Published in final edited form as:

J Voice. 2013 July ; 27(4): 463–472. doi:10.1016/j.jvoice.2013.03.004.

***In Vivo* Measurement of Pediatric Vocal Fold Motion Using Structured Light Laser Projection**

Rita R. Patel^{*}, Kevin D. Donohue[†], Daniel Lau[†], and Harikrishnan Unnikrishnan[†]

^{*}Department Speech and Hearing Sciences, College of Arts and Sciences, Indiana University, Bloomington, Indiana

[†]Center for Visualization and Virtual Environments, College of Engineering, University of Kentucky, Lexington, Kentucky

Summary

Objective—The aim of the study was to present the development of a miniature structured light laser projection endoscope and to quantify vocal fold length and vibratory features related to impact stress of the pediatric glottis using high-speed imaging.

Study Design—The custom-developed laser projection system consists of a green laser with a 4-mm diameter optics module at the tip of the endoscope, projecting 20 vertical laser lines on the glottis. Measurements of absolute phonatory vocal fold length, membranous vocal fold length, peak amplitude, amplitude-to-length ratio, average closing velocity, and impact velocity were obtained in five children (6–9 years), two adult male and three adult female participants without voice disorders, and one child (10 years) with bilateral vocal fold nodules during modal phonation.

Results—Independent measurements made on the glottal length of a vocal fold phantom demonstrated a 0.13 mm bias error with a standard deviation of 0.23 mm, indicating adequate precision and accuracy for measuring vocal fold structures and displacement. First, *in vivo* measurements of amplitude-to-length ratio, peak closing velocity, and impact velocity during phonation in pediatric population and a child with vocal fold nodules are reported.

Conclusion—The proposed laser projection system can be used to obtain *in vivo* measurements of absolute length and vibratory features in children and adults. Children have large amplitude-to-length ratio compared with typically developing adults, whereas nodules result in larger peak amplitude, amplitude-to-length ratio, average closing velocity, and impact velocity compared with typically developing children.

Keywords

High speed laryngeal imaging; Pediatric voice; Laser endoscopy; Vocal fold vibrations

INTRODUCTION

Clinically, endoscopic laryngeal imaging techniques, such as stroboscopy, video kymography, and high-speed digital imaging, provide direct assessment of vocal fold structure and function, critical for evaluation and treatment of voice disorders. A complete understanding of the body-cover relationship¹ is vital for clinically reconciling vocal fold pathology with patient symptoms. However, limited empirical investigations of vocal fold

vibratory motion exists in the pediatric population.² Quantitative characterization of vocal fold motion can greatly enhance our understanding of the clinical impact of growth and development in children. Additionally, data from quantitative *in vivo* assessment of vocal fold length and vibratory features can be used to advance our knowledge of laryngeal physiology and add to the theories of voice, especially for normal and disordered pediatric voice production.

Data on vocal fold motion from adults are valuable but cannot be used for direct interpretation of pediatric phonation, as laryngeal anatomy and the vocal fold layered structures^{3–6} in the pediatric population differ considerably when compared with adult. Although phonatory features of amplitude and closing velocity have been used to quantify voice production in adults,^{7–11} these features have been understudied in children. Clinically, expectations about pediatric laryngeal function cannot be based on models of adult physiology as there are known differences not just in scale but anatomical/functional differences in the respiratory^{12–15} and laryngeal systems^{2,16} used in voice production in children. Age-specific models of vocal fold oscillations need to be developed that take into account the kinematic motion resulting from the unique less-differentiated multilayered vocal fold structure. High-speed digital imaging coupled with laser projection provides the first opportunity to quantify vocal fold motion as it relates to the kinematic activity because of increased temporal resolution and nonreliance on the estimated fundamental period to track vibratory motion.^{17–19}

Investigations thus far using custom-made laser systems with rigid and flexible endoscopes²⁰ have demonstrated that objective measurements of vocal fold structure and function can be obtained when coupled with stroboscopy,^{11,21} kymography,^{9,22} and high-speed imaging^{7,10,23,24} in adults. Except for Larsson and Hertegard⁷ and Popolo and Titze,¹¹ the laser projection systems were coupled with a 90° rigid endoscope, compared with the 70° rigid endoscope, which is a standard in the United States. These measurements use the principle of laser triangulation in which patterns like dots, lines, or a regular grid of dots are projected onto the vocal folds from a point offset from the endoscope camera. The earlier systems using a single laser beam^{7,8,25} were difficult to maneuver and target the laser dot onto the vocal folds. The laser dot pattern was also difficult to distinguish from the specular reflections common in the endoscopic high-speed recordings. A horizontal line, which is a more distinguishable pattern than dots, was projected by George et al⁹ to record three-dimensional (3D) vibratory motion using kymography. An alternate approach that simplifies the calibration is to use the distance between two laser dots as the reference.^{10,11,24,26,27} Because the two laser dots projected at a fixed distance apart do not diverge significantly, the distance between the dots in the image can be used as a ruler. The two-dot systems have the same difficulty of the laser spot detection and targeting as the one-dot instruments. A system consisting of two horizontal parallel laser lines was used by Wurzbacher et al²³ to quantify vocal fold dynamics from high-speed imaging with improved robustness over the single dot systems. The tilt of the imaging plane compared with the optical axis was the main source of error for all the techniques that apply the parallel patterns. In work by Patel et al²⁶ using irregular shaped laser dots, the tilt of the vocal fold imaging plane in relation to the endoscope optical axis was accounted by calibrating at different tilts against the dot distortions on the imaging plane. The laser projecting endoscope was calibrated by imaging a grid pattern at depths 6–10 cm and tilt angles of 15°–25°. The laser patterns obtained during the calibration process were used as templates for pattern matching with the *in vivo* recordings. This first laser endoscope for measurement of pediatric phonation²⁶ allowed for the development of the calibration procedure to compensate for variations in endoscopic distances and projection angles. It was applied to a 9-year-old child during modal phonation to measure vocal fold length (6.84 ± 0.0002 mm), membranous length (3.66 ± 0.0004 mm), and vibratory amplitude (0.29 ± 0.06 mm).²⁶ However, the laser housing added 5 mm to the

