A method for assessing the regional vibratory pattern of vocal folds by analysing the video recording of stroboscopy

J. S. Lee[1] E. Kim[2] M.-W. Sung[3] K. H. Kim[3] M. Y. Sung[2] K. S. Park[1]

[1]Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul, Korea
[2]Department of Computer Science, Inchon University, Inchon, Korea
[3]Department of Otolaryngology Head and Neck Surgery, Seoul National University College of Medicine, Seoul, Korea

Abstract—*Stroboscopy and kymography have been used to examine the motional abnormality of vocal folds and to visualise their regional vibratory pattern. In a previous study (Laryngoscope, 1999), we introduced the conceptual idea of videostrobokymography, in which we applied the concept of kymography on the prerecorded video images using stroboscopy, and showed its possible clinical application to various disorders in vocal folds. However, a more detailed description about the software and the mathematical formulation used in this system is needed for the reproduction of similar systems. The composition of hardwares, user-interface and detail procedures including mathematical equations in videostrobokymography software is presented in this study. As an initial clinical trial, videostrobokymography was applied to the preoperative and postoperative videostroboscopic images of 15 patients with Reinke's edema. On preoperative examination, videostrobokymograms showed irregular pattern of mucosal wave and, in some patients, a relatively constant glottic gap during phonation. After the operation, the voice quality of all patients was improved in acoustic and aerodynamic assessments, and videostrobokymography showed clearly improved mucosal waves (change in open quotient: mean $\pm$ SD $=$ 0.11 $\pm$ 0.05).*

Keywords—*Videostrobokymography, Stroboscopy, Vocal fold, Larynx*

Med. Biol. Eng. Comput., 2001, **39**, 000–000

1 Introduction

IN PHONATION, the vocal folds play a pivotal role in controlling the airflow through the larynx. To create the vibration of airflow, they form an autonomous oscillator by interacting with acoustic pressures in the respiratory tract. Because the larynx and vocal folds are situated deep in the neck, they are relatively inaccessible. Thus, special techniques are required to examine their structure and to evaluate their function. Despite the fact that accurate images of vibrating vocal folds can be obtained using ultra-high-speed cinematography, it is expensive and time-consuming (FUJIMURA, 1988; GAUFFIN and HAMMARBERG, 1991).

Stroboscopy, which overcomes the drawbacks of ultra-high-speed cinematography, is useful in examining the vibratory pattern of vocal folds and is available for clinical procedures (TRAPP and BERKE, 1988; KARNELL, 1989; BLITZER *et al.*, 1992; HIRANO and BLESS, 1993). During speech, vocal folds vibrate 100 times per second or more, which is too fast to be seen by the naked eye. In order to overcome this difficulty, a special light source called a strobe light is used to illuminate vocal folds in slow motion. The strobe emits a very bright and very short flash of light. If the firing rate is adjusted so that it is a little faster or slower than the vibration rate of the folds, the folds will appear to move in slow motion. This motion can be videotaped and replayed for further analysis. This type of examination is important because it allows mucosa, a constituent of the vocal folds, to be examined during voice production. Early changes or asymmetry in mucosal vibration can signify a vocal fold weakness, a mucosal problem, tremor, spasm, or an infiltrative process inhibiting the motion (BLITZER *et al.*, 1992).

In spite of these advantages, interpretation by an experienced physician is necessary for accurate diagnosis due to the very subjective nature of the interpretation. Other problems are that since stroboscopy is recorded on videotape, it is time-consuming to find data to follow up the progress of a patient and a large amount of storage capacity is needed.

