33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Real-time Laryngoscopic Measurements of Vocal-fold Vibrations

Idaku Ishii, Shoichi Takemoto, Takeshi Takaki, Muneo Takamoto, Kentaro Imon, and Katsuhiro Hirakawa

***Abstract*— We have developed a high-frame-rate laryngoscope that can measure the vibration distribution of a human vocal fold in real time at hundreds of hertz. Our laryngoscope can extract a vocal-fold contour at 4000 fps as 20 pairs of its left and right border points from 256×512-pixel images to quantify left–right asymmetry of vocal-fold vibrations. Experiments on artificial vocal-fold-like vibrations of a silicon rubber membrane were performed to confirm the laryngoscope's effectiveness, and the vocal folds of human subjects, including patients with laryngeal diseases, were examined under clinical conditions.**

I. Introduction

To quantify degrees of laryngeal diseases such as laryngeal cancer and nodules, there is a strong demand in clinical otolaryngology for the examination of vocal-fold vibrations as well as the static appearance of vocal folds. A human vocal fold is self-excitedly vibrated at hundreds of hertz or more by breathing; this is too fast for human eyes and standard NTSC cameras (30 frames per second (fps)). Recently, many studies of the vibration dynamics of vocal folds have been reported. These studies involved the use of offline high-frame-rate (HFR) videos; laryngeal dynamics analysis by phonovibrography [1]; videokymography, which represents temporal brightness changes on horizontally scanned lines in an image [2]; laryngeal topography for pixel-level frequency analysis in brightness [3]; and modal analysis using laryngeal topography [4]. However, most of these offline HFR-video-based analyses need a long image-processing computation time to obtain several vocal-fold parameters. This non-simultaneousness limits the application range of HFR-video-based vocal-fold diagnosis under clinical conditions, where simultaneous online diagnosis is often required.

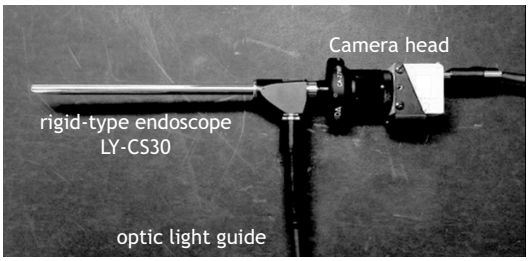
In this study, we developed an HFR-video-based laryngoscope that can simultaneously extract and visualize vocal-fold vibrations by introducing a high-speed vision platform that we have designed for real-time image processing of 256×512-pixel images at 4000 fps; we evaluated the laryngoscope under clinical conditions by quantifying the left–right asymmetry of the vocal-fold vibrations of human subjects, some of whom were patients with laryngeal diseases.

II. Real-time HFR-video-based Laryngoscope

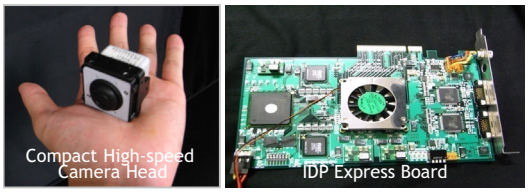
A. Configuration

Our developed HFR-video-based laryngoscope consists of an IDP Express high-speed vision platform and a rigid-type endoscope. Figure 1(a) shows an overview.

I. Ishii, S. Takemoto, and T. Takaki are with Graduate School of Engineering, Hiroshima University, Hiroshima 739-8527, Japan. M. Takamoto, K. Imon, and K. Hirakawa are with Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima 734-8551, Japan. (corresponding author (I. Ishii) Tel: +81-82-424-7692; e-mail: iishii@robotics.hiroshima-u.ac.jp).



(a) overview



(b) IDP Express

Fig. 1. Real-time HFR-video-based laryngoscope.

IDP Express [5] is a high-speed vision platform for processing and recording HFR videos. It consists of a camera head, a dedicated Field Programmable Gate Array (FPGA) processing board (IDP Express board), and a personal computer (PC). Figure 1(b) shows an overview. Eight-bit gray images are transferred from the camera head at 2000 fps (for 512×512-pixel images) and 4000 fps (for 512×256-pixel images). The IDP Express board is designed to implement image-processing algorithms on FPGAs. It can transfer input images and their processed results to a PC memory at thousands of frames per second. On the PC, various Application Programming Interface (API) functions that help in accessing memory-mapped data can be used on a Windows XP OS for application-software development. In this study, the algorithms (given below) used in measuring vocal-fold vibrations are implemented using an application-software program; we used a PC with an Intel Core 2 Quad CPU 2.50 GHz×4, an ASUSTeK P5Q(P45) ATX motherboard, 3.25-GB memory, and two 16-lane PCI-E 2.0 buses.

