2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Reconstruction of Three-dimensinal Laryngeal Dynamics via Stereo-matching Technique

Isao Tokuda[1], Masahiro Iwawaki[2], Ken-Ichi Sakakibara[3,4], Hiroshi Imagawa[4], Takaharu Nito[4], Tatsuya Yamasoba[4], Niro Tayama[5]

1: Department of Micro System Technology, Ritsumeikan University, Shiga, 525-8577 Japan
2: School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, 923-1292 Japan
3: Department of Communication Disorders, Health Sciences University of Hokkaido, Hokkaido 002-8072 Japan
4: Department of Otolaryngology, University of Tokyo, Tokyo 113-8655 Japan
5: National Center for Global Health and Medicine, Tokyo 162-8655 Japan

Abstract—Towards quantitative measurement of three-dimensional dynamics of human vocal folds, stereo matching technique is applied to stereo high-speed filming data. Opening phase as well as closing phase was successfully reconstructed for one male subject in good agreement with the laryngeal mechanism known from the standard knowledge.

1. Introduction

Observation of the vocal fold vibration is indispensable for understanding the voice production in particular in singing voice and voice pathology. Various methods have been developed for observing the laryngeal mechanism such as electroglottography (EGG) [1], high-speed imaging [2], videokymography [3], and excised larynx experiment [4]. These techniques are however based on partial measurement of the vocal folds, and actual measurements of the three-dimensional dynamics of the larynx have been done only *in vitro*. In this paper, stereo high-speed filming technique [5, 6, 7] is combined with stereo matching [8] to recover three-dimensional dynamics of the vocal folds *in vivo*.

2. Stereo High-speed Recording

The stereo-endoscope in previous studies [5, 6, 7] 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.

As a high-speed digital camera, Photron Fastcam 1024PCI was employed. The stereo-endoscopic high-speed digital recording was carried out with an image resolution of 768 (horizontal) × 352 (vertical), a frame rate of 3750 fps, and a recording duration of 10.12s.

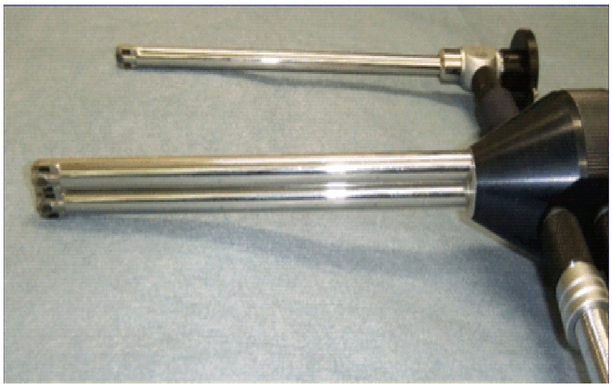

Figure 1: Picture of stereo-endoscope.

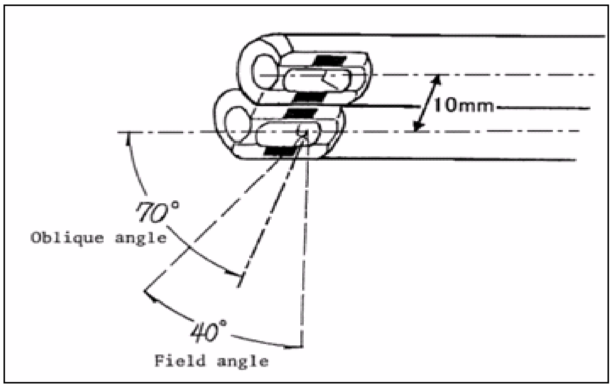

Figure 2: Dimension of stereo-endoscope.