To solve these problems, many researchers have tried to extract quantitative and objective measures from stroboscopic records. WENDLER *et al.* (1986) measured features such as maximal width, minimal width, open quotient, and speed index or stroboscopic images. WOO (1996) plotted glottal area waveform and extracted features such as peak glottal area, duration of opening and closing, and rates of vocal fold opening and closing from the waveform. They analysed pre-recorded video signals using a conventional videostroboscopy

Correspondence should be addressed to Dr Kwang Suk Park; email: kspark@snuvh.snu.ac.kr

First received 6 November 2000 and in final form 19 February 2001

MBEC online number: 20013578

system. Since the parameters were determined by analysing the global movement of vocal folds, they could not reflect regional dysfunction, such as local left-right asymmetry in vocal fold vibration.

GALL and HANSON (1973) and GROSS (1985) introduced the technique of kymography, which is not limited to periodic vibration, and ŠVEC and SHUTTE (1995) further developed the concept of kymography and provided a means to visualize the regional vibratory pattern of vocal folds by displaying successive line images selected by the operator in such a way that the time dimension of the selected area is displayed in a vertical direction. Although the method of ŠVEC and SHUTTE (videokymography) can provide information on the actual movement of one line of interest, a modified CCD camera to acquire the line images of vocal folds with a rate of almost 8000 line images sec^{-1} is needed. This system, moreover, cannot select and display multiple lines during the monitoring of vocal fold vibration.

We have therefore developed a low-cost system, videostrobokymography (SUNG *et al.*, 1999), in which Švec's method is applied to pre-recorded video images using conventional stroboscopy. By analysing the pre-recorded image, we can examine multiple regions simultaneously in a more compatible environment after routine clinical procedures. In a previous report (SUNG *et al.*, 1999), we introduced the concept of the videostrobokymography, the fundamental principle of which is similar to that of the laryngostroboggraphy developed by ISOGAI (1994, 1996), and showed its possible clinical application to various disorders in vocal folds. However, a more detailed description of the software and the mathematical formulation used in this system is required for the reproduction of similar systems, and these technical issues are addressed in this study.

2 System components and user interface

The hardware of the videostrobokymography system consists of a VCR, an image-grabbing board*, and a processing computer (Pentium processor, 266 MHz CPU and 128 MB memory). The image-grabbing board was used to digitise analogue video signals for further analysis using a computer. Software operating in Windows 95† was developed using Visual C++‡ for image-grabbing and analysis. A C++ library provided with the image-grabbing board was used to control the board.

One hundred successive frames of videostroboscopic signal, corresponding to about four seconds, were captured and loaded into computer memory via the image-grabbing board. They were saved as an AVI-type movie. Four seconds was sufficient to permit capture of about four to six cycles of vibration of vocal folds. This number of cycles was considered to be sufficient to observe the pattern of vibration.

A videostrobokymogram using the digitised video images was composed as follows. First, one static frame was selected from the dynamic movie image, and a rectangular region of interest (ROI) was selected to include the region containing both vocal folds. A new window, then, appeared on the upper left corner of the screen and showed the first frame of the dynamic image including only the selected region. The ROI was chosen to minimise computation time and memory requirement for image data storage. The user clicked the mouse on the horizontal line of interest within the small window. The same lines in all frames

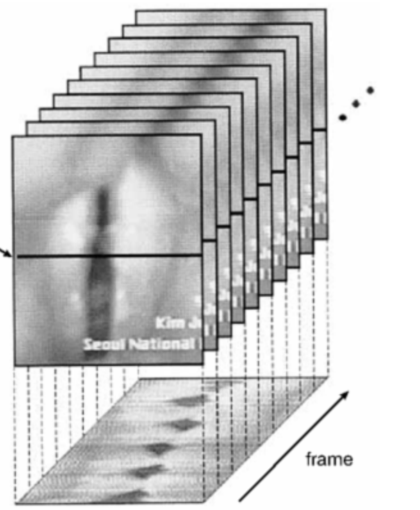

Fig. 1 *Stroboscopic image and selected line used to compose videostrobokymogram (upper) and videostrobokymogram of the selected line (lower)*

were automatically selected and displayed along the vertical direction so that the vocal fold vibration can be described vertically as a waveform (Fig. 1). By selecting different lines of interest, the vibratory patterns of multiple regions can be displayed sequentially (Fig. 2). The program was designed with backward and forward buttons so that users were able to choose any one frame among the grabbed dynamic images. When each videostrobokymogram was displayed in a separate window, the selected horizontal line and the window number were presented on the original stroboscopic image. User-friendly software was designed for this videostrobokymographic system. The user interface of this system is shown in Fig. 2.