existing endoscope causing some discomfort and difficulty in positioning the laser dots on a relatively flat and small area of the pediatric vocal folds. Contrary to the two laser dots, multiple laser stripes would allow for a complete and dense coverage of the vocal fold structures without extra effort on part of the clinician to aim the pattern on regular surfaces of the vocal folds. Relative to all other orientations, vertical lines change pixel positions as a function of camera distance from the image plane, thereby maximizing the sensitivity with changes in depth, resulting in more accurate pixel density estimates. The periodic pattern of multiple stripes also serves as a special coding, where stripes can be enhanced over structures in the background image through filtering to improve pixel location accuracy of the projected stripes.

This article presents the development of a novel miniaturized custom-built laser system using multiple vertical lines coupled with a 70° rigid endoscope for clinical measurement of pediatric vocal fold motion. The device performance was assessed through accuracy and precision estimates on vocal fold phantoms with known dimensions. Furthermore, *in vivo* measurements on human subjects were performed to demonstrate feasibility and potential clinical significance. *In vivo* measurements of phonatory length, membranous length, peak amplitude, amplitude-to-length ratio, average closing velocity, and impact velocity were obtained using high-speed laryngeal imaging of typically developing children, adult male and female, and one child with vocal fold nodules during modal phonation.

METHODS

Laser projection device

The laser projection device was custom-built (Green Light Optics, Cincinnati, OH) to achieve the critical patterns on the expected depth range of the image plane while limiting the extension of the device to 4 mm when mounted on the side of the endoscope. The device consists of three main components: (1) the laser, (2) beam tube, and (3) optics module (Figure 1). The main case houses the laser consisting of a 300-mW green laser, electrical power input plug, and an on/off switch. The hollow beam tube of 150.5 mm in length and 3 mm in diameter allows the laser to transmit from the main case to the optics module without interference. The optics module is mounted in parallel to the tip of the endoscope and contains the Ronchi grating and projection optics. The optics module consists of five elements: (1) the glass plate, which is 1.5 mm thick and 4 mm in diameter; (2) field stop of spring steel 0.127 mm thick and 4 mm diameter with a wire Electrical Discharge Machining (EDM) cut with 0.8×0.5 opening offset by 0.273 mm from center; (3) a lens of optical grade acrylic of 0.87-mm center thickness with 4 mm diameter, which is a convex sphere on the entrance and a concave sphere on the exit surface; (4) aperture stop, which is a spring steel of 0.217 mm thickness and 4 mm diameter using an EDM cut with 0.44-mm diameter opening; and (5) prism lens of optical grade acrylic has a 10-mm base diameter with a 45° hypotenuse, a concave sphere exit surface, and a plano entrance surface.

The light from the laser is transferred to the diffuser through the beam tube. The diffuser spreads the beam by 30° making the beam wide enough for the Ronchi grating deposited on the glass substrate to create the vertical stripe patterns. The light coming out from the beam stop is modulated with stripes creating a rectangular cross section of 0.8×0.5 mm. The structured light is diffused further and reflected perpendicularly by the prism lens. The offset between the light projection and endoscope viewing cause the projection and the viewing field to cross each other. The projected image is centered within the endoscope field of view at a distance of 30 mm from the exit window. The projected pattern shifts horizontally as the image plane distance changes. The projection consists of 21 symmetric and linearly placed vertical light strips on the vocal fold plane (Figure 2). Each individual stripe has a width of 0.3 mm and the pattern period of 1 mm. At a depth of 30 mm, the laser energy distribution is

1.1 mW/mm² in the absence of any optical losses of the projected pattern. Therefore, the maximum exposure would always be lesser than this level.

Given the constraints of obtaining endoscopic imaging in children (resulting from smaller sizes and lower tolerance), multiple stripes were used to allow for a complete and dense coverage of the vocal fold structures without extra effort on the part of the clinician to aim the pattern on regular surfaces of the vocal folds. The vertical lines were used so that the projection system could be mounted to the side of the endoscope, thereby increasing the dimension in a direction least likely to increase the discomfort. For this orientation between the projector and camera, horizontal lines show little change as a function of depth, whereas the vertical lines exhibit a maximum shift in response to depth changes. The larger the pattern shift across pixels, the more likely the change in depth will be detected with an improved robustness to pixel location errors.

