et al.*,[37] the TA stimulation (as measured by electromyography or EMG) is a sigmoidal, or step function of the RLN stimulation current (or voltage). In particular, over a very narrow range of RLN stimulation current, the TA activity varied from 0% to 100% of maximum stimulation. Thus, with only a small amount of input noise, the resultant TA activity may be completely unknown. Perhaps due to this phenomenon, in the present study, the recorded vibration patterns occurred randomly with respect to the input RLN stimulation current.

Nevertheless, the Nasri *et al.*[37] study also reported that robust measures of TA activity (expressed in the percentage of maximum stimulation) could be estimated based on the dynamical range of subglottal pressure. For the present study, the dynamical range of subglottal pressures was 1.2–3.7 kPa. Using the empirical relationship from Nasri *et al.*,[37] periodic, chest-like phonation was observed in the range of 12%–37% of maximum TA stimulation (corresponding to a subglottal pressure of 1.2–2.2 kPa). Fry-like phonation was observed in the range of 37%–100% of maximum TA stimulation (corresponding to a subglottal pressure of 2.2–3.7 kPa), characterized by low fundamental frequencies (30–80 Hz), irregular oscillations, and pulse-like vibrations, with relatively long periods of glottal closure. Within the range of chest-like vibrations, fundamental frequency increased with increasing TA activity. However, beyond a certain threshold of TA activity, fundamental frequency abruptly dropped, resulting in fry-like vibrations. Similar phenomena have been previously observed and predicted.[38–41] For example, it is known that TA activity causes the TA muscle to simultaneously stiffen and contract. If the muscle itself were a significant component of the vibrating tissue (presumably this would be the case in chest voice), as TA activity was increased, the increased stiffness of the TA would result in an increase in the fundamental frequency. Because abrupt jumps between registers (bifurcations) commonly occur for small changes in some input parameter, it is not surprising that an abrupt jump is observed between chest-like and fry-like vibrations for some critical threshold of maximum TA stimulation. However, if the tissue vibrations only occurred in the cover (presumably this would be the case in vocal fry), as TA activity were increased, the contraction of the TA would also shorten the cover, decreasing the stiffness in the cover, and thereby decreasing the fundamental frequency. Future studies will report further details regarding the dynamics of the fry-like vibrations observed in the *in vivo* canine larynx.

TABLE I. The TA stimulation level, the RLN stimulation, the measured subglottal pressure (sP), the applied air flow, and the measured fundamental frequency for the investigated recordings.

Recording	TA stimulation (%)	RLN (mA)	sP (kPa)	Flow (ml/s)	F_0 (Hz)
R1	33	0.07	2.1	650	220
R2	12	0.04	1.6	675	180

In this study, we focus on the details of two chest-like vibrations at 33% (R1) and 12% (R2) of TA activity, respectively. The details regarding these two phonations are summarized in Table I. Exemplarily, the reconstructed physical positions of the sutures for one frame within R1 are illustrated in Fig. 4. For orientation in this figure, glottal airflow initiated inferiorly and proceeded superiorly through the hemiglottis. Zero in the ''lateral'' dimension denoted the glottal midline, or the position of the glass plate. The axis coordinates denoted distances from the chosen origin: the left upper corner of the cube at the glass plate; see Fig. 3. Maximum displacements, velocities and accelerations and decelerations occurred midway between the anterior and posterior extremes of the membranous vocal fold, near the superior extreme of the medial surface (Suture Nos. 12, 13, 14; see Fig. 4), the location where the mucosal wave is expected to have the largest displacement.[5] The largest components of displacements occurred in the lateral direction, and the smallest components occurred in the anterior–posterior direction, as previously reported.[13,42] The sutures with the smallest displacements were near the anterior–inferior edge of the medial surface (Suture Nos. 1, 2; see Fig. 4). For both phonatory utterances, the numerical values of extrema for displacements, velocities, and accelerations are summarized in Table II.