3 Initial trials

To test the clinical applicability of this system, videostrobokymographic investigations were performed in normal subjects and patients with vocal fold disorders. While the subjects phonated constant vowel /i/, video images of vocal fold movement were recorded using a commercial videostroboscopy system§ with a 70° telescope. The images were recorded using CCD camera attached to the end of a laryngoscope. The fundamental frequency of vocal fold vibration was detected using a condenser type microphone placed on each of the thyroid alae, and introduced to the stroboscopy so as to trigger the stroboscopic light source.

Videostroboscopic images and videostrobokymograms of a normal subject are shown in Fig. 3. The upper row (Fig. 3*a*) shows stroboscopic images corresponding to one cycle of vibration. The numbers on the second image (line b–d) represent the selected lines for the reconstruction of the videostrobokymograms in the lower row (Figs 3*b*, *c*, and *d*). The videostrobokymograms show regular periodic waveforms, which are represented as symmetric diamond shapes in the areas of open period.

Fig. 4 shows the images of a patient with sulcus vocalis. Sulcus vocalis is a condition in which a furrow along the edge of the vocal fold causes dysphonia. The furrow (indicated by the arrow) located in the internal margin of the left vocal fold (right side in Fig. 4) was seen in the stroboscopic image. Minimal touch of both mucosal edges in the posterior part of the glottis (upper side in Fig. 4) could be seen in the videostrobokymogram (Fig. 4*b*) as

* Raptor, Bit flow Inc., USA.
† Microsoft Co., Redmond, WA.
‡ Microsoft Co.

§ Kay Elemetrics, Lincoln Park, NJ.

