Estimation of Glottal Area Function Using Stereo-endoscopic High-Speed Digital Imaging

Hiroshi Imagawa[1], *Ken-Ichi Sakakibara*[1,2], *Isao T. Tokuda*[3], *Mamiko Otsuka*[4], *Niro Tayama*[1,5]

[1] Department of Otolaryngology, University of Tokyo, Japan
[2] Department of Communication Disorders, Health Sciences University of Hokkaido, Japan
[3]Japan Advanced Institute of Science and Technology, Japan
[4]Kumada Clinic, Japan
[5]Department of Otolaryngology, Head and Neck Surgery,
National Center for Global Health and Medicine, Japan

imagawa@m.u-tokyo.ac.jp, kis@hoku-iryo-u.ac.jp, isao@jaist.ac.jp, ma-mi@wb3.so-net.ne.jp, ntayama@hosp.ncgm.go.jp

Abstract

In this paper, a novel stereo-endoscopic high-speed digital imaging system and a method to estimate the glottal area function are proposed. Glottal length, width, and area of one female participant were estimated in three different fundamental frequencies (F_0s).

Index Terms: glottal area function, stereoscopy, high-speed imaging

1. Introduction

Glottal area function estimation plays an important role in clarifying a physical mechanism of vocal fold vibration and investigating voice qualities in a quantitative manner. There have been various methods for estimating glottal area function, however, most of them estimate relative glottal area functions, and actual measurements of glottal area have been done only in vitro. In this paper, estimation of glottal area function in vivo based on actual measurement by stereoscopic high-speed digital imaging is proposed.

2. Stereo-endoscopic high-speed digital imaging and calculation of the three-dimensional coordinates

2.1. Stereo-endoscope

The stereo-endoscope in previous studies [3, 4, 5], manufactured by Nagashima Medical Instrument Corporation in 1980 (Figure 1), was employed in this study. The stereo-endoscope includes two independent ordinary rigid optical systems with a diameter of 9 mm, a fiber-optic light guide, an optical connector, a light source and a camera. The tips of the optical systems house objective lenses with prisms designed for 70° oblique-angled view, with a field angle of 40° (Figure 2). The distance between the optical axes of the tips was 10 mm. The stereo-endoscope was attached to a CCTV lens of 50 mm, and the CCTV lens was connected to the high-speed digital camera.

Figure 1: *Stereo-endoscope.*

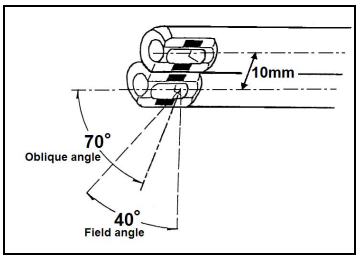

Figure 2: *Dimension of Stereo-endoscope.*

2.2. High-speed imaging and calculation of three-dimensional coordinates

The high-speed digital camera employed in this study was Photron Fastcam 1024PCI with the following specifications: an image sensor size of 17.4 mm×17.4 mm, a full image resolution of 1024×1024 pixels, a temporal resolution of 1000 fps at a full image resolution of 1024×1024, 8-bit grayscale, and a memory size of 12 GB allowing recording of 9600 frames at a full image resolution (a sample duration of 9.6s at the maximum speed). Reducing an image resolution of the image sensor in recording makes it possible to record at a higher frame rate.

In stereo-endoscopic high-speed digital recordings, the high-speed camera captured images at an image resolution of 768 (horizontal) × 352 (vertical), a frame rate of 3750 fps, and sample duration of 10.12s. Figure 3 shows an example of a pair of stereoscopic images of the larynx. A pair of images was formed side-by-side on the image sensor.