We used a rigid-type laryngoscope LY-CS30 manufactured by the Machida Endoscope Co., Ltd., Japan; its diameter, length, field of view, and observation depth are 8 mm, 155 mm, 40° in the direction angled at 70°, and 30–70 mm, respectively. The camera head of the IDP Express was mounted at the end of the LY-CS30 via a C-mount lens adaptor. The dimensions and weight of the camera head were 35 mm×35 mm×34 mm and 300 g, respectively. A metallic halide light-source ILH-2A (Olympus Co., Ltd., Japan) delivered bright-light illumination to the LY-CS30 via a fiber-optic light-guide.

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

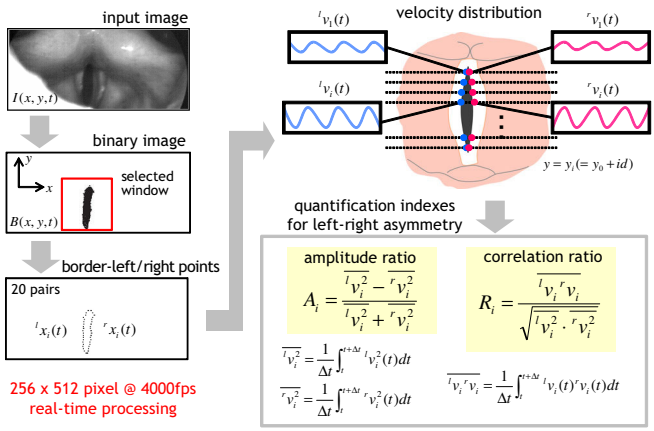


Fig. 2. Quantification algorithm for vocal-fold vibrations.

B. Implemented Algorithm

On the IDP Express, we implemented a contour-extraction algorithm by software to obtain a velocity distribution on the left and right edges of a vocal fold, and introduced quantification indexes for left–right asymmetry of vocal-fold vibrations to indicate the degree of laryngeal diseases.

1) Velocity distribution of vocal-fold vibrations

The velocity distribution on the left and right edges of a vibrating vocal fold can be detected as shown in Figure 2.

(a) Binarize input images

Binary images $B(x,y,t)$ are obtained as vocal-fold regions by binarizing input images $I(x,y,t)$ with a threshold θ.

$$B(x,y,t) = \begin{cases} 1 & (I(x,y,t) < \theta) \\ 0 & (\text{otherwise}) \end{cases}. \tag{1}$$

(b) Extract left and right edges of the vocal fold

N pairs of left and right edge positions of the vocal fold are extracted as ${}^{l}x_i(t)$ and ${}^{r}x_i(t)$, respectively, by scanning changes in $B(x,y,t)$ on $y = y_i (i = 1,\ldots,N)$ at intervals of d in a window to be tracked.

(c) Calculate velocities on the edges of the vocal fold

By differentiating ${}^{l}x_i(t)$ and ${}^{r}x_i(t)$ with respect to time, the velocity distribution of the vocal-fold vibration is calculated as follows:

$$({}^{l}v_i(t), {}^{r}v_i(t)) = \left(\frac{d}{dt}{}^{l}x_i(t), \frac{d}{dt}{}^{r}x_i(t)\right). \tag{2}$$

In this study, 20 pairs ($N = 20$) of left and right edge positions of a vocal fold were extracted, and the velocity distribution was calculated for 256×512-pixel images in real time at 4000 fps on our developed laryngoscope.