The influence of the 9.1% linearization error of the calibration on the dynamical variables of displacement, velocity, and acceleration was also computed. This was done by recomputing the dynamical variables after randomly varying

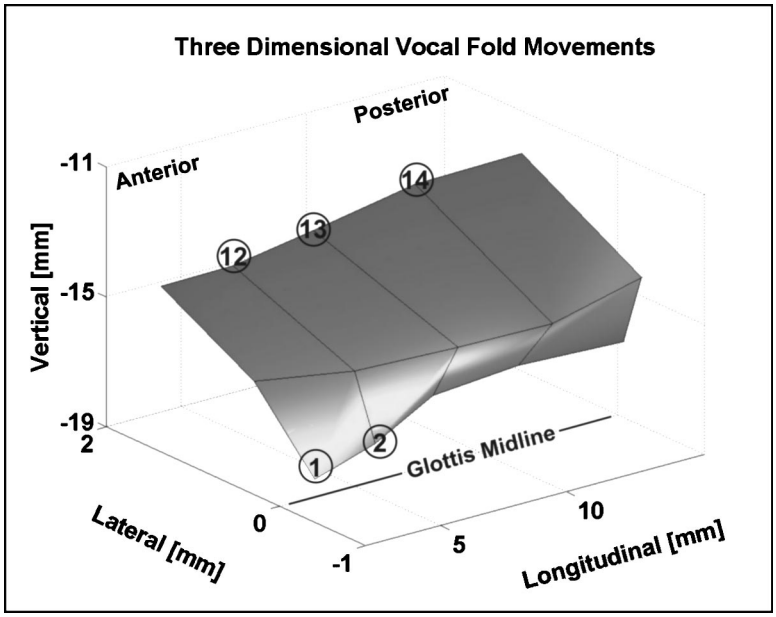

FIG. 4. A three-dimensional reconstruction of the vocal fold. The physical locations of extreme values are denoted by the highlighted sutures. The magnitudes of the corresponding extrema are enumerated in Table II. Zero on the lateral axis corresponds to the glottal midline or the position of the glass plate.

the entries of the mapping matrix F up to 6%, which matched the discrepancy to the former reported calibration error of 3.9%.[20] Across the two examples, the maximum deviations of the maximum displacements due to the increased linearization error were as follows: 0.11 (lateral), 0.07 (vertical), and 0.04 mm (anterior–posterior). The maximum discrepancies were 0.06 mm/ms for maximum velocity, and 0.02 mm/ms^2 for maximum acceleration and deceleration. The stability of the mapping matrix F against small distortions was achieved by using orthogonal basis vectors and an optimized condition number.[20] The condition number reflects the linear independency of the matrix entries. With small condition numbers (i.e., the minimal value is 1) in the mapping matrix F on the order of 3.2–3.6, small disturbances in F resulted in small disturbances in the image.[20]

Empirical Eigenfunctions, or the underlying modes of vibration, were also extracted from the two vibration sequences (see Table III). For R1, the largest EEF (EEF1) captured 69.5% of the energy, the second largest EEF (EEF2) captured 23.6% of the energy, and the third largest EEF (EEF3) 1.6% of the energy. Hence, the two largest EEFs captured 93.1% of the total energy, and the largest three EEFs captured 94.7%. The two dominant eigenfunctions entrained at a fundamental frequency of 220 Hz and the weaker eigenfunction captured the second harmonic at 440 Hz.

For R2, the largest EEF captured 60.2% of the energy, EEF2 captured 33.1% of the energy, and EEF3 2.5% of the energy. Hence, the two largest EEFs captured 93.3% of the total energy, and the largest three EEFs captured 95.8%. The two dominant eigenfunctions entrained at a fundamental frequency of 180 Hz, and the weaker eigenfunction vibrated at an independent frequency of 70 Hz.