Data collection

Adult and children participants were recruited from Institutional Review Board (IRB)-approved advertisements and fliers placed around the University of Kentucky Campus and at University of Kentucky Children's Hospital. A total of 11 volunteer participants were recruited for the study after signing an IRB-approved informed consent/assent forms, at the University of Kentucky, Vocal Physiology and Imaging Laboratory. Participants without voice disorders were included in the study if they met the following criterion: had negative histories of vocal pathology, were not professional voice users, and were perceptually judged to have normal voice by a certified speech-language pathologist specializing in voice disorders. Children experiencing puberty as identified via case history were excluded. Selection criteria for adult controls were similar to those of pediatric group, except that the adult controls had a negative history of smoking. Ten participants without voice disorders were included in the study: five children (age range: 6–9 years; three girls and two boys) and five adults (age range: 21–45 years, three females and two males). One female child, aged 10 years, with bilateral vocal fold nodules was included in the study.

Sustained phonation on the vowel /i/ was recorded at participant's typical pitch and loudness. Typical voice production was defined as phonation that is close to the average speaking fundamental frequency and loudness for individual subjects. The examiner judged the level of typical voice production through perceptual judgments of sustained phonation and conversation sample. The recordings were performed by a digital gray scale KayPENTAX high-speed system model 9710 with a sampling rate of 4000 frames per second with a spatial resolution of 512 × 256 pixels for a maximum duration of 4.094 seconds and at 60 frames per second with the laser endoscope. Children tolerated the laser endoscope well, however required considerable pre-endoscopic preparation, typical of endoscopic examination in routine clinical practice. The camera was coupled to a 70° KayPENTAX endoscope. Simultaneous acoustic signal was recorded at 50 kHz. Acoustic recording was used to confirm the presence of steady state phonation and participants' task performance of typical pitch and loudness.

Physical measurement estimation

Calibration—Calibration was performed to estimate the distance of the vocal folds from the endoscope by associating the laser stripe positions with endoscope camera depth. Information from the calibration procedure is then used to convert pixel dimensions to millimeters. The calibration is a one-time process for each device and need not be repeated for each recording. The calibration apparatus (Figure 3) consists of a calibration grid mounted on a stepper motor, where the calibration grid was a regular pattern of dots separated by 5 mm. The motor positions the grid through a rail at high-precision steps via a

computer-controlled interface. Figure 3A shows the stripes projected on the grid, and Figure 3B shows the endoscope resting on a stable support with the calibration grid mounted on the stepper motor rail. The stripe edge pixel locations and image pixel to real-world coordinate correspondences are linearly dependent on the depth. These linear relationships are established by imaging the calibration grid at four depths within the expected range for children's vocal fold. A more detailed description of the calibration process is provided in the Appendix.

Procedure for measurements on the vocal folds—Stripe contrasts required for reliable edge detection were obtained for lower video speeds. Therefore, the sizing of the invariant vocal fold features was performed at lower speeds to create pixel-to-millimeter conversion factors in the higher speed videos at 4000 frames per second. Pixel coordinates at multiple points along the stripe edge are selected by a human observer using an interactive image interface for a low-speed image frame with projected stripes. This interface allows for frame selection based on the clarity of the features needed for a glottal length estimate. The depth of the vocal folds is subsequently estimated from the resulting edge coordinates. The depths along with pixel locations are combined to get the physical measurements (x and y). The pixel-to-millimeter conversion factor for the image plane is estimated from averaging the length ratios between all pairwise points. The glottal length is then marked in both 60 and 4000 frames per second recordings. The pixel-to-millimeter conversion factor for 4000 frames per second is computed by the following equations:

$$L_{\text{mm},60} = \gamma_{60} \times L_{\text{pix},60} \quad (1)$$

$$\gamma_{4000} = \frac{L_{\text{mm},60}}{L_{\text{pix},4000}}, \quad (2)$$

where L is the glottal length and subscripts denote the frame rate (60 or 4000 frames per second) and the unit of length (pixels or millimeter). The displacement and length features are measured in pixels and converted to millimeters by multiplying with 4000.

To further enhance the stripe contrast against the anatomical image, band-pass filtering was applied based on the spatial periodicity of the pattern. Because the stripe patterns are relatively stationary over sequential frames, a number of frames were averaged together to enhance the stripe contrast against the underlying anatomical image. The averaged image was subsequently band-pass filtered in the horizontal direction over the range of the expected stripe frequencies (5–10 pixels per cycle). An example of the stripe-enhanced image is shown in Figure 4. The sensitivity of the stripe position to depth in the Figure 4 is observed by noting the stripes above the vocal folds, which are on tissues closer to the camera in depth, causing a horizontal shifting from the stripes that fall on the folds. A right edge of a stripe is seen on the right fold in this example (Figure 4), which is used to determine depths and the pixel-to-millimeter conversion. The image can be scaled and enhanced by human observer to improve identification of edge pixels over the surface of interest.