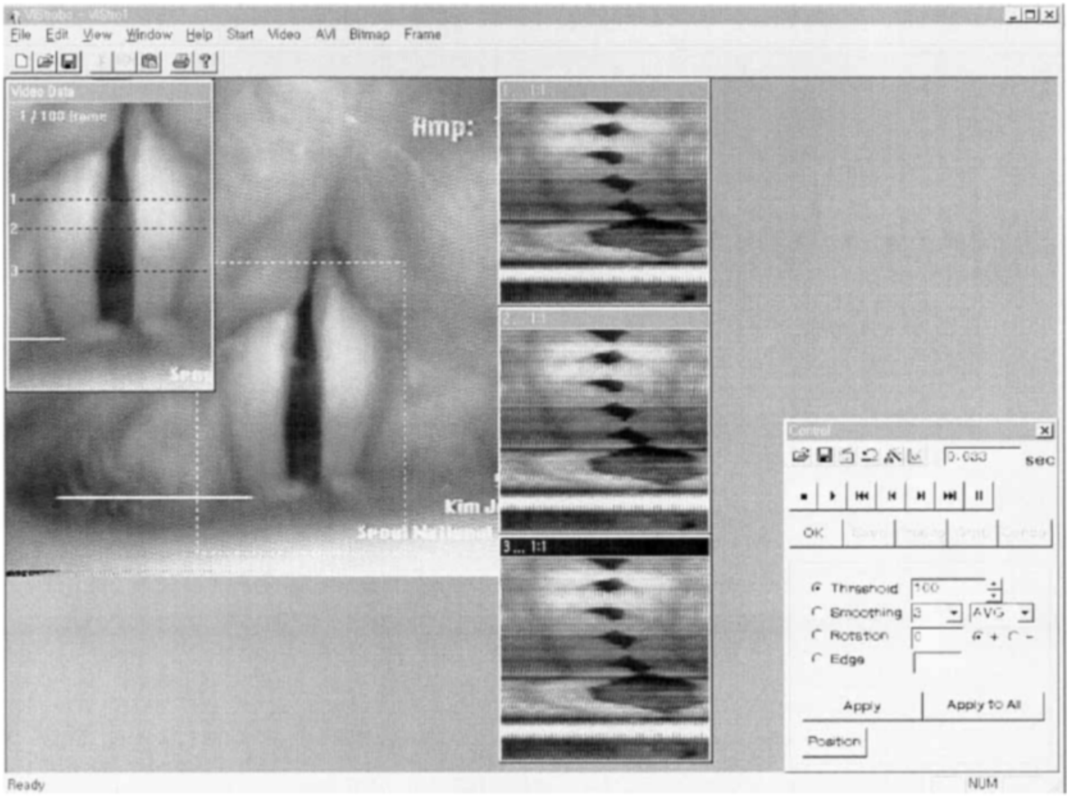

Fig. 2 *User interface of videostrobokymographic system*

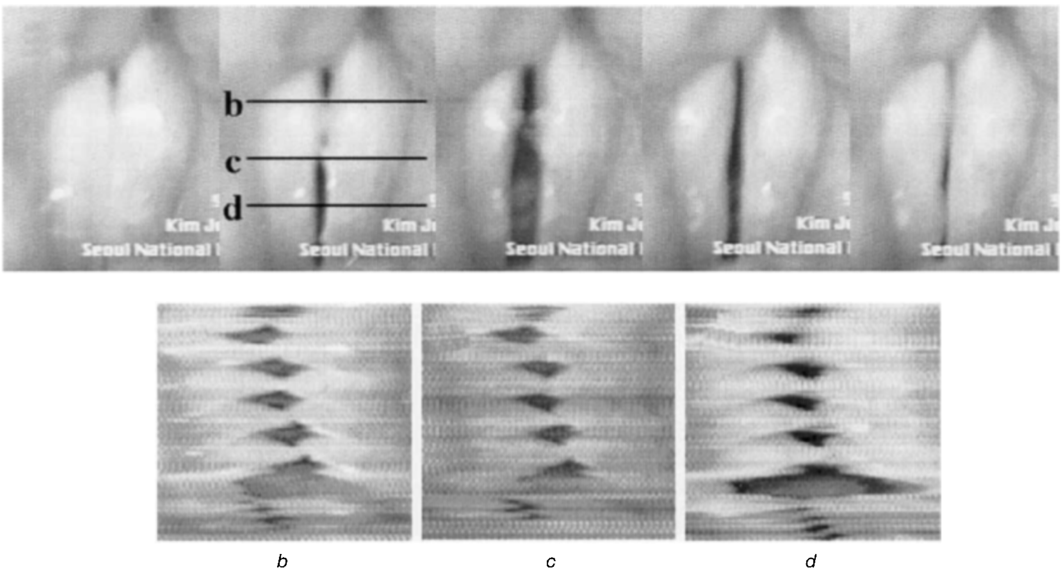

Fig. 3 *Stroboscopic images and videostrobokymograms of a normal control. (a) Stroboscopic images: regular vibratory pattern is shown; (b), (c), (d) videostrobokymograms of the lines represented in (a): regular periodic waveforms are shown*

well as in the stroboscopic image (Fig. 4*a*). Moreover, it was found that the vibration of the mucosal edges in this area was not synchronised with the middle and lower part of the videostrobokymogram. The videostrobokymograms from lines that cross the furrow (Figs 4*c* and *d*) show that the edge of the left vocal fold (right side in Fig. 4) did not move to the right (left side in Fig. 4).