At 33% TA activity (R1), the spatial EEFs are displayed in Fig. 5 from a frontal aspect, and in Fig. 6 from a sagittal aspect. In Fig. 5, the three horizontal rows correspond to the three different eigenfunctions EEF1 to EEF3. The five columns correspond to the five vertical columns of sutures on the vocal fold. Zero on the lateral axis indicates the glottal midline. Values larger than zero resulted from computational inaccuracies, which only affect the actual position of the suture, but not the vibrational patterns. Axes are referenced from a chosen origin (i.e., upper left corner of the cube). The arrows indicate the direction of the suture movement. EEF1 captured an alternating convergent/divergent glottis, e.g., the dotted line captured a relatively divergent glottis and the dashed line captured a relatively convergent glottis. Similar to previous studies, the convergent glottis was associated with a relatively elevated vertical position. EEF2 primarily captured the lateral in-phase movements of the medial surface. For EEF3, it was difficult to discern a clear vibration because of the small amplitudes of vibration. Within R2,

TABLE II. The computed extremal values of displacement, velocity, and acceleration, and their positions; see Fig. 4.

	Recording 1 (R1)		Recording 2 (R2)	
	Value	Suture	Value	Suture
Max. displacement longitudinal (mm)	0.32	13	0.28	12
Max. displacement vertical (mm)	1.36	13	1.30	14
Max. displacement lateral (mm)	1.47	13	1.57	13
Min. displacement longitudinal (mm)	0.03	1	0.07	2
Min. displacement vertical (mm)	0.06	1	0.24	2
Min. displacement lateral (mm)	0.02	1	0.20	2
Max. absolute velocity (mm/ms)	1.62	13	1.04	14
Max. absolute acceleration (mm/ms^2)	0.61	13	0.40	14
Max. absolute deceleration (mm/ms^2)	0.58	13	0.38	13,14

TABLE III. Percentage parts of the computed three largest Empirical Eigenfunctions and the corresponding frequencies.

	Recording 1 (R1)		Recording 2 (R2)	
	(%)	Frequency (Hz)	(%)	Frequency (Hz)
EEF1	69.5	220	60.2	180
EEF2	23.6	220	33.1	180
EEF3	1.6	440	2.5	70
Σ	94.7		95.8	