2) Quantification indexes

Based on the vibration velocities on the left and right edges of a vocal fold, an amplitude ratio A_i and correlation ratio R_i are calculated for $y = y_i (i = 1,\ldots,N)$ as quantification indexes of left–right asymmetry in vocal-fold vibrations.

$$A_i = \frac{\overline{{}^{l}v_i^2} - \overline{{}^{r}v_i^2}}{\overline{{}^{l}v_i^2} + \overline{{}^{r}v_i^2}}, \qquad R_i = \frac{\overline{{}^{l}v_i {}^{r}v_i}}{\sqrt{\overline{{}^{l}v_i^2} \cdot \overline{{}^{r}v_i^2}}}, \tag{3}$$

where $\overline{{}^{l}v_i^2}$, $\overline{{}^{r}v_i^2}$, and $\overline{{}^{l}v_i {}^{r}v_i}$ are given as time-correlation functions of left-edge and right-edge velocities in a time span $t \sim t + \Delta t$ as follows:

$$\overline{{}^{l}v_i^2} = \frac{1}{\Delta t}\int_t^{t+\Delta t} {}^{l}v_i^2(t)dt, \quad \overline{{}^{r}v_i^2} = \frac{1}{\Delta t}\int_t^{t+\Delta t} {}^{r}v_i^2(t)dt. \tag{4}$$

$$\overline{{}^{l}v_i {}^{r}v_i} = \frac{1}{\Delta t}\int_t^{t+\Delta t} {}^{l}v_i(t) {}^{r}v_i(t)dt. \tag{5}$$

A_i and R_i are sensitive to the amplitude differences and phase differences, respectively, between the left-side and right-side vibrations of a vocal fold, and they are normalized into the range of -1 and 1. When a vocal fold is vibrated left–right symmetrically, A_i and R_i become 0 and -1, respectively. When A_i and R_i differ from these standard values, these deviations indicate asymmetric vocal-fold vibrations, which are often caused by laryngeal diseases.

III. Evaluation of Artificial Vocal-Fold-Like Vibrations of a Silicon Rubber Membrane

To verify the performance of our laryngoscope for real-time vocal-fold-like vibration measurements at hundreds of hertz, a silicon rubber membrane with a slit incision was evaluated as an artificial vocal-fold model. The experimental setup is shown in Figure 3. The silicon rubber membrane had a slit incision 14 mm in length; this is the average length of an adult male's vocal cord. To simulate vocal-fold-like vibrations, the slit incision was vibrated using an air jet from an air compressor. In the evaluation, the vibration frequency was set to approximately 400 Hz, corresponding to that of a human vocal-fold. To simulate left–right asymmetry in vocal-fold-like vibrations, paper weights were attached on the left side of the slit incision.

Figure 4 shows the captured images and waveforms of the average positions of 20 pairs of left and right edge positions in 0.1 s for stationary vocal-fold-like vibrations with no weight and with a 4.0 g weight; the images and waveforms were extracted from 256×512-pixel images in real time at 4000 fps. Our laryngoscope can measure 400-Hz vibrations of both the left and right sides of the slit incision, and detect left–right asymmetry in the vocal-fold-like vibrations when there is a weight around the slit incision. Figure 5 shows the left-side and right-side envelopes of the average velocity distribution for an interval of $\Delta t = 0.1$ s for vocal-fold-vibrations with weights of 0.0 g, 1.0 g, 2.0 g, and 3.0 g, plotted as 20 pairs of averaged velocities, ${}^{l}V_i = (\overline{{}^{l}v_i^2(t)})^{\frac{1}{2}}$,

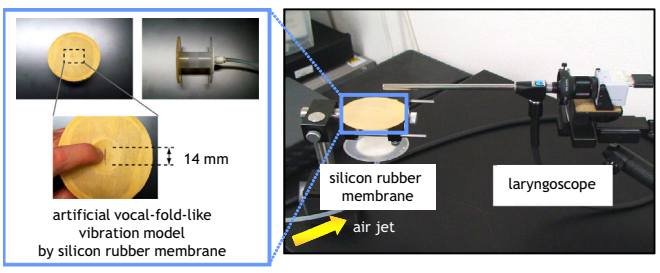


Fig. 3. Silicon rubber membrane with a vibrating slit incision.