Precision and accuracy

The precision and the accuracy of the device were computed by performing measurements on a custom vocal fold phantom. The glottal length on the model was measured as 17.81 mm using an electronic caliper. The image of the phantom model kept approximately at 40 mm from the endoscope was captured with 15 independent recordings. The phantom model as recorded by the laser projection endoscope is shown in Figure 5. The red dots on Figure 5A are the user-selected estimate of the location of the laser stripes. The red line on Figure

5B represents the user input for the glottal length. The depth and the physical dimensions are subsequently computed using Equations A1 and A4 of the Appendix, respectively. The bias or the deviation from the expected result is the mean difference between the estimated glottal lengths and the actual length. The variability, denoting the repeatability, is represented by the standard deviation (SD) of the glottal length estimations.

Quantitative measurement of vibratory features

The laser projection system was used to estimate physical units for phonatory length,²⁸ membranous length,²⁸ peak vibratory amplitude, amplitude-to-length ratio, peak closing velocity, and the impact velocity. Peak vibratory amplitude was defined as the maximum horizontal excursion of the vocal folds from the medial glottal line corresponding to the peak opening phase of the glottal cycle. Phonatory length and membranous length of the vocal folds were calculated in millimeters based on Hirano et al²⁸ anatomical definition. Points to calculate the phonatory and membranous lengths of the vocal folds were marked during the maximum open phase of the phonatory cycle, where the vocal folds are in a slightly open position. Amplitude-to-length ratio is defined as the ratio of peak vibratory amplitude to the phonatory vocal fold length.²⁹ The peak velocity attained during the closing phase of the vocal fold oscillation and the velocity at the first instance of vocal fold closure (impact velocity) was also computed. Measurement of peak vibratory amplitude, amplitude-to-length ratio, peak closing velocity, and impact velocity were calculated from the mid-point of the membranous portion of the vocal fold. The displacement waveforms were denoised and interpolated/extrapolated over every open cycle to estimate the closing and opening instants with added accuracy. Details regarding the image processing algorithm and feature extraction are presented in Unnikrishnan et al.³⁰ The features computed for the left and right vocal folds are averaged to obtain a representative value.

RESULTS

Phonatory vocal fold length, membranous vocal fold length, amplitude-to-length ratio, peak amplitude, peak closing velocity, and impact velocity at modal phonation were calculated for five children and five adults without voice disorders and for one child with bilateral vocal fold nodules. The average vocal fold length during phonation for adult males was 14.95 mm, for adult females was 11.38 mm, and for children was 6.35 mm (Table 1). The membranous length of the vocal folds, as expected was smaller than the phonatory length of the vocal fold (Table 1). The mean peak amplitude was smallest for typically developing children (0.31 mm), compared with adult females (0.64 mm) and adult males (1.00 mm). Children exhibited comparably large amplitude-to-length ratio (0.05) compared with adult male and female (0.06) participants. Similarly, average closing velocity was smallest for typically developing children compared with adult females and adult males (Table 1). The child with vocal fold nodules exhibited large peak amplitude (0.53 mm), amplitude-to-length ratio (0.10), average closing velocity (0.48 meters per second), and impact velocity (0.68 meters per second).

The measurement precision and accuracy calculated using the phantom vocal fold model revealed an error of +0.13 mm and SD of ± 0.23 mm over the multiple measurements. For scans performed in this work, the pixel size was approximately 0.045 mm. Therefore, the error was approximately three times the pixel size suggesting that the error and variability have some contribution from the glottal length pixel selection as well as from imperfections of the laser stripes and calibration process. A variability of 5 pixels can be expected based on the clarity of the pixel edges. This level of variability accounts for most of the variability in the measurement process. The percent error for estimating the glottal length is 0.77% with SD on the order of 1.34%. This measurement is critical for the pixel-to-millimeter conversions on the high-speed video because a similar level of error can be expected for the

measurements of vibratory features. The bias error suggests that measurement performed with this device is overestimated approximately 1%, most of which is masked by the variability from the measurement precision.

DISCUSSION

The results obtained from human subjects demonstrate the feasibility of proposed laser projection system for obtaining absolute measures of the vocal fold kinematics in the pediatric population. The use of vertical lines allowed for the projection system to be mounted in such a manner, so as to add only 4 mm to the endoscope diameter resulting in little to no additional discomfort during the imaging process. This is especially important for use in the smaller pediatric larynx, where comfort of the subjects is critical for completing the scan. The usage of the device is not significantly different from that used in conventional endoscopic examinations. Because the stripes are present in all parts of the image, special maneuvering is not required to bring the stripes closer to the vocal fold. In addition, the extended periodic pattern allows for spatial filtering to enhance the projected pattern contrast, thereby reducing the edge detection error and reducing the variability from ambiguous pixel selection.

Quantitative measurements of vocal fold lengths, vibratory amplitudes, and closing velocity during phonation for adults were similar to those reported in the literature.^{7,9,11,24} The measures of vibratory length and vibratory amplitude in typically developing children are similar to the value reported in our previous article.²⁶ However, the reported values here are from a larger number of participants ($n = 5$) compared with a single subject in our previous work. As expected, children showed shorter vocal fold length and reduced vibratory amplitude (Figure 6) than the adults.