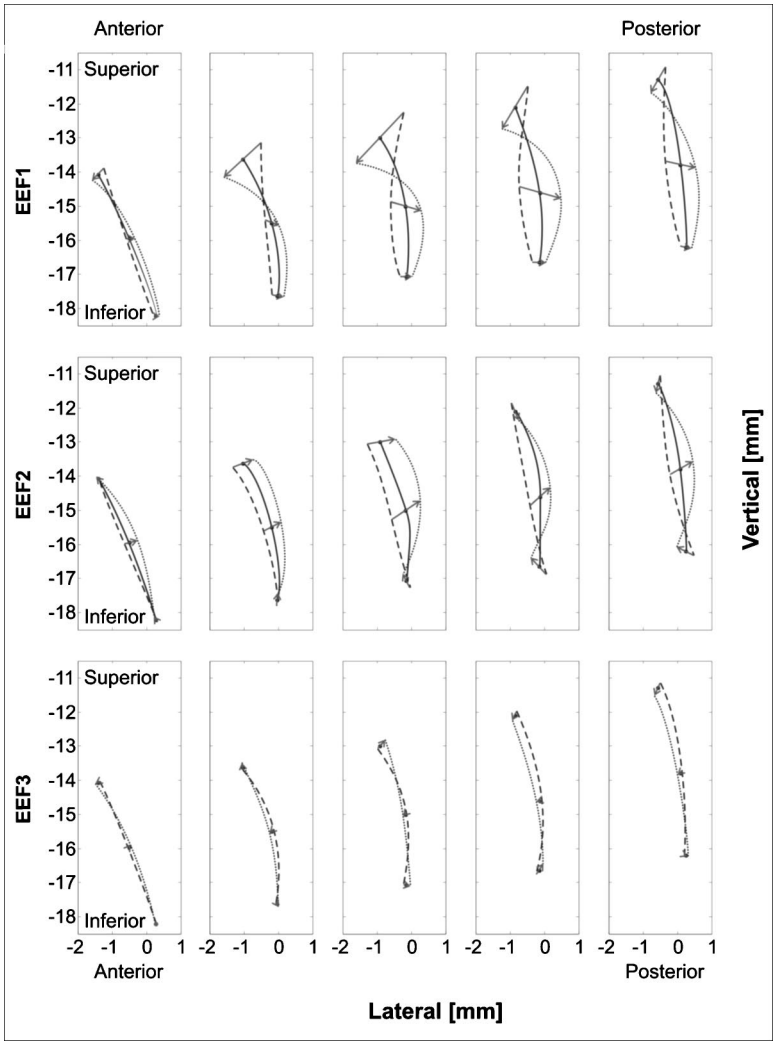

FIG. 5. Empirical Eigenfunctions for R1 (Suture Nos. 1–15): The maximum (dotted line) and minimum (dashed line) projections of the three largest EEFs about the equilibrium position (solid line) from a frontal aspect, for each of the five vertical columns of sutures. Zero on the lateral axis corresponds to the glottal midline. The units are measured from a chosen origin. The arrows indicate the direction of suture movement captured by the illustrated EEF. For EEF3, the mean value (solid line) has been omitted so as not obscure the maximum/minmum projections.

similar vibration and entrainment patterns were observed for EEF1 and EEF2.

While the 9.1% linearization error of the calibration procedure had a small impact on the computation of the dynamical variables, the linearization error had negligible impact on the computation of the eigenfunctions. The eigenvalues, which quantify the energy or the statistical variance captured by the empirical eigenfunctions, deviated by a maximum amount of 0.2% as a result of the linearization error. The dot product of EEF1 computed with and without the additionally induced linearization error was 0.998, also yielding an error of 0.2%. EEF2 yielded identical results. A similar dot product using EEF3 yielded 0.997, or an error of 0.3%. Theses results were consistent with previous reports, which showed that the computation of empirical eigenfunctions was a highly robust procedure, even in the presence of substantial background noise.[43]

Former computational studies[10,17] and experimental studies with excised larynges[16] have yielded similar underlying eigenfunctions for EEF1 and EEF2, with similar entrainment patterns, which are known to be critical for sustained oscillation. For example, the divergent glottis of EEF1 possesses a relatively low intraglottal pressure, while the alternating convergent glottis possesses a relatively high intraglottal pressure.[44] Thus, this eigenfunction correlated strongly with the intraglottal pressure. The elevated position of the convergent glottis was probably induced by the relatively high subglottal pressure associated with that glottal shape. On the other hand, EEF2 captured the net lateral movement of the medial surface.

For the fluid–structure interactions of concern, the transfer of energy from the airflow to the tissue would be optimized if the intraglottal pressure (governed by EEF1) were in-phase with the net lateral velocity of the vocal fold (governed by EEF2). Thus, the entrainment or synchronization of EEF1 and EEF2 would facilitate this transfer of energy, enabling sustained vocal fold oscillations.

Whether EEF1 must capture the alternating convergent/divergent glottal shape or whether it may switch roles with EEF2 is not entirely clear from the experiments to date. The present studies with an *in vivo* larynx, along with all previous computational studies[10,17] have indicated that EEF1 captures the alternating divergent/convergent glottis. On the other hand, in a former excised larynx experiment,[16] EEF1 captured the net lateral movement of the vocal fold, and EEF2 captured the alternating convergent/divergent glottis. So it may be that the roles of EEF1 and EEF2 may be reversed, but that both eigenfunctions are essential for sustained oscillation.