3. Stereo Matching

Stereo-matching is one of the standard techniques in the field of image analysis [8]. The idea of the stereo matching is to find the corresponding points on the left and right images to measure its depth. Among conventional techniques, area-based stereo matching technique combined

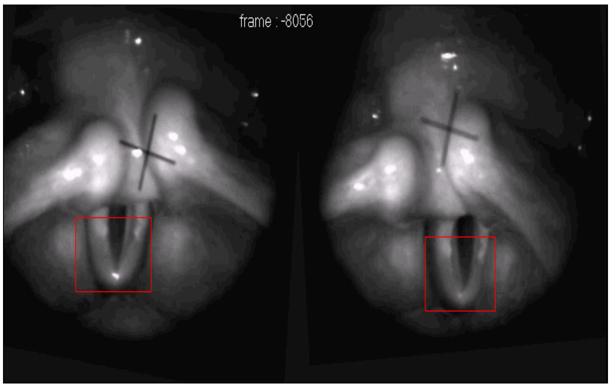

Figure 3: A pair of stereoscopic images of the larynx.

with a prior selection of landmarks was employed. Once the pairs of matching points are selected, three-dimensional shape can be recovered. Let D_L and D_R be horizontal distances of the matching points from the centers of the left and right optical fields, respectively, and D_V be a vertical distance of both points from the center of the left and right optical fields. Then, location of the matched points in three-dimensional space $p = (x_p, y_p, z_p)$ is given by the following formulas:

$$
\begin{aligned}
z_p &= \frac{1}{k_1(D_L - D_R) + k_2}, \\
x_p &= k_3 z_p D_L, \\
y_p &= k_3 z_p D_V,
\end{aligned}
$$

where k_1, k_2, k_3 are calibration constants empirically determined by photographing a Cartesian graph paper. These formulas are valid for idealistic situation. In reality, however, a photographic lens causes optical distortion, which violates the above formulas. To correct such distortion, modified formulas were utilized in our computation.

4. Results

4.1. Recording

One male subject with no laryngeal pathology was asked to produce a modal voice with a sustained vowel and observed by stereo-endoscopic high-speed digital imaging. Figure 3 shows static laryngeal views during the phonation.

4.2. Three-dimensional reconstruction

With respect to the pair of stereoscopic images of figure 3, stereo-matching technique was applied. Figure 4 shows the result of three-dimensional reconstruction of the vocal folds. Smooth surfaces of the left and right vocal folds can be clearly recognized. Glottal opening area is also seen in the middle as a deep blue area.

Application of this technique to a series of stereo images produced a movie of three-dimensional laryngeal dynamics. Figure 5 shows snap shots of the vocal folds movement viewed in a frontal section. From $t = 0$ to $t = 2.4$, the vocal folds move close to each other and they collide with each other at $t = 4.8$. After the complete closure of the glottis ($t = 5.87, 7.20$), the vocal folds start to open at $t = 7.47$. It should be noted that in the closing phase the lower edge of the vocal fold precedes the upper edge. This represents a phase shift between the lower and upper edges of the vocal folds, which is well known for efficient energy transfer from the airflow to the vibrating vocal folds [9]. Because the lower edge is hidden by the upper dege during the opening phase, the phase shift is discernible mainly during the closing phase.

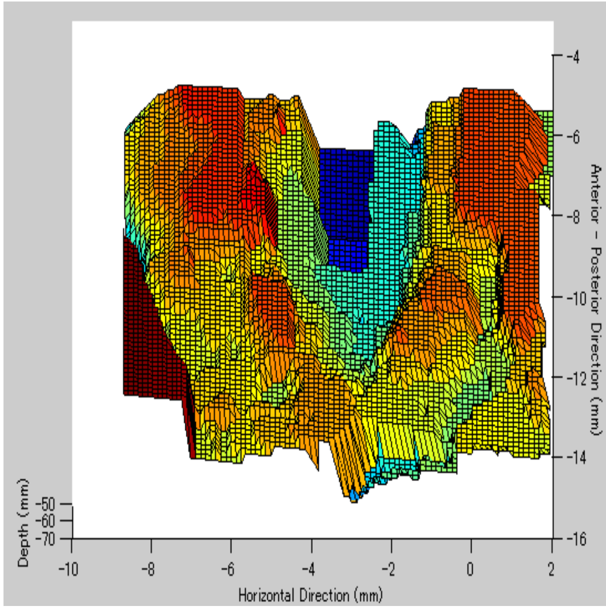

Figure 4: Three-dimensional configuration of the larynx reconstructed from the stereoscopic images.