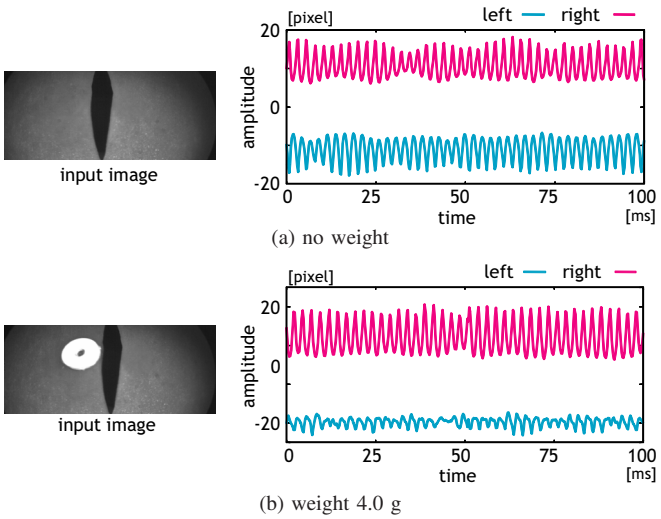


Fig. 4. Input images and left and right edge positions of a vibrating slit incision.

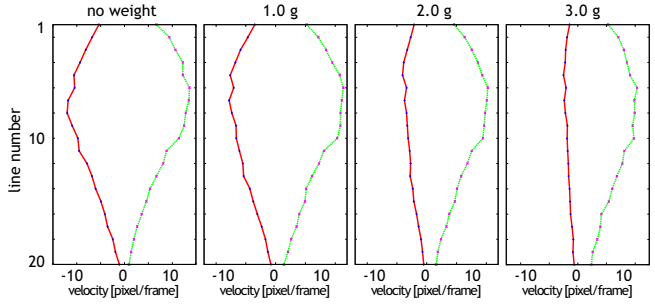


Fig. 5. Velocity distributions of artificial vocal-fold-like vibrations.

${}^{r}V_i = (\overline{{}^{r}v_i^2(t)})^{\frac{1}{2}}$ $(i = 1, \ldots, 20)$. When there was no weight attached, the velocity distribution was almost left–right symmetric. When the mass of the weight was increased, the amplitudes of the velocities on the left side gradually decreased in proportion to the mass of the weight.

Figures 6 and 7 show the amplitude ratios and correlation ratios, respectively, for quantifying left–right asymmetry in vocal-fold-like vibrations when we changed the mass of the weight on the left side of the slit incision from 0.0 g to 3.0 g. Subfigures (a) show the distributions of A_i and R_i, respectively. Subfigures (b) show the spatially averaged ratios $\bar{A} = \sum_{i=1}^{20} A_i/20$ and $\bar{R} = \sum_{i=1}^{20} R_i/20$, respectively. You can observe that the amplitude ratios were around 0 and the relative ratios were around -1 when there was no weight, and deviations from these standard values caused by left–right asymmetric vocal-fold-like vibrations increased in proportion to the mass of the weight on the left side of the slit incision. From these evaluation results, it was confirmed that our developed laryngoscope can simultaneously extract and quantify left–right asymmetric vocal-fold-like vibrations of a silicon rubber membrane at hundreds of hertz.

IV. Examination of Human Vocal-fold Vibrations

Using our laryngoscope, we examined the vocal-fold vibrations of five human subjects (A, B, C, D, and E) under

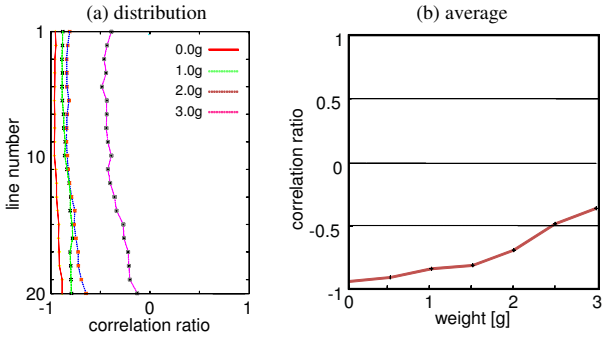


Fig. 6. Amplitude ratios of artificial vocal-fold-like vibrations.