A comparison of the relative strengths of eigenfunctions

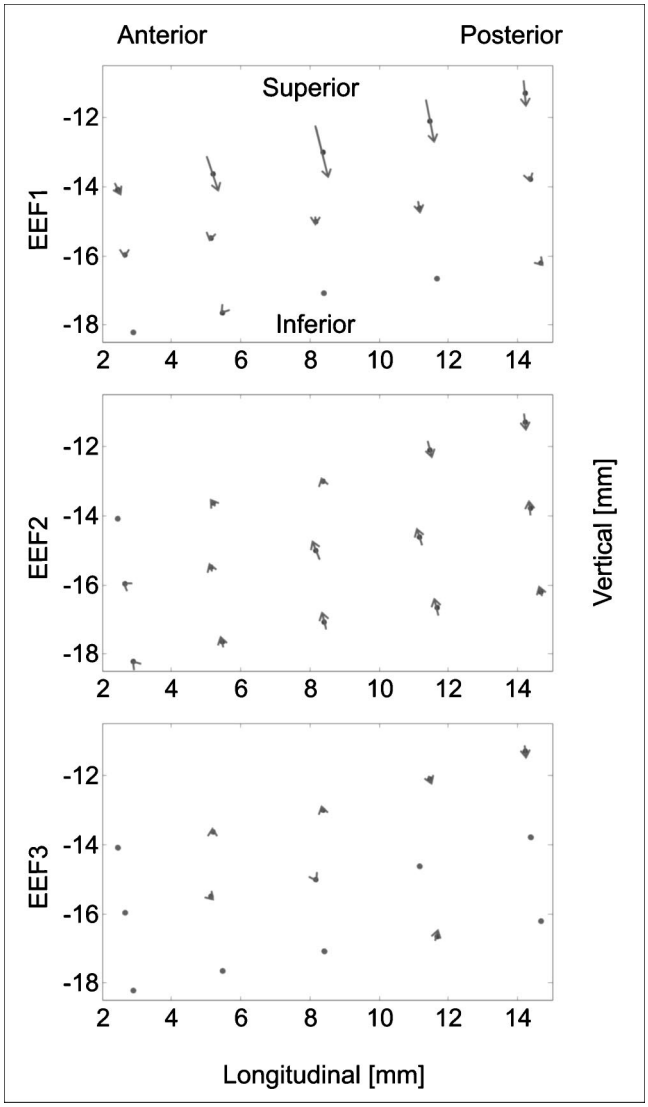

FIG. 6. The three largest EEFs for R1 from a sagittal aspect, for each of the five vertical columns of sutures (Suture Nos. 1–15). The units correspond to a chosen origin (i.e., the left upper corner of the cube). The arrows indicate the direction of the movement. The dots represent the mean values of the suture positions.

in several investigations is also of note. In clinical endoscopic studies of vocal fold vibration,[36] EEF1 captured 97% of the energy for normal phonation. In this study, it is important to note that only lateral movement, or one-dimensional vibrations were analyzed along the anterior–posterior length of the vocal fold. The EEF1 captured a half wavelength of a sinusoid along the anterior–posterior length of the folds, a characteristic feature of normal vocal fold vibration, as observed from a superior view. Higher-order anterior–posterior modes usually only become significant in an irregular voice.[36]

EEFs have also been extracted from finite element models of vocal fold vibration.[10,17] Because anterior–posterior vibrations were neglected in this model, two-dimensional vibrations were analyzed. In these studies, the two largest EEFs captured about 98% of the total energy. In an excised canine larynx experiment (normal phonation), which also reported two-dimensional vibrations (e.g., anterior–posterior vibrations were neglected here as well) the two largest EEFs again, captured 98% of the energy.[16] These results are to be compared with the present three-dimensional study in which the first two eigenfunctions explained about 93% of the energy and the first three eigenfunctions explained about 95% of the energy.

As just mentioned, previous theoretical and experimental studies have neglected anterior–posterior vibrations.[10,16,17] The improved spatial resolution (1024 $\times$1024 pixel) of the high-speed digital camera employed in this investigation enabled the study of full three-dimensional movement, including anterior–posterior vibrations. The results are shown in Fig. 6 from a sagittal view. For EEF1, the vertical movement of the superior row was clearly observable. In the lower two rows, little vertical or longitudinal movement was discernable. However, weak anterior–posterior vibrations did exist. Because EEF2 captured primarily lateral vibrations, both vertical and anterior–posterior vibrations were weak, but on the same order of magnitude. Again for EEF3, no unique behavior was discernable.