Measurements and a procedure of calculation are based on those reported in [1, 2, 3] (Figure 4). The two tips are assumed to be set coplanar, and are mutually inclined to a mid-axis by a small angle α. The distance between the optical axes at the tips is d_T. A rectangular coordinate system is defined with the origin at the tip of the left endoscope. The z-axis is along the optical axis of the left endoscope. The x-axis passes through

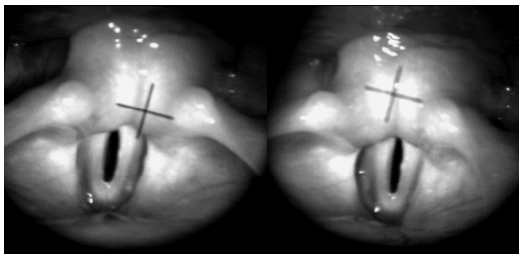

Figure 3: *A pair of stereoscopic images of the larynx.*

the two endoscope tips, and the *y*-axis is orthogonal to the *x*-axis and the *z*-axis (out of the page).
Vectors to the object point $p = (x_p, y_p, z_p)$ form angles θ_R and θ_L with the left and right optical axes respectively. Let D_L and D_R be horizontal distances of the images of p from the centers of the left and right optical fields respectively, and D_V be a vertical distance of the image of p from the center of the left and right optical fields, then coordinates of p are calculated by the following formulas:

$$z_p = \frac{1}{k_1(D_L - D_R) + k_2} \tag{1}$$

$$x_p = k_3 z_p D_L \tag{2}$$

$$y_p = k_3 z_p D_V \tag{3}$$

where k_1, k_2, k_3 are calibration constants empirically determined by photographing a Cartesian graph paper. The above calculations are true if D_L and D_R are proportional to $\tan\theta_L$ and $\tan\theta_R$, respectively. In reality, however, a photographic lens causes optical distortion and hence, the relationships such that D_L and D_R are proportional to $\tan\theta_L$ and $\tan\theta_R$ are less likely to be satisfied. Therefore, further calibration and correction of optical distortion are desired. After including correction of optical distortion, the modified formulas to calculate the coordinates are as follows:

$$z_p = \frac{1}{k_1(D_L - c_1 D_R + c_2 D_V) - k_2} \tag{4}$$

$$x_p = k_3 f(z_p, D_L) + k_4$$

$$\text{where} \quad f(z_p, D_L) = \frac{D_L - c_3 {z_p}^2 + c_4 z_p + c_5}{c_6 {z_p}^{-c_7}} \tag{5}$$

$$y_p = k_5 z_p D_V + k_6 \tag{6}$$

where k_i and c_i are calibration constants. The procedure to determine constants k_i and c_i was as follows: (i) D_L and D_R were measured by changing distance between the tips of endoscope and the 5 mm Cartesian graph paper from 14 mm to 84 mm; (ii) the regression lines of D_L on x_p and D_R on x_p were calculated for each z_p; (iii) D_L and D_R were represented as functions both having parameters of x_p and z_p; (iv) the regression plane of (D_L, D_R, D_V) was obtained.
As a result, distribution of errors between real coordinates and estimated coordinates in the three-dimensional Euclid space had a median of 0.55 mm (5 percentile of 0.15 mm, 95 percentile of 2.96 mm). The errors of x_p, and y_p were less than 15% of the error of z_p.

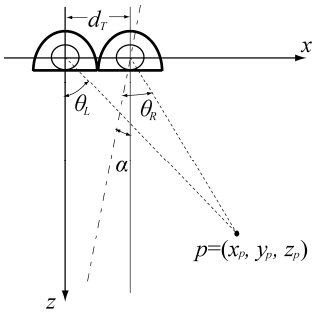

Figure 4: *Geometrical quantities defined for calculation of three dimensional coordinates*