The results represent the first report of *in vivo* measurements of amplitude-to-length ratio, average closing velocity (meters per second), and impact velocity (meters per second) in children and in one child with vocal fold nodules. The study finding of comparably large amplitude-to-length-ratio in children compared with adults lend preliminary evidence to the statement that children have “difficulty making pitch and loudness adjustments independently of each other.”²⁹ (p. 181) Large amplitude-to-length ratio along with high lung pressure¹³ in children would predispose the children to struggle to sing a crescendo at high pitch.²⁹ The average closing velocity for typically developing children is less than that of adult males or females. This is expected as children’s vocal folds travel less distance per cycle. The child with nodules has a higher average closing velocity compared with typically developing children. However, the impact velocity for typically developing children is approximately similar to the adult values. The measurement of impact velocity is an indicator of the force applied on the vocal folds by each other during collision (contact), suggesting that typically developing children have similar impact velocity compared with adults. Assuming more or less the same material property for the vocal fold tissue for all the children, higher impact velocity could correspond to higher impact stress. Comparable value of impact velocity as adults could suggest that typically developing children could be at high risk for developing vocal fold lesions like vocal fold nodules related to high-impact stress.³¹ Vocal fold nodules are reported to be the most frequently occurring vocal fold pathology in *school age* children (7–16 years) with the overall prevalence of 30.2%.³²

The child with the vocal fold nodules has larger peak vibratory amplitude and average closing velocity compared with typically developing children and large amplitude-to-length ratio and impact velocity compared with both typically developing children and adults in the study. The child with nodules achieved similar instantaneous velocity (Figure 7) compared with the adult males. These are the first *in vivo* findings in a child with vocal fold nodule,

quantifying impact stress. Large group studies quantifying these kinematic features with the use of laser projection with high-speed digital imaging are warranted to establish empirical relationship between the measures of impact stress and size of vocal nodules in children. Results from this study also suggest that impact velocity, estimated through the information provided by the laser endoscope and image processing methods during the closing phase of the glottal cycle may provide better estimates of impact stress compared with average closing velocity. Future studies on quantitative measurements of such vibratory features may lead to early identification and development of physiologically based treatment approaches for early identification and management of pediatric dysphonia.

Measurement accuracy and precision can be affected by factors related to: (1) the laser calibration process and (2) the vocal fold edge estimation. The identification of glottal lengths by the human observer, at high speed and at low frame rates, for measurement of vocal fold displacement leads to a level of variability. However, the variability from pixel selection is common to all methods, even automatic methods. High resolution and distinct spatial properties of the features can reduce this variability. For the estimates reported in this work, the variability from pixel selection was less than 5 pixels, which is approximately equal to the SD of the phantom measurements. The impact of this variability is reduced by using a large structure for the reference, such as the vocal fold length, which covers many pixels (200–400) along the region most critical for estimating the vocal fold edge displacement. The result is a more robust measure of the pixel density compared with measuring smaller structures or those further away from the structures of interest. Another source of variability related to the laser calibration process comes from estimating the depth of the vocal folds from the endoscope. The proposed method compensates for changes in depth between frames; however, it does assume that all pixels of interest lie in the same plane. Because there is motion in the vertical direction of the vocal fold edges during phonation as well as some variations in the angle between the camera and the image plane, not all the pixels of interest lie in the same plane. Results from the calibration indicate an approximate 2% change in the pixel-to-millimeter conversion factor for every millimeter the fold structure extends in the vertical direction outside of the glottal length plane. If the extensions in the depth are limited to ± 2.5 mm for any given frame, then error from out of plane extension is less than $\pm 5\%$. In addition, the features measured in this work were averaged over 30 cycles further reducing the variability resulting from individual measurements.

The most significant limitation is the low contrast of the projection pattern at high frame rates. The additional step for a low-resolution scan to measure the glottal length confounds the scanning procedure. If patterns could be seen at the higher frame rate, it would not only simplify the scanning procedure, it would also allow for 3D motion estimation. Although laser projections have the advantage of maintaining consistent patterns over changes in depth, it would not be practical to increase the power beyond its current level. However, a calibration procedure can potentially be extended to compensate for changes occurring with broadband light sources. Depth can be tracked in a similar manner by tracking horizontal movement of the projected vertical edges and computing the depth directly for each frame and pixel.

Other sources affecting accuracy and precision of measurement related to edge estimation include the finite spatial resolution (pixel size) and discrete temporal sampling (frame rate). For fundamental frequencies of 277 Hz (typical of 6-year-old children)³³ at 4000 frames per second, each sample interval represents 6.9% of the total cycle resulting in an expected error of 6.9% while estimating the times for the vocal fold opening, closing, or the duration of the peak extension. The measurement accuracy is also reduced due to the size of pixels. The peak pixel displacement for the recordings is in the range of 8 (children) to 14 pixels

(males). Hence, each pixel represents 12.5–7% of the peak amplitude. The edge detection algorithm does not significantly add to these errors, especially at the mid-glottal region, where the algorithm has been established to have subpixel accuracy.³⁰ Another limitation of the study is the size of the attached laser sleeve. Although the attachment only added 4 mm to the existing endoscope, it is still large to image children across all age ranges, including infants.

An important contribution of this article in terms of the laser endoscope device is the application of spatial filtering to enhance the contrast of the projected periodic pattern. This can be applied to any technique for reducing the variability of identifying pattern edges or centers on the pixel grid. A higher pattern contrast through filtering can also enable the use of automatic edge detection and tracking techniques, such as those used on the fold edge tracking. The laser endoscope can be used with standard endoscopic light for physical estimation of glottal gaps and lesion size in the pediatric population. This article also provides preliminary evidence of differential kinematic characteristics of typically developing children compared with a child with vocal fold nodules and typically developing adults, requiring further large-scale studies.

