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

Hiroshi Imagawa¹, Ken-Ichi Sakakibara¹,², Isao T. Tokuda³, Mamiko Otsuka⁴, Niro Tayama¹,⁵

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
⁴Kumada Clinic, Japan
⁵Department 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 (F₀s).

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.

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 dₓ. 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...
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 \( \theta_L \) and \( \theta_R \) 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 \( D_D \) 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:

\[
\begin{align*}
    z_p &= \frac{1}{k_1(D_L - D_R) + k_2} \quad (1) \\
    x_p &= k_3z_pD_L \quad (2) \\
    y_p &= k_3z_pD_D \quad (3)
\end{align*}
\]

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:

\[
\begin{align*}
    z_p &= \frac{1}{k_1(D_L - c_1D_R + c_2D_D) - k_2} \quad (4) \\
    x_p &= k_3f(z_p, D_L) + k_4 \\
    y_p &= k_5z_pD_D + k_6
\end{align*}
\]

where \( c_1 \) and \( c_2 \) 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 \( (D_L, D_R, D_D) \) 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 \( z_p \).
Using the proposed method, the area of the glottis was 35.6 mm² after smoothing, and 29.5 mm² 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 $F_0$s (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).](image)

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 $\langle d z/d x \rangle_{\text{mean}}$ was -0.3. The mean lateral inclinations in the cases of high and low $F_0$s 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 mm², 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 8: (a): glottis after smoothing and (b): glottis after smoothing and planar approximation in $F_0 = 230$ Hz.](image)

![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.](image)

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.

The glottal area function was estimated using the stereoendoscopic 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.

5. Conclusions

The glottal area function was estimated using the stereoendoscopic 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