IV. CONCLUSIONS

This work reported the three-dimensional medial surface dynamics of an *in vivo* canine vocal fold as a function of RLN stimulation. The RLN stimulation current was not a robust input variable to differentiate the observed vibration patterns. However, an estimate of TA activity did reliably differentiate the vibration types, as did subglottal pressure. It should be noted that RLN innervation stimulates other muscles in addition to the TA, including the PCA (posterior cricoarytenoid), the LCA (lateral cricoarytenoid), and the IA (interarytenoid). However, because the PCA is an abductory muscle that tends to discourage phonation, this branch was clipped for this experiment. In addition, the IA branch was unavoidably clipped with the removal of one vocal fold to implement the hemilarynx procedure. Finally, if the LCA muscle did not adequately adduct the folds to produce phonation, additional external forces are applied to adduct the arytenoid. Thus, in our setup, the TA muscle was the primary muscle and perhaps the only laryngeal muscle dependent solley on RLN innervation.

Nevertheless, in future work, TA activity (or a similar measure such as the complex action potential) should be measured directly, rather than estimated. The hemilarynx methodology, previously designed for the excised larynx, was adapted for an *in vivo* canine larynx with thyroarytenoid muscle contraction. For the larynx in this study, chest-like vibrations were observed in the range of 12%–37% of TA activity (based on the percentage of maximum stimulation). Fry-like vibrations were observed in the range of 37%–100% of TA activity. During the chest-like vibrations, the fundamental frequency increased as a function of increasing TA stimulation. However, beyond a certain threshold in TA stimulation, an abrupt drop in fundamental frequency occurred resulting in fry-like vibration, an expected outcome for high TA and low cricothyroid (CT) stimulation (in the present study, since there was no SLN innervation, there was no presumably stimulation to the CT). This phenomenon could be explained in terms of existing theories of TA

function.[38–41] While the underlying modes of vibration were extracted and mechanisms of sustained oscillation were discussed, some difficulties were encountered (e.g., airflow leakage, and the inability to simultaneously image the superior and medial surfaces of the vocal fold). Nevertheless, areas with high mobility were identified along the medial surface of the vocal fold (Fig. 4). Furthermore, the EEFs were extracted and mechanisms of sustained oscillation were discussed, showing consistency with previous investigations of computational models[10,17] and excised larynx experiments.[16]

Future refinements need to be made in the imaging procedures in order to track sutures simultaneously on both the superior and medial surfaces of the vocal fold, as has been reported for the excised human larynx.[20] Indeed, this task has been implemented with little difficulty on excised human larynx studies, but was presumably more difficult here because of a small vocal lip that sometimes occurs on the canine vocal fold at the junction of the superior–medial surface.[45] Perhaps this difficulty in imaging the canine larynx can be circumvented by replacing the glass prism, with a more flexible set-up using mirrors. Other concerns included sealing airflow leakage in the posterior glottis, and properly hydrating the folds throughout the experiment. Future studies will also explicitly examine the observed fry-like vibrations of the *in vivo* canine vocal fold. Future studies will also covary several input variables, such as both RLN and SLN, which will allow other vocal registers, such as a falsetto, to be observed. It is expected that such studies will further illuminate the mechanisms of vocal production, particularly the poorly understood role of the thyroarytenoid muscle in phonation.[46] Ultimately, it would be desirable to extend these studies to independently estimate the mechanical stresses of the various layers of the vocal fold tissues (including the body, cover, and ligament) as a prelude to predicting tissue dynamics. However, at present, no reliable methods exist to estimate or measure such differential stresses in laryngeal tissue layers.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC03072 from NIH/NIDCD. The authors would also like to thank Norma Antonanzas-Barroso for her helpful comments.