CONCLUSION

The custom-built laser device resulted in quantitative measurements of vocal fold length and vibratory features with precision and accuracy sufficient for observing distinctions between related features of children and adults. The challenges of projecting patterns for measurement with minimal extensions of the endoscope tip were addressed with a sidemounted projector extending the tip 4 mm to the side and a vertical stripe pattern to maximize the sensitivity with changes in depth. Although low contrast limited stripe visibility, novel applications of averaging and band-pass filtering were used to reliably extract stripes to successfully extract measurements from children and adults. The device introduced in the article enabled the extraction of features that reflect actual distances and velocities for characterizing the kinematics of the vocal folds. Results from the limited population showed clear distinctions between adults and children as well as differences for the one child with nodules. In terms of features based on pixel ratios, larger amplitude-to-length ratios were observed for typically developing children relative to the adult population. In addition, features based on physical distance measurements showed interesting differences as well. Although children operate at a higher pitch, smaller closing velocities were observed (due to smaller peak displacements), except in the case of the child with nodules. In this case, velocities on the order of those in adults were observed, suggesting a higher impact stress and offering a potential explanation for the presence of the nodules. Although the purpose of the study was to develop a pediatric friendly laser endoscopic device, with further research similar device could be used for both adults as well as children.

Future clinical application

Considerable research is required for the clinical applicability of laser endoscopy coupled with high-speed imaging in a typical pediatric voice clinic. Future laser endoscopic devices need to consider the adequacy of the laser light intensity on vocal fold tissue and the visibility of the laser projection with high capture rates of up to 8000 frames per second under high illumination of 300-W Xenon light source that is used with high-speed imaging. Although the reported laser device is small, further miniaturization would make it feasible to potentially perform laser endoscopy on children younger than 5 years of age. Due to technological limitations, the current commercially available high-speed systems do not have an option of coupling a flexible endoscope, thereby overall limiting the application of high-speed imaging to only children who are able to tolerate rigid endoscopy. Future research on laser endoscopy coupled with high-speed laryngeal imaging has the potential to

greatly enhance clinical in-office practice in pediatric voice disorders by providing quantitative outcome measurements before and after therapy or surgical interventions, thereby providing evidence-based assessments of behavioral and surgical management of children with voice disorders that currently rely on perceptual or indirect assessment of vibratory motion.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The project was supported by National Institutes of Health (NIH)/National Institutes of Deafness and Other Communication Disorders R03DC11360-01. The project was also supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH, through grant UL1RR033173.

APPENDIX: Calibration process

The laser device projects a regular pattern of vertical stripes onto the field of view, which change their pixel position as the depth of the image plane changes relative to the endoscope. Let l_k be the column pixel coordinate of the k^{th} stripe edge, where $k = 1, 3, 5, \dots$ represents the left edges and $k = 2, 4, 6, \dots$ represents the right edges. A linear relationship exists between the depth and column (x -dimension) for each stripe edge denoted by:

$$l_k = \alpha(k) z_k + \beta(k), \quad (\text{A1})$$

where z_k is the depth at which the stripe l_k is incident on the image plane. The relationship in Equation A1 can be used to estimate depth from the stripe location once the parameters are determined for each stripe edge through a calibration procedure.

The depth z_k can then be used to estimate the orthogonal x and y coordinates of each pixel based on a simple pinhole camera model, in which the real-world coordinates are mapped into pixel coordinates by rotation, translation, and scaling. This transformation is represented by:

$$\underbrace{\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}}_p = \underbrace{\begin{bmatrix} r_{11}(z) & r_{12}(z) & t_x(z) \\ r_{21}(z) & r_{22}(z) & t_y(z) \\ 0 & 0 & 1 \end{bmatrix}}_{R_z} \underbrace{\begin{bmatrix} s_x(f, z) & 0 & 0 \\ 0 & s_y(f, z) & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{S_z(f)} \underbrace{\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}_w, \quad (\text{A2})$$

where $(x, y, \text{ and } z)$ are the real-world coordinates and $(x'$ and $y')$ are the image coordinates. R_z represents the depth-dependent rotation and translation, whereas $S_z(f)$ represents the scaling, which is dependent on depth (z) and focal length (f). w is the column matrix $[x, y, 1]$, where x , and y are the real world coordinates. For a constant focal length, the locus of real-world coordinates associated with a given image coordinate p forms a line in space as a function of depth, as shown in Figure 8. In a real image recording, however, the focal length will be adjusted to keep the image in focus. Therefore, consider an image point adjusted for the focal length so that the S_z is only dependent on depth. The reverse projection relationship can then be written as:

$$S_z^{-1}(f) R_z^{-1} p = \tilde{S}_z^{-1} \tilde{R}_z^{-1} \tilde{p} = w, \quad (\text{A3})$$

where $\tilde{p}$ denotes the scaled camera coordinates to account for focal length changes and $\tilde{S}_z$ denotes the new scaling matrix. The effect of varying focal length can be compensated by using the circular frame present in all the endoscopic images (Figure 9) as a reference. The

circular frame, which is at a constant depth from the camera, is resized to a fixed size (500×500) to revert any scaling due to a change in the focal length. Equation A3 can now be recast as a linear relationship between the real-world coordinates and depth for each camera coordinate:

$$\begin{bmatrix} x \\ y \end{bmatrix} = C \begin{bmatrix} x' \\ y' \end{bmatrix}_{2 \times 2} \begin{bmatrix} z \\ 1 \end{bmatrix}, \quad (\text{A4})$$

where $C[x', y']$ can be estimated from a minimum of two points along the projection line illustrated in Figure 8.