3. Estimation of Glottal Area Function

3.1. Method

3.1.1. Glottal edge detection

First, glottal edges in the left and right images both are detected to estimate a glottal area in each frame. On each horizontal line, the edges are automatically determined as the points with maximal brightness derivative among the points with minimum brightness.
To represent the glottis as a plane in the three-dimensional space, the following steps were processed: (i) smoothing the estimated left and right edges independently along *y*-axis by a predetermined window function, reasonably assuming that the edge of glottis is a smooth curve in three-dimensional space, here, the 7-point weighted mean with a length of 7 pixels (0.7 mm at the distance of 50 mm from the endoscope tips in the real space) was employed for smoothing; (ii) determining a regression line of z on y for each edge after the smoothing, and rewriting z in such a manner that the left and right glottal edges were represented as two lines; (iii) for each y, picking up middle m_y point of the left and right glottal edges, then a linear approximation C of a curve $\{ m_y \}$ was obtained by linear regression of z on y; (iv) calculating lateral inclination from the left edge (x_L, y, z_L) to the right edge (x_R, y, z_R) for each y, and the mean of the lateral inclinations, denoted by $(dz/dx)_{mean}$; (v) obtaining the glottal hyperplane as the plane including the line approximation C and the hyperplane's cotangent vector is orthogonal to $(dz/dx)_{mean}$. The glottal edge points were obtained as points of projection along z-axis of (x_L, y, z_L) and (x_R, y, z_R) on the glottal hyperplane.

3.1.2. Verification of the method

To verify the proposed method for estimation of the glottal edges, the proposed method was applied to estimate a rectangular slit obtained by cutting through a thick paper (Figure 5). The rectangular slit had the length of 13.2 mm, the width of 2.2 mm, and the depth of 0.25 mm. Hence, the area of the slit was 29.04 mm^2.

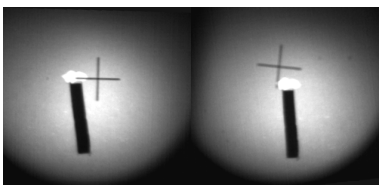

Figure 5: *A pair of stereoscopic images of the larynx.*

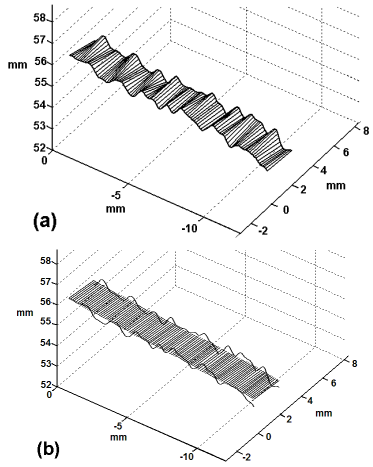

Figure 6: *(a)*: *glottis of the slit after edge smoothing and (b): glottis of the slit after smoothing and planar approximation.*

Using the proposed method, the area of the glottis was 35.6 mm^2 after smoothing, and 29.5 mm^2 after smoothing and planar approximation. Figure 6 illustrates the glottis after smoothing in (a) and the estimated glottis after smoothing and planar approximation in (b).

3.2. Experiments

One female participant without any vocal problems performed three different tasks: in different F_0s (middle, high, and low), with the same sustained vowel (almost /e/ by reason of insertion of endoscope into the mouth) and observed by stereo-endoscopic high-speed digital imaging. Figure 7 shows static laryngeal views during phonations in three different registers. The measured F_0 of voices in middle, high, and low F_0 were 230 Hz, 450 Hz, and 100 Hz, respectively. The low F_0 voice was perceived as vocal fry and the other two voices were perceived as modal.

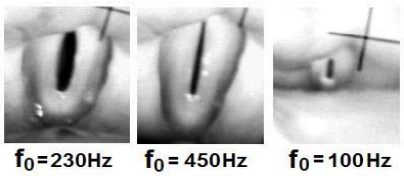

Figure 7: *The laryngeal views during middle F_0 (Left), high F_0 (middle), and low F_0 (right).*

3.3. Results

Figure 8 shows the glottis after smoothing in (a) and the estimated glottis after smoothing and planar approximation in (b) at F_0 = 230 Hz. In this case, the mean lateral inclination $(dz/dx)_{mean}$ was -0.3. The mean lateral inclinations in the cases of high and low F_0s were also equal to -0.3.