We applied videostrobokymography to the preoperative and postoperative videostroboscopic images of 15 patients with Reinke's edema ($M/F = 11/4$, mean age $= 55.0$ years). Reinke's edema of the vocal fold is defined as a swelling or accumulation of fluid confined to the space underneath the epithelial layer, known as Reinke's space. The mass of the vocal folds is greatly increased, and abnormally low fundamental frequency and high subglottic pressure are the characteristic findings from voice analysis. The patients underwent laryngomicrosurgery with mucosa-sparing technique, and a postoperative videostroboscopic examination was performed 4 to 32 months (mean 12.4 months) after the operation.

On preoperative examination, videostrobokymograms showed an irregular pattern of mucosal wave and, in some patients, a relatively constant glottic gap during phonation. Open quotients (see Section 4.2 for the definition of open quotient) derived from the videostrobokymograms were quite different, even in the same patient, dependent on the location along the vocal folds. The most edematous portion had a very small OQ, however, the portion posterior to the edema had a very large OQ. After the operation, the voice quality of all patients was improved in acoustic and aerodynamic assessments, and videostrobokymography showed apparent improvements in mucosal movement.

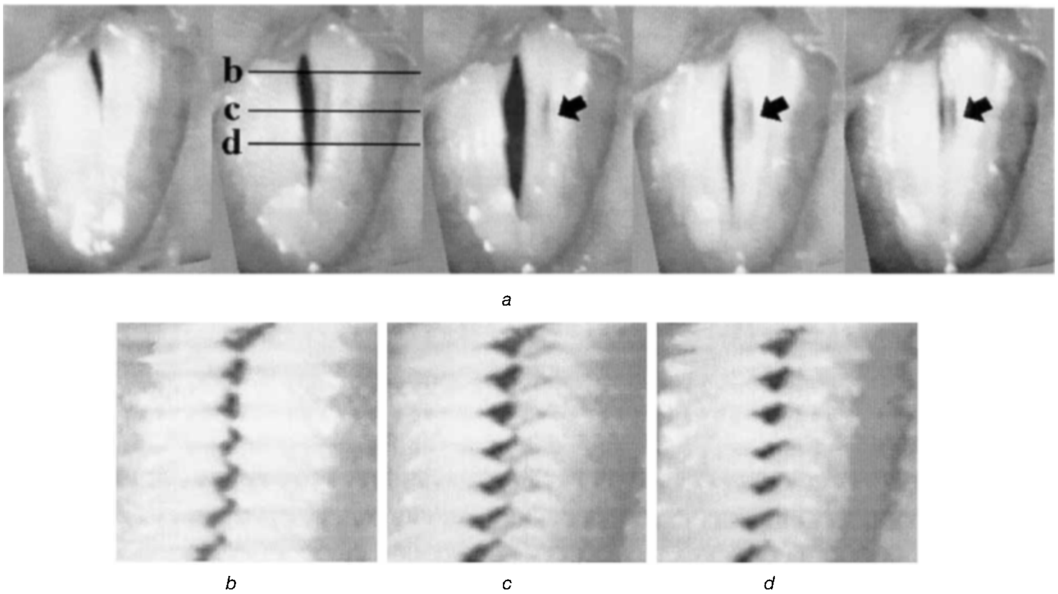

Fig. 4 *Stroboscopic images and videostrobokymograms of a patient with sulcus vocalis. (a) Stroboscopic images: a furrow along the internal margin of the vocal fold is shown; (b), (c), (d) videostrobokymograms of the lines represented in (a): this irregular pattern is well described*

4 Image processing and parameter estimation

4.1 *Rotation to compensate the obliqueness of glottal area*

The obliqueness of the glottal areas can result in horizontal mismatch of the mirrored region in the resulting videostrobokymogram. This obliqueness might be due to not only the relative orientation of the patient's head and laryngoscope during the recording of videostroboscopy, but also the camera position on the telescope, or could also be inborn. Prior to composing the videostrobokymogram, each frame was rotated automatically so that the longitudinal axis of glottal area could align with the y-axis of monitor screen. To calculate the angle between both axes, the maximum glottal area was assumed to be an ellipsoid and the second-order moment of this area was computed. The longer principal axis of the ellipsoid was considered as the longitudinal axis of glottal area. The glottal area was defined as the region enveloped within the margin of both vocal folds and the maximum glottal area was extracted from the frame in which the vocal folds were opened to maximum extent. The procedure to calculate the rotation angle is as follows (Fig. 5) (BALLARD and BROWN, 1982).