The unknown coefficients of $C[x', y']$ (Equation A4), R_z (Equation A3), (k) , and (k) Equation A1 are obtained through the calibration process. The stripe depth z_k and the real-world coordinates of the points where the stripe is incident are then computed from Equations A1 and A4. The relationship between the first seven stripe edge pixel locations and the corresponding depth is shown in Figure 10.

REFERENCES

1. Hirano M. Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatr* (Basel). 1974; 26:89–94. [PubMed: 4845615]
2. Patel RR, Dixon A, Richmond A, Donohue KD. Pediatric high speed digital imaging of vocal fold vibration: a normative pilot study of glottal closure and phase closure characteristics. *Int J Pediatr Otorhinolaryngol*. 2012; 76:954–959. [PubMed: 22445799]
3. Hirano, M.; Kurita, S.; Nakashima, T. Growth, development, and aging of human vocal fold. In: Bless, DM.; Abbs, JH., editors. *Vocal Physiology Contemporary Research and Clinical Issues*. College-Hill Press; San Diego, CA: 1983. p. 22–43.
4. Hartnick CJ, Rehbar R, Prasad V. Development and maturation of the pediatric human vocal fold lamina propria. *Laryngoscope*. 2005; 115:4–15. [PubMed: 15630357]
5. Boseley ME, Hartnick CJ. Development of the human true vocal fold: depth of cell layers and quantifying cell types within the lamina propria. *Ann Otol Rhinol Laryngol*. 2006; 115:784–788. [PubMed: 17076102]
6. Sato K, Hirano M, Nakashima T. Fine structure of the human newborn and infant vocal fold mucosae. *Ann Otol Rhinol Laryngol*. 2001; 110(5 pt 1):417–424. [PubMed: 11372924]
7. Larsson H, Hertegard S. Calibration of high-speed imaging by laser triangulation. *Logoped Phoniatr Vocol*. 2004; 29:154–161. [PubMed: 15764209]
8. Doellinger M, Kunduk M, Kaltenbacher M, et al. Analysis of vocal fold function from acoustic data simultaneously recorded with high-speed endoscopy. *J Voice*. 2012; 26:726–733. [PubMed: 22632795]
9. George NA, de Mul FF, Qiu Q, Rakhorst G, Schutte HK. Depth-kymography: high-speed calibrated 3D imaging of human vocal fold vibration dynamics. *Phys Med Biol*. 2008; 53:2667–2675. [PubMed: 18443389]
10. Hoppe U, Rosanowski F, Döllinger M, Lohscheller J, Schuster M, Eysholdt U. Glissando: laryngeal motorics and acoustics. *J Voice*. 2003; 17:370–376. [PubMed: 14513959]
11. Popolo PS, Titze IR. Qualification of a quantitative laryngeal imaging system using videostroboscopy and videokymography. *Ann Otol Rhinol Laryngol*. 2008; 117:404–412. [PubMed: 18646436]
12. Sapienza CM, Stathopoulos ET. Respiratory and laryngeal measures of children and women with bilateral vocal fold nodules. *J Speech Hear Res*. 1994; 37:1229–1243. [PubMed: 7877283]
13. Stathopoulos ET. Relationship between intraoral air pressure and vocal intensity in children and adults. *J Speech Lang Hear Res*. 1986; 29:71–74.
14. Tang J, Stathopoulos ET. Vocal efficiency as a function of vocal intensity: a study of children, women, and men. *J Acoust Soc Am*. 1995; 97:1885–1892. [PubMed: 7699170]