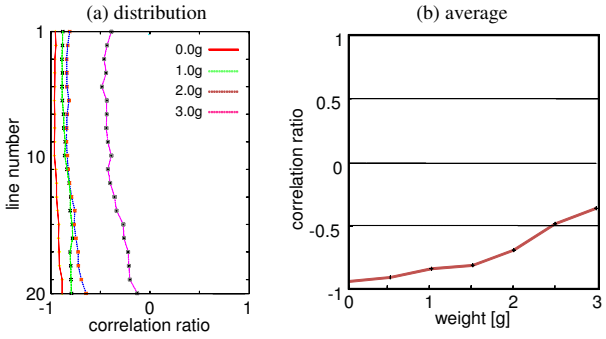


Fig. 7. Correlation ratios of artificial vocal-fold-like vibrations.

clinical conditions. Subjects A, B, and C were healthy males in their twenties. Subject D was a female patient in her fifties with a polypous vocal cord, and subject E was a male patient in his thirties who had had a vocal-cord nodule removed by an operation. In these real-time examinations under clinical conditions, the image size and frame rate of our developed laryngoscope were set to 512×512 pixels and 2000 fps, respectively, and 20 pairs of left and right border points of their vocal folds were stably extracted. The vibration frequencies of the vocal folds of subjects A, B, C, D, and E were approximately 180 Hz, 200 Hz, 140 Hz, 160 Hz, and 180 Hz, respectively.

For subjects A–E, Figure 8 shows the 512×512-pixel input images and time-averaged velocities ${}^{l}V_i$ and ${}^{r}V_i$ on the 20 scanned lines on the left and right edges of their vocal folds in 0.1 s. Figure 9 shows their amplitude ratios A_i and correlation ratios R_i on the 20 lines; these indicate the degree of asymmetry in the vocal-fold vibrations. For subjects A, B, and C, who were healthy, the amplitude ratios were around 0 and the correlation ratios were around -1. These tendencies indicate that their vocal folds vibrated almost left–right symmetrically. For subject D, it can be observed that the left–right symmetry in the vocal-fold vibrations was strongly disrupted because a polyp on the right side of the vocal fold disturbed vocal-fold vibrations. For subject E, the vocal-fold vibrations were observed to be almost left–right symmetric, but the correlation ratios were slightly less than those of the healthy subjects A, B, and C. This tendency corresponds to the fact that subject E felt an incompatibility in mucous-membrane movement around the site of the nodule removed

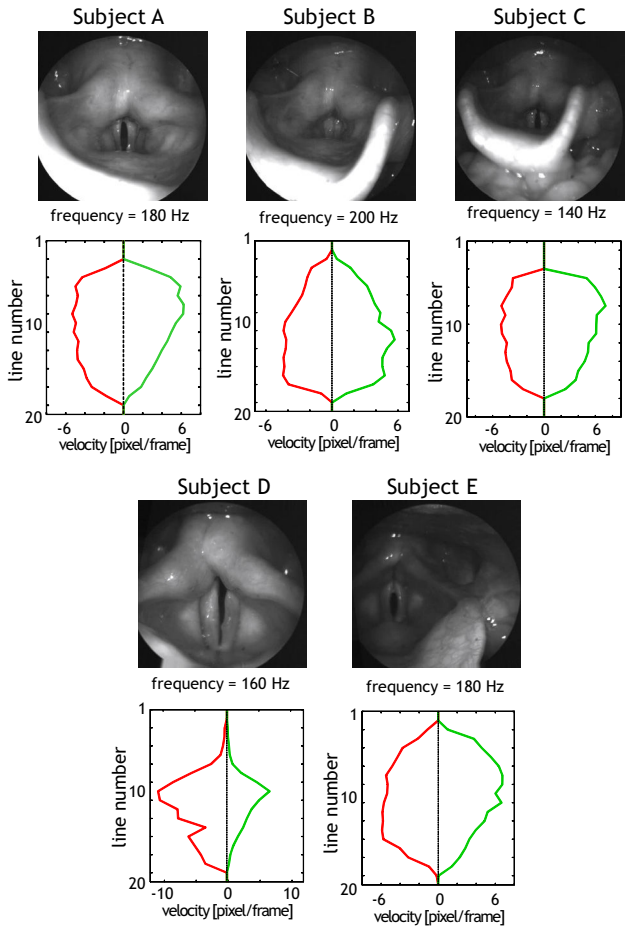


Fig. 8. Input images and velocity distributions of human vocal-folds.