[1] P. Mergell, H. Herzel, and I. Titze, ''Irregular vocal-fold vibration—high speed observation and modelling,'' J. Acoust. Soc. Am. **108**, 2996–3002 (2000).
[2] J. G. Svec, J. Horacek, F. Sram, and J. Vesely, ''Resonance properties of the vocal folds: *In vivo* laryngoscopic investigation of the externally excited laryngeal vibrations,'' J. Acoust. Soc. Am. **108**, 1397–1407 (2000).
[3] M. Döllinger, T. Braunschweig, J. Lohscheller, U. Eysholdt, and U. Hoppe, ''Normal voice production: Computation of driving parameters from endoscopic digital high speed images,'' Methods Inf. Med. **42**, 271–276 (2003).
[4] E. Schönhärl, *Die Stroboskopie in Der Praktischen Laryngologie* (Thieme-Verlag, Stuttgart, 1960).
[5] I. R. Titze, ''Parametrization of the glottal area, glottal flow, and vocal fold contact area,'' J. Acoust. Soc. Am. **75**, 570–580 (1984).
[6] Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissues* (Springer-Verlag, New York, 1993).
[7] R. Wilhemls-Tricarico, ''Physiological modeling of speech production: methods for modeling soft tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098 (1995).
[8] P. Mergell and H. Herzel, ''Modelling biphonation—the role of the vocal tract,'' Speech Commun. **22**, 141–154 (1997).
[9] M. P. de Vries, H. K. Schutte, and G. J. Verkerke, ''Determination of parameters for lumped parameter models of the vocal folds using a finite-element method approach,'' J. Acoust. Soc. Am. **106**, 3620–3628 (1999).
[10] F. Alipour, D. A. Berry, and I. R. Titze, ''A finite element model of vocal fold vibration,'' J. Acoust. Soc. Am. **108**, 3003–3012 (2000).
[11] E. J. Hunter, I. R. Titze, and F. Alipour, ''A three-dimensional model of vocal fold abduction/adduction,'' J. Acoust. Soc. Am. **115**, 1747–1759 (2004).
[12] T. Baer, ''Investigation of phonation using excised larynges,'' Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA, 1975.
[13] S. Saito, H. Fukuda, S. Kitahira, Y. Tsuzuki, H. Muta, E. Takayama, T. Fujika, N. Kokawa, and K. Makino, ''Pellet tracking in the vocal fold while phonating: experimental study using canine larynges with muscle activity,'' in *Vocal Fold Physiology*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO, 1985), pp. 169–182.
[14] J. J. Jiang and I. R. Titze, ''A methodological study of hemilaryngeal phonation,'' Laryngoscope **103**, 872–882 (1993).
[15] G. M. Gardner, J. Castracane, M. Conerty, and S. M. Parnes, ''Electronic speckle pattern interferometry of the vibrating larynx,'' Ann. Otol. Rhinol. Laryngol. **104**, 5–12 (1995).
[16] D. A. Berry, D. W. Montequin, and N. Tayama, ''High-speed digital imaging of the medial surface of the vocal folds,'' J. Acoust. Soc. Am. **110**, 2539–2545 (2001).
[17] D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, ''Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions,'' J. Acoust. Soc. Am. **95**, 3595–3604 (1994).
[18] G. S. Berke and B. R. Gerratt, ''Laryngeal biomechanics: An overview of mucosal wave mechanics,'' J. Voice **7**, 123–128 (1993).
[19] F. Alipour-Haghighi and I. R. Titze, ''Elastic models of vocal fold tissues,'' J. Acoust. Soc. Am. **90**, 1326–1331 (1991).
[20] M. Döllinger and D. A. Berry, ''Computation of the three-dimensional medial surface dynamics of the vocal folds,'' J. Biomech. (in press).
[21] M. Döllinger and D. A. Berry, ''Three-dimensional medical surface dynamics of the vocal folds using high-speed digital imaging,'' in *Proceedings of the 26th World Congress of the International Association of Logopedics and Phoniatrics* (Speech Pathology Australia, Melbourne—Australia, 2004), p. CD.
[22] G. S. Berke, D. M. Moore, P. A. Monkewitz, D. G. Hanson, and B. R. Garrett, ''A preliminary study of particle velocity during phonation in an in vivo canine model,'' J. Voice **3**, 306–313 (1989).
[23] S. Bielamowicz, G. S. Berke, J. Kreiman, and B. R. Gerratt, ''Exit jet particle velocity in the in vivo canine laryngeal model with variable nerve stimulation,'' J. Voice **13**, 153–160 (1999).
[24] M. Hirano, ''Structure and vibratory behavior of the vocal folds,'' in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and S. Franklin (Univ. of Tokyo, Tokyo, 1977), pp. 13–30.
[25] Y. I. Abdel-Aziz and H. M. Karara, ''Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry,'' in *Proceedings of the Symposium on Close-Range Photogrammetry* (American Society of Photogrammetry, Falls Church, VA, 1971), pp. 1–18.
[26] J. Lohscheller, M. Döllinger, M. Schuster, U. Eysholdt, and U. Hoppe, '' The laryngectomee substitute voice: Image processing of endoscopic recordings by fusion with acoustic signals,'' Methods Inf. Med. **42**, 277–281 (2003).
[27] N. R. Miller, R. Shapiro, and T. M. McLaughlin, ''A technique for obtaining spatial kinematic parameters of segments of biomechanical systems from cinematographic data,'' J. Biomech. **13**, 535–547 (1980).
[28] H. Hatze, ''High-precision three-dimensional photogrammetric calibration and object reconstruction using a modified DLT-approach,'' J. Biomech. **15**, 11–19 (1988).
[29] L. Chen, C. W. Armstrong, and D. D. Raftopoulos, ''An investigation on the accuracy of three-dimensional space using the Direct Linear Transformation,'' J. Biomech. **27**, 493–500 (1994).
[30] Y. Xunhua and L. Ryd, ''Accuracy analysis for RSA: a computer simulation on 3D marker reconstruction,'' J. Biomech. **30**, 493–498 (2000).
[31] G. A. Wood and R. N. Marshall, ''The accuracy of DLT extrapolation in three-dimensional film analysis,'' J. Biomech. **19**, 781–785 (1986).