Figure 9 shows a time-varying function of the glottal width and length (solid and dotted lines, respectively, in the top graph), and the glottal area (in the bottom graph) at F_0 = 230 Hz. The maximum glottal area was 11 mm^2, and the maximum glottal width was 2 mm. The glottal length function was not significantly triangulated, and the maximum length was 7 mm. By observing the glottal area and width, the closing phase was slightly shorter than the opening phase in each period.

Figure 10 shows time-varying glottal characteristics at F_0 = 450 Hz. The maximum glottal area was about 5 mm^2, and the maximum glottal length was approximately 8 mm. The glottis at F_0 = 450 Hz was 1 mm longer than that at F_0 = 230 Hz. The maximum glottal width was 1.0 mm and significantly narrower than that at F_0 = 230.

Figure 11 shows time-varying glottal characteristics at F_0 = 100 Hz in vocal fry. The maximum glottal area was 1.3 mm^2, and the maximum glottal length was approximately 2 mm.

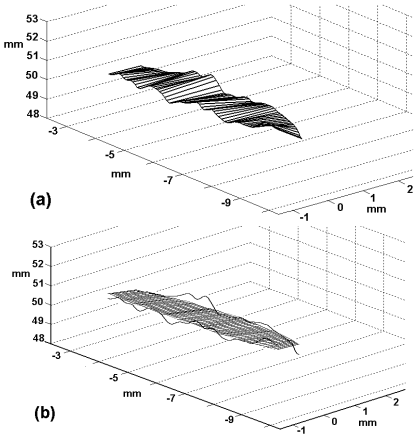

Figure 8: *(a)*: *glottis after smoothing and (b): glottis after smoothing and planar approximation in F_0 = 230 Hz.*

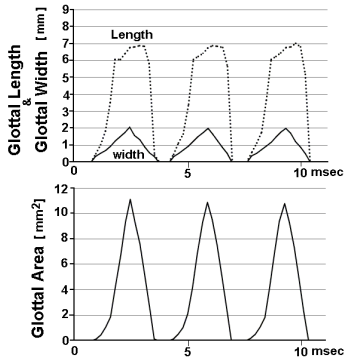

Figure 9: *Glottal length (solid line at the above), glottal length (dotted line at the above), and glottal area (at the bottom) in F_0 = 230 Hz*

4. Discussion

The glottal area functions estimated by the proposed method with stereo-endoscopic high-speed digital imaging were in accordance with known results. The estimated values of the maximum glottal lengths at F_0 = 230 Hz and at 450 Hz showed good accordance with those in [4, 5]. In the future, it is necessary to improve the method for estimating the glottal area function from the theoretical and instrumental viewpoints. For example, the glottis has been commonly defined as a space between two vocal folds, however, the space between the vocal folds in reality is three-dimensional and has a volume. As a result the glottal area is defined as a certain section of the glottal space and its size may vary depending on a plane which gives the glottal section.

The time-varying function of the mean lateral inclination is illustrated at the top picture in Figure 12. This figure shows that the maximum lateral inclination is at the open phase. This

phenomena implies that the edge of the lower lip of the vocal fold is detected as the glottal edge in one side, and the edge of the upper lip is detected as the glottal edge in the other side, by inclination of the optical axis at the open phase. However, it is reasonable to assume that the lateral inclination of the glottis is insignificant.

The bottom picture of Figure 12 shows a comparison between the glottal area function with time-varying lateral inclination (dotted line) and with constant lateral inclination which is obtained by averaging inclinations in each period (solid line). The maximum area with the constant lateral inclination of the glottis is 20% smaller than the maximum area with time-varying inclination. Ideally, the endoscope must be set to minimize inclination of the optical axis. However, it is difficult to set the position and direction of the endoscope tip in the small pharyngeal space. Thus, the proposed method using a constant value for the lateral inclination of the glottis is likely to provide a solution to obtain more accurate estimation against this difficulty in the endoscope setting.