15. Stathopoulos ET, Sapienza CM. Developmental changes in laryngeal and respiratory function with variations in sound pressure level. *J Speech Lang Hear Res.* 1997; 40:595–614. [PubMed: 9210117]
16. Kahane JC. Growth of the human prepubertal and pubertal larynx. *J Speech Hear Res.* 1982; 25:446–455. [PubMed: 7176619]
17. Patel R, Dailey S, Bless D. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Ann Otol Rhinol Laryngol.* 2008; 117:413–424. [PubMed: 18646437]
18. Deliyski DD, Hillman RE. State of the art laryngeal imaging: research and clinical implications. *Curr Opin Otolaryngol Head Neck Surg.* 2010; 18:147–152. [PubMed: 20463479]
19. Wittenberg T, Moser M, Tigges M, Eysholdt U. Recording, processing, and analysis of digital high-speed sequences in glottography. *Mach Vis Appl.* 1995; 8:399–404.
20. Kobler JB, Rosen DI, Burns JA, et al. Comparison of a flexible laryngoscope with calibrated sizing function to intraoperative measurements. *Ann Otol Rhinol Laryngol.* 2006; 115:733–740. [PubMed: 17076094]
21. Kobler JB, Hillman RE, Zeitels SM, Kuo J. Assessment of vocal function using simultaneous aerodynamic and calibrated videostroboscopic measures. *Ann Otol Rhinol Laryngol.* 1998; 107:477–485. [PubMed: 9635457]
22. George NA, de Mul FF, Qiu Q, Rakhorst G, Schutte HK. New laryngoscope for quantitative high-speed imaging of human vocal folds vibration in the horizontal and vertical direction. *J Biomed Opt.* 2008; 13:064024. [PubMed: 19123670]
23. Wurzbacher T, Voigt I, Schwarz R, et al. Calibration of laryngeal endoscopic high-speed image sequences by an automated detection of parallel laser line projections. *Med Image Anal.* 2008; 12:300–317. [PubMed: 18373942]
24. Schubert S, Hoppe U, Döllinger M, Lohscheller J, Eysholdt U. High-precision measurement of the vocal fold length and vibratory amplitudes. *Laryngoscope.* 2002; 112:1043–1049. [PubMed: 12160271]
25. Shaw HS, Deliyski DD. Mucosal wave: a normophonic study across visualization techniques. *J Voice.* 2008; 22:23–33. [PubMed: 17014988]
26. Patel RR, Donohue KD, Johnson WC, Archer SM. Laser projection imaging for measurement of pediatric voice. *Laryngoscope.* 2011; 121:2411–2417. [PubMed: 21993904]
27. Dollinger M, Berry DA, Luegmair G, Huttner B, Bohr C. Effects of the epilarynx area on vocal fold dynamics and the primary voice signal. *J Voice.* 2012; 26:285–292. [PubMed: 21708451]
28. Hirano, M.; Kiyokawa, K.; Kurita, S. Laryngeal muscles and glottic shaping. In: Fujimura, O., editor. *Vocal Physiology: Voice Production, Mechanisms and Functions.* Raven Press Ltd.; New York, NY: 1988.
29. Titze, IR. *Principles of voice production.* Prentice Hall; Englewood Cliffs, NJ: 1994.
30. Unnikrishnan, H.; Donohue, KD.; Patel, R. Analysis of high-speed phonoscopy pediatric images. *Proceedings of the International Society for Optics and Photonics; SPIE;* 2012. p. 1Q-13Q.
31. Dejonckere PH, Kob M. Pathogenesis of vocal fold nodules: new insights from a modelling approach. *Folia Phoniatr Logop.* 2009; 61:171–179. [PubMed: 19571551]
32. Akif Kilic M, Okur E, Yildirim I, Guzelsoy S. The prevalence of vocal fold nodules in school age children. *Int J Pediatr Otorhinolaryngol.* 2004; 68:409–412. [PubMed: 15013605]
33. Maturo S, Hill C, Bunting G, Ballif C, Maurer R, Hartnick C. Establishment of a normative pediatric acoustic database. *Arch Otolaryngol Head Neck Surg.* 2012; 138:956–961. [PubMed: 22986573]

FIGURE 1.
Pediatric structured light laser projection endoscope.

FIGURE 2.
Attachment of laser assembly parallel to the endoscope.

FIGURE 3. Structured light endoscope calibration apparatus. (A) Laser stripes projected onto a grid. (B) Endoscope resting on a stable support with the calibration grid mounted on the stepper motor rail.

FIGURE 4.
Enhanced laser stripe contrast after band-pass filtering and averaging.

FIGURE 5.

Phantom model as viewed through the endoscope. (A) Stripe points close to the vocal fold edge are marked by the red dots. (B) The glottal length is marked using a red line. (A color figure can be found in the online version of this article.)

FIGURE 6. Instantaneous amplitude in millimeters in adult male, adult female, typical child, and child with vocal fold nodule.

FIGURE 7. Instantaneous velocity in meters per second in adult male, adult female, typical child, and child with vocal fold nodule.

FIGURE 8.

Ray diagram illustrating the relationship between the depth, image coordinates, and the real-world coordinates.

FIGURE 9.
The vertical stripes projected to a plane at 3-mm depth as viewed through the endoscope.

FIGURE 10.
Laser stripe edge location and corresponding depths for the first seven stripes.

TABLE 1

Vocal Fold Lengths and Vibratory Feature Estimates

Features	Children (n = 5), Mean (SD)	Adult Female (n = 3), Mean (SD)	Adult Male (n = 2), Mean (SD)	Child With Nodules (n = 1)
Phonatory length (mm)	6.35 (1.18)	11.38 (0.81)	14.95 (1.43)	5.16
Membranous length (mm)	4.66 (0.97)	8.2 (0.33)	11.36 (0.99)	4.08
Peak amplitude (mm)	0.31 (0.08)	0.64 (0.11)	1.00 (0.73)	0.53
Amplitude-to-length ratio	0.05 (0.02)	0.06 (0.01)	0.06 (0.02)	0.10
Avg. closing velocity (m/s)	0.34 (0.08)	0.45 (0.06)	0.52 (0.02)	0.48
Impact velocity (m/s)	0.50 (0.08)	0.63 (0.1)	0.56 (0.01)	0.68