5. Conclusions and discussions

Stereo matching technique has been applied to stereo high-speed filming data to recover three-dimensional dynamics of the vocal folds. The results imply that the reconstructed dynamics agrees quite well with the laryngeal mechanism known from the standard knowledge. This provides a promising step towards quantitative measurement of three-dimensional laryngeal motion. Our study is preliminary in the sense that it was applied only to a single phonation data with one subject. Various data including several subjects with different registers should be examined in our future work.

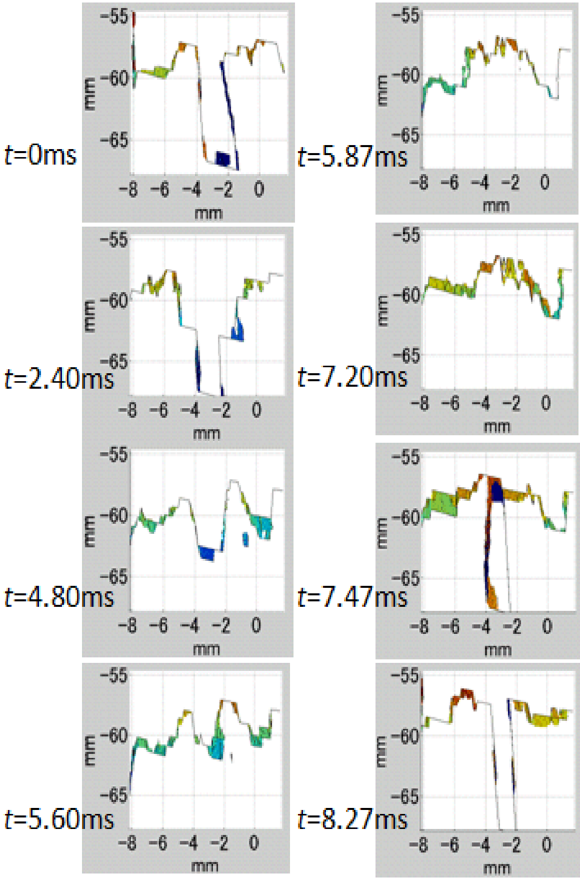

Figure 5: Movement of the vocal folds in frontal section.

Acknowledgments

This research was partly supported by Grant-in-Aid (KAKENHI: 20500161) from the MEXT of Japan, SCOPE (071705501) of MIC of Japan, and JAIST Grant for Exploratory Research.

References

[1] T. Baer, A. Löfqvist, and N.S. McGarr, "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques", *J. Acoust. Soc. Am.*, vol.73, p.1304, 1983.

[2] S. Kiritani, H. Hirose, and H. Imagawa, "High-speed digital imaging analysis of vocal fold vibration in diplophonia," *Speech. Comm.*, vol.13, p.23, 1993.

[3] J.G. Švec and H.K Schutte, "Videokymography: High-Speed Line Scanning of Vocal Fold Vibration," *J. Voice*, vol.10, pp.201-205, 1996.

[4] J. van den Berg and T.S. Tan, "Results of experiments with human larynxes," *Pract. Otorhinolaryngol.* (Basel), vol.21, pp.425-450, 1959.

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

[6] 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*, vol.15, pp.9-16, 1981.

[7] 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, pp.264-276, 1983.

[8] R. Nevatia, *Machine Perception* (Prentice-Hall, Englewood Cliffs, N.J., 1982).

[9] I.R. Titze, "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.*, vol.83, pp.1536-1552, 1988.