First, we extracted the glottal area (Fig. 5*b*) using a simple threshold method (pixels with 30% or less of the global maximum pixel value were considered to be glottal area) and defined the binary matrix representing the geometric shape of glottal area as follows:

$$B(x, y) = \begin{cases} 1 & \text{if } (x, y) \text{ is in the glottal area} \\ 0 & \text{if } (x, y) \text{ is not in the glottal area} \end{cases} \quad (1)$$

Then, the co-ordinates of the centre of mass (x_g, y_g) and the second moments of the glottal area (I_{xx}, I_{yy}, I_{xy}) were calculated as follows:

$$x_g = \frac{\sum_{x,y} xB(x, y)}{\sum_{x,y} B(x, y)} \quad (2)$$

$$y_g = \frac{\sum_{x,y} yB(x, y)}{\sum_{x,y} B(x, y)} \quad (3)$$

$$I_{xx} = \sum_{x,y} (x - x_g)^2 B(x, y) \quad (4)$$

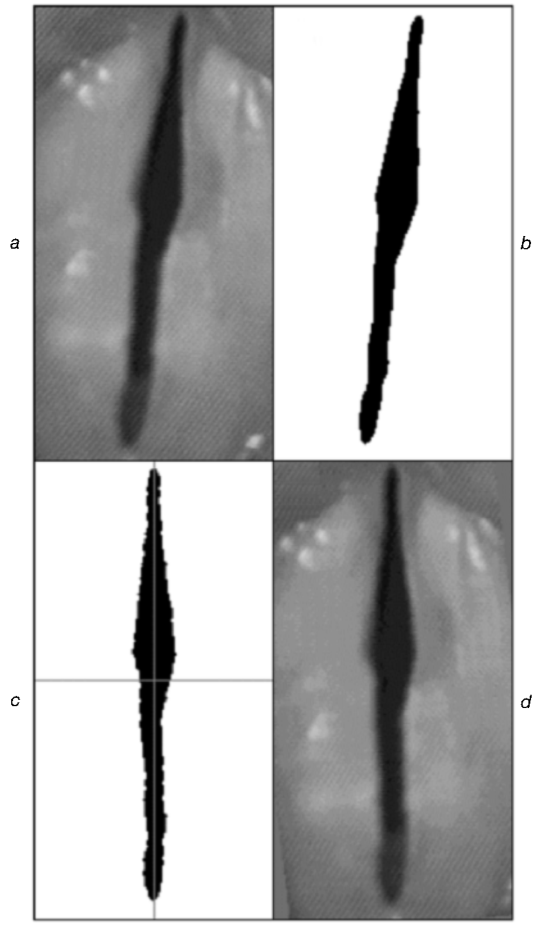

Fig. 5 *Procedure to compensate for the obliqueness of vocal folds. (a) Stroboscopic image with maximum glottal area; (b) binary image of glottal area extracted using simple thresholding method; (c) automatically rotated binary image of glottal area; (d) obliqueness-compensated stroboscopic image*