[32] A. M. T. Choo and T. R. Oxland, ''Improved RSA accuracy with DLT and balanced calibration marker distributions with an assessment of initial-calibration,'' J. Biomech. **36**, 259–264 (2003).
[33] W. Murray, *Numerical Methods for Unconstrained Optimization* (Academic, London, 1972).
[34] G. Golub and C. V. Loan, *Matrix Computations*, 2nd ed. (The John Hopkins University, Baltimore, MD, 1989).
[35] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt, ''Vocal fold parameter extraction using the two-mass-model,'' IEEE Biomed. Eng. **49**, 773–781 (2002).
[36] J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel, ''Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes,'' J. Acoust. Soc. Am. **110**, 3179–3192 (2001).
[37] S. N. Nasri, P. Dulguerov, J. Damrose, M. Ye, J. Kreiman, and G. S. Berke, ''Relation of recurrent laryngeal nerve compound action potential to laryngeal biomechanics,'' Laryngoscope **105**, 639–643 (1995).
[38] M. Hirano, ''Morphological structure of the vocal cord as a vibrator and its variations,'' Folia Phoniatr. **26**, 89–94 (1974).
[39] M. Hirano, ''Phonosurgery: Basic and clinical investigations,'' Otologia (Fukuoka) **21**, 239–440 (1975).
[40] I. R. Titze, J. Jiang, and D. Druker, ''Preliminaries to the body-cover theory of pitch control.,'' J. Voice **1**, 314–319 (1988).
[41] I. R. Titze, *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ, 1994).
[42] T. Baer, ''Investigation of the phonatory mechanism,'' ASHA Report 11, 1981, pp. 38–47.
[43] K. S. Breuer and L. Sirovich, ''The use of the Karhunen–Loeve procedure for the calculation of linear eigenfunctions,'' J. Comput. Phys. **96**, 277–296 (1991).
[44] I. R. Titze, ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1551 (1988).
[45] T. Riede, H. Herzel, D. Mehwald, W. Seidner, E. Trumler, G. Böhme, and G. Tembrock, ''Nonlinear phenomena in the natural howling of a dog–wolf mix,'' J. Acoust. Soc. Am. **108**, 1435–1442 (2000).
[46] I. Sanders, Y. Han, J. Wang, and H. Biller, ''Muscle spindles are concentrated in the superior vocalis subcompartment of the human thyroarytenoid muscle,'' J. Voice **12**, 7–16 (1998).