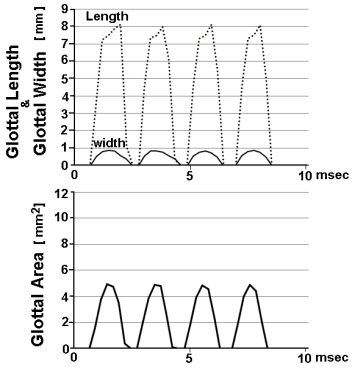

Figure 10: *Glottal length (solid line at the above), glottal length (dotted line at the above), and glottal area (at the bottom) in F_0 = 450 Hz*

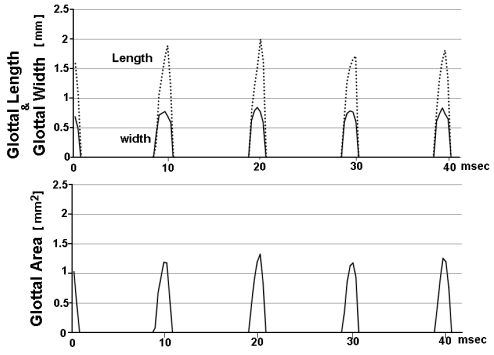

Figure 11: *Glottal length (solid line at the above), glottal length (dotted line at the above), and glottal area (at the bottom) in F_0 = 100 Hz*

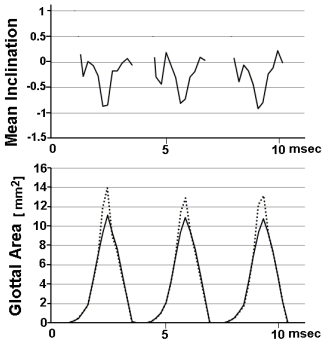

Figure 12: *mean lateral inclination (at the above), glottal area with planar approximation with time-varying mean lateral inclinations (dotted line at the above), and with time-averaged lateral inclination (solid line). F_0 = 230 Hz*

5. Conclusions

The glottal area function was estimated using the stereo-endoscopic high-speed digital imaging system. The estimated values and functions seem to be reasonably in good accordance with known results. However, there still exist many difficulties in accurately estimating glottal area to overcome, theoretically and practically. Further improvement of the estimation method and three-dimensional reconstruction of laryngeal views are addressed in the future studies.

6. Acknowledgements

The authors would like to thank Hisayuki Yokonishi for their helpful discussions, and Tatsuya Yamasoba and Takaharu Nito for their supports on our research. The authors also thank Mika Ito for her valuable comments on this paper. This research was partly supported by Japan and Grant-in-Aid (KAKENHI: 20500161) from the MEXT, Japan, SCOPE (071705501) of MIC, Japan, and JAIST Grant for Exploratory Research.

7. References

[1] M. Sawashima and S. Miyazaki, "Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberscopes", *Ann. Bull. RILP*, 8:7—10, 1974. http://www.umin.ac.jp/memorial/rilp-tokyo/

[2] O. Fujimura, T. Baer, and S. Niimi, "A stereo-fiberscope with a magnetic interlens bridge for laryngeal observation", *J. Acoust. Soc. Am.*, 65(2):478—480, 1979.

[3] K. Honda, S.R. Hibi, S. Kiritani, S. Niimi, and H. Hirose, "Stereoendoscopic measurement of the laryngeal structure", *Ann. Bull. RILP*, 14:73—78, 1980.

[4] M. Sawashima, H. Hirose, S. Hibi, H. Yoshioka, N. Kawase, and M. Yamada, "Measurements of the vocal fold length by use of stereoendoscope – a preliminary study", *Ann. Bull. RILP*, 15:9—16, 1981.

[5] M. Sawashima, H. Hirose, K. Honda, H.Yoshioka, S. R. Hibi, N. Kawase, and M. Yamada, "Stereoendoscopic Measurement of the Laryngeal Structure", *Vocal Fold Physiology, Contemporary research & Clinical issues*, Edited by Diane M. Bless and James H. Abbs, Collegge-Hill Press, 264-276, 1983.