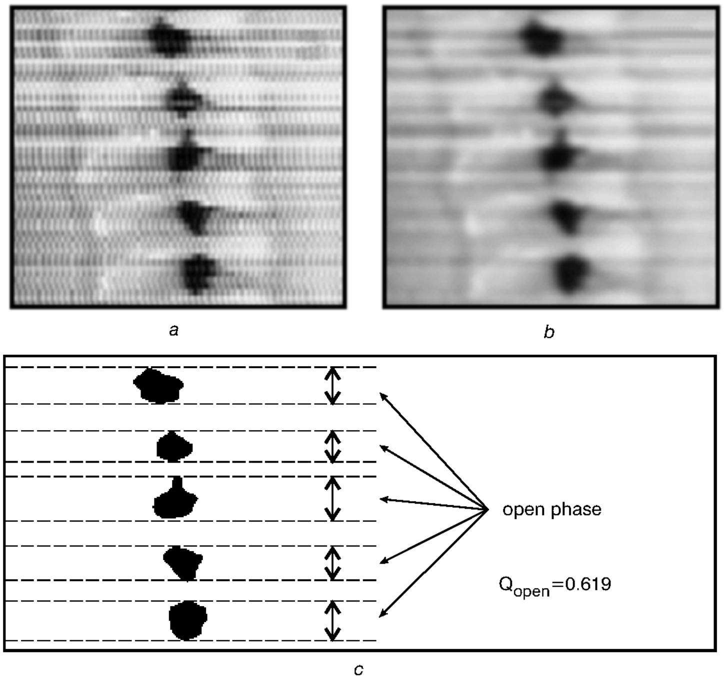

Fig. 6 *Procedure for calculating open quotient. (a) Original; (b) smoothed; (c) segmented videostrobokymogram and phase transition*

$$I_{yy} = \sum_{x,y}(y - y_g)^2 B(x, y) \tag{5}$$

$$I_{xy} = \sum_{x,y}(x - x_g)(y - y_g)B(x, y) \tag{6}$$

With the second moments, the rotation angle between the longitudinal axis of the glottal area and the y-axis of monitor screen was calculated using the following equation:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2I_{xy}}{I_{xx} - I_{xy}}\right) + \frac{\pi}{2} \tag{7}$$

4.2 *Open quotient*

Open quotient is the period of time the vocal folds are separated (open phase) relative to the entire duration (total phase) of a single cycle of vocal fold vibration, and was calculated by analysing the videostrobokymogram. The procedure is shown in Fig. 6 The original videostrobokymogram (Fig. 6*a*) was smoothed (Fig. 6*b*) using a Gaussian kernel and segmented using a simple thresholding method. From the resulting binary image (Fig. 6*c*), the number of frames of open and closed phases was estimated. The black areas in the segmented image corresponded to the open phase. Open quotient was calculated using the equation below, where Q_{open} is the open quotient and where *open phase*, *closed phase* and *total phase* are the numbers of frames of the corresponding phases.

$$Q_{open} = \frac{open\ phase}{total\ phase} = \frac{open\ phase}{open\ phase + closed\ phase} \tag{8}$$

5 Discussion and conclusions

The present study reports technical details concerning the videostrobokymography (Sung *et al.*, 1999) developed to assess the regional vibratory pattern of vocal folds, in which video recordings of stroboscopic examinations were analysed to visualise this pattern as a static image. A user-friendly windows program was designed to help the examiner to perform the operation more carefully in a more relaxed environment after stroboscopic examination so that accurate and meaningful analysis might be possible.

In the videostrobokymogram, limitations related to the use of stroboscopic light, allow investigation only of nearly-periodic vibrations of the vocal folds. Irregularity in the vibration pattern of patients with severe voice disorders will lead to a random 'stroboscopical' image frame selection and degrade the quality of the videostrobokymogram information. In contrast, the videokymography of Švec and Shutte (1996) detects and displays irregular vibration patterns realistically, because it does not rely on the stroboscopy effect and it displays all cycles. Although this difference may be considered a major drawback of the videostrobokymogram, nevertheless, the videostrobokymogram gives a helpful plot of sequential oscillations on a single two-dimensional figure. That is indeed an advantage, because it enables the information from a dynamic process to be visualised on a single print-out. Moreover, videostrobokymography enables evaluation of the vibrations of several vocal fold regions at the same time, and also allows one to obtain objective parameters to describe the vibration patterns.

The user interface and detailed procedures for videostrobokymography, and the mathematical equations provided in this paper will be helpful to investigators who want to reproduce similar systems to use in monitoring and analysing pathologic changes in the movement of vocal folds.