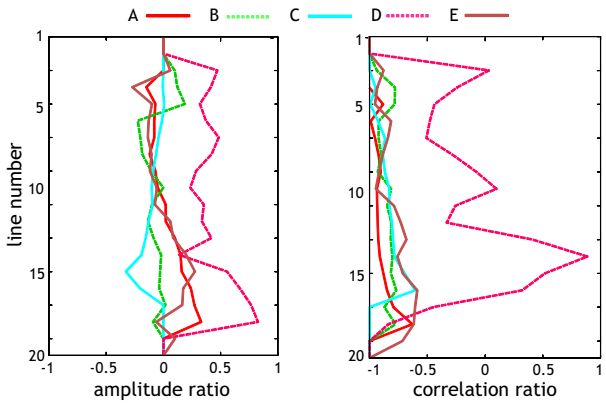


Fig. 9. Amplitude ratios A_i and correlation ratios R_i of human vocal-fold vibrations.

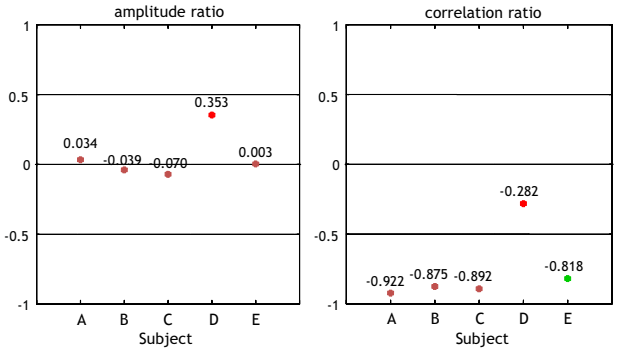


Fig. 10. Spatially averaged amplitude ratios $\bar{A}$ and correlation ratios $\bar{R}$ of human vocal-fold vibrations.

from the left vocal-fold.

Figure 10 shows the spatially averaged amplitude ratios $\bar{A}$ and correlation ratios $\bar{R}$ for subjects A–E. For subjects A, B, and C, $\bar{A}$ were 0.034, 0.039, and -0.070, respectively, and $\bar{R}$ were -0.922, -0.875, and -0.892, respectively. These ratios approximately matched the standard values ($\bar{A} = 0$, $\bar{R} = -1$) when vocal-fold vibration is left–right symmetric. For subject D, $\bar{A}$ and $\bar{R}$ were 0.353 and -0.282, respectively. These ratios deviated from the standard values, indicating that there was a significant left–right asymmetric vocal-fold vibration; this was caused by a polyp on the right side of the vocal fold. For subject E, the averaged amplitude ratio $\bar{A}$ was 0.003, i.e., close to zero, and the averaged correlation ratio $\bar{R}$ was -0.818, i.e., it deviated slightly from the values for the healthy subjects A, B, and C. It can be considered that the slight difference between the left-side and right-side vocal-fold vibrations results in incompatible mucous-membrane movement around the site of the removed nodule.

V. Conclusion

Real-time measurement of vocal-fold vibrations was achieved by developing an HFR-video-based laryngoscope that can quantify left–right asymmetry in vocal-fold vibrations using real-time image-processing at 4000 fps. Its performance was verified by quantifying vocal-fold vibrations for human subjects under clinical conditions; some of the subjects were patients with laryngeal diseases. We are now planning to quantify various types of vocal-fold vibrations for many more patients with laryngeal diseases, using the simultaneousness of our HFR-video-based laryngoscope to measure vocal-fold vibrations, and make a detailed database of vocal-fold vibrations that can provide more accurate and objective clinical diagnoses of laryngeal diseases.

References

[1] J. Lohscheller, U. Eysholdt, H. Toy, and M. Dollinger, "Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics," *IEEE Trans. Med. Imaging*, 27(3), 300–309, 2008.

[2] J.G. Svec and H.K. Schutte, "Videokymography: high-speed line scanning of vocal fold vibration," *J. Voice*, 10(2), 201–205, 1996.

[3] S. Granqvist and P. Lindestad, "A method of applying Fourier analysis to high-speed laryngoscopy," *J. Acoust. Soc. Am.*, 110(6), 3193–3197, 2001.

[4] K.-I. Sakakibara, H. Imagawa, Y. Yokonishi, and N. Tayama, "Laryngotopograph for high-speed digital images of normal and pathological vocal fold vibratory patterns," *28th World Congr. Int. Assoc. Logoped. Phoniat.*, 2010.

[5] I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, T. Takaki, and K. Tajima, "2000 fps real-time vision system with high-frame-rate video recording," *Proc. Int. Conf. Rob. Autom.*, 1536–1541, 2010.