Acknowledgments—This work was supported in part by the Ministry of Health and Welfare under grant HMP-95-G-1-16, and in part by BK21 Project for Human Life Sciences.

References

Ballard, D. H., and Brown, C. M. (1982): 'Computer vision' (Prentice-Hall, Englewood Cliffs, NJ), pp. 254–261

Blitzer, A., Brin, M. F., Sasaki, C. T., Fahn, S., and Harris, K. S. (1992): 'Neurologic disorders of the larynx' (Tieme Medical Publishers, New York)

FUJIMURA, O. (1988): 'Vocal physiology: voice production, mechanisms and functions' (Raven Press, New York), pp. 225–345
GALL, V., and HANSON, J. (1973): 'Bestimmung physikalischer Parameter der Stimmlippenschwingungen mit Hilfe der Larynxphotokymographie', *Folia Phoniartr* (Basel), **25**, pp. 450–459
GAUFFIN, J., and HAMMARBERG, B. (1991): 'Vocal fold physiology. Acoustic, perceptual, and physiological aspects of voice mechanism' (Singular Publishing Group, San Diego), pp. 213–216
GROSS, M. (1985): 'Larynxfotokymographie', *Sprache-Stimme-Gehör*, **9**, pp. 112–113
HIRANO, M., and BLESS, D. (1993): 'Videostroboscopic examination of the larynx' (Singular Publishing Group, San Diego), pp. 23–24
KARNELL, M. P. (1989): 'Synchronized videostroboscopy and electroglottography', *J. Voice*, **3**, pp. 68–75
ISOGAI, Y. (1994): 'Laryngostrobography by the newly developed strobo-motion analyzer', *Larynx Japan*, **6**, pp. 91–96
ISOGAI, Y. (1996): 'Laryngostroboscopy-observation of the traveling wave of the mucous membrane', *Larynx Japan*, **8**, pp. 84–91
SUNG, M.-W., KIM, K. H., KOH, T.-Y., KWON, T.-Y., MO, J.-H., CHOI, S.-H., LEE, J. S., PARK, K. S., KIM, E. J., and SUNG, M. Y. (1999): 'Videostrobokymography: a new method for the quantitative analysis of vocal fold vibration', *Laryngoscope*, **109**, pp. 1859–1863
ŠVEC, J. G., and SCHUTTE, H. K. (1996): 'Videokymography: high-speed line scanning of vocal fold vibration', *J. Voice*, **10**, pp. 201–205
TRAPP, T. K., and BERKE, G. S. (1988): 'Photoelectric measurement of laryngeal paralyses correlated with videostroboscopy', *Laryngoscope*, **98**, pp. 486–492
WENDLER, J., KÖPPEN, K., and FISCHER, S. (1986): 'Validity of stroboscopic data in terms of quantitative measures', Proc. Int. Conf. Voice, pp. 36–43
WOO, P. (1996): 'Quantification of videostrobolaryngoscopic findings – measurements of the normal glottal cycle', *Laryngoscope*, **106**, pp. 1–27

Authors' biographies

JAE SUNG LEE received his BS in Electrical Engineering in 1996, and MS in 1998 and PhD in 2001 in Biomedical Engineering from Seoul National University, Seoul, Korea. Since 1996, he has been involved in designing and development of software and hardware equipments for the diagnosis of laryngeal disorders. He has also been a Research Member at the Department of Nuclear Medicine and the Institute of Radiation Medicine, Seoul National University College of Medicine since 1995, and was a Visiting Scientist at the Technische Universität München, Munich, Germany in 1999.

KWANG SUK PARK graduated from Seoul National University in 1980. He received his MS in 1982, and PhD in 1985 from Seoul National University in Electronic Engineering. He is now Chairman and Associate Professor in the Department of Biomedical Engineering, Seoul National University College of Medicine. His research interests are biomedical signal and information processing.