



<b>Title</b>	<b>A high-speed quantitative analysis of vocal fold vibration in normal and dysphonic subjects</b>
<b>Other Contributor(s)</b>	<b>University of Hong Kong.</b>
<b>Author(s)</b>	<b>Cheung, Pui-yin, Natalie</b>
<b>Citation</b>	
<b>Issued Date</b>	<b>2007</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/55491">http://hdl.handle.net/10722/55491</a></b>
<b>Rights</b>	<b>Creative Commons: Attribution 3.0 Hong Kong License</b>

**A High-speed Quantitative Analysis of Vocal Fold Vibration in Normal and Dysphonic  
Subjects**

**Cheung Pui Yin, Natalie**

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science  
(Speech and Hearing Sciences), The University of Hong Kong, June 30, 2007.

# **A High-speed Quantitative Analysis of Vocal Fold Vibration in Normal and Dysphonic**

## **Subjects**

Cheung Pui Yin, Natalie

## **Abstract**

The study investigated the periodicity of vocal fold vibration quantitatively. Modal register in normal and dysphonic (pathological) subjects was investigated. Computer system of High Speed Video Processing (HSVP) program was used in the high-speed laryngoscopic analysis, which captured 1941 frames per second. Eight ratio indices representing glottal area, length, and width were computed. Five ratio indices were found to be distinctive indicators of dysphonia. They were glottal area ratio index, left glottal area ratio index, right glottal area ratio index, right glottal width ratio index, and posterior glottal length ratio index.

## INTRODUCTION

Vibration of the vocal folds is a high-speed and complex movement. It determines the quality of voice. Understanding such vibratory characteristics is crucial in studying voice production and for clinical diagnosis of voice disorders.

For a healthy voice, both vocal folds must be of equal tension and thickness in order to produce a symmetrical or periodic oscillation during phonation. During adduction, both vocal folds must meet at the midline. Any disturbance of vocal fold vibrations causes asymmetrical or aperiodic oscillation, and resulted in dysphonic voice qualities. For example, hoarseness is commonly regarded as turbulences and air loss due to incomplete glottic closure, and irregular vibrations of the vocal folds (Fig 1). The important vibratory characteristics include different vocal folds dimension in open phase, opening phase, closed phase, closing phase, mucosal waves propagation, and left-right vocal folds vibratory phase differences and amplitudes.

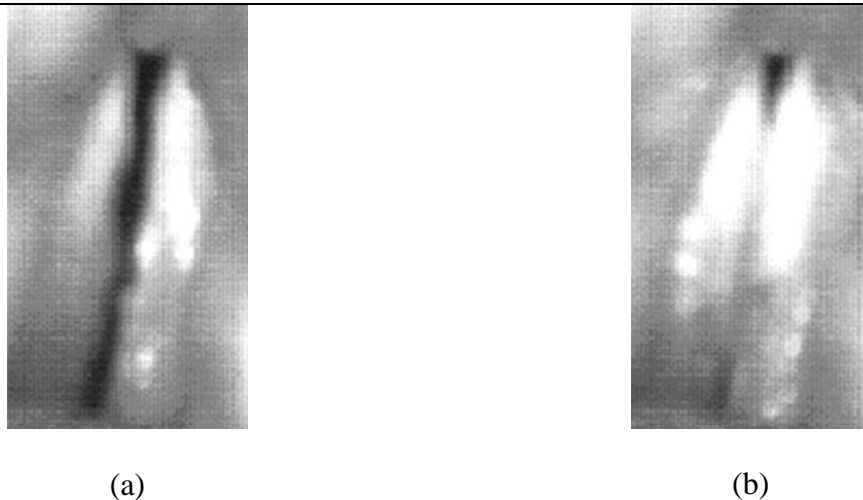


Figure 1. The open phase (a) and closed phase (b) taken from a subject with hoarse voice. It demonstrates the irregular vocal folds vibration.

---

At normal human pitch, the vocal folds oscillate with a fundamental frequency of 100-400 Hz. As the human visual perception is restricted to a time resolution of about 15 images per second, advanced medical measurements and imaging technologies are needed to investigate phonation qualitatively and quantitatively. High-speed laryngoscopy is a more recent technology to acquire physiological data through direct visual inspection of the sound source using digital recordings during voice production at a high speed. It can capture from 2000 to 8000 (depending on the system) consecutive frames per second, which is fast enough to resolve vibrations of the vocal folds during sustained phonation of a vowel sound. This allows understanding of phonatory vocal fold dynamics possible.

Previously, there were a number of drawbacks in high-speed imaging, e.g., limited recording time capacity, time-consuming analysis procedures, illumination problems. Technological advancement brings the visualization of vocal fold dynamics possible in an accurate and effective manner. It gives automatic and precise quantification of asymmetry, degree of glottic closure, glottic area, periodicity of vibration, and speed quotients.

Some parameters defined and computed using different high-speed image processing program in laryngoscopic study have been found to be good indicators for different vibratory patterns across different types of phonation and articulation these years. Koster et al. studied the time of vibration onset in 1999. The result revealed that there was a large individual difference of vocal folds oscillation at voice onset of syllables with variable breathiness and stress. In another study of different aspirated and unaspirated stop consonants, the voice onset of glottal vibrations, the

subglottal pressure during phonation, and the prephonatory glottal width were found to be crucial linguistic content of speech (Hong, Kim, & Niimi, 2002). In a study about people with unilateral vocal fold atrophy (Hertegard, Larsson, & Wittenberg, 2003), it was revealed that hoarse and breathy voice was due to desynchronization between two folds. This finding was demonstrated by automatic analysis of glottal area variations using high-speed technique. Lindestad et al. investigated Mongolian throat singing in 2001. They found that ventricular fold vibration involvement was present during such kind of throat singing, resulted in increased subharmonics and reduced glottal flow in certain vibratory periods.

High-speed laryngoscopy has been used with other technologies to analyze mechanisms of different phonation as well. Parallel laser beam projection at the vocal fold together with high-speed imaging technique was used in one of the voice research done by Hoppe et al. in 2003 and Schuberth et al. in 2002. The result of Hoppe et al. (2003) demonstrated that the amplitude of glottal vibrations and glottal area variations decreased with increased pitch during glissando phonation. They found that vocal fold tension was one of the contributions to regulate vocal fold length. That was, the vocal fold was lengthened by increased isometric tension on folds, and hence increased the pitch. Similar result was found in Schuberth's study, i.e., membranous vocal fold length increased with increasing fundamental frequency. In another project done by Granqvist et al. (2003), they combined high-speed system with inverse filtering and electroglottography. It was found that there was a delay between glottal flow waveform and glottal area waveform during glottal closure. It was suggested that the breathy voice was due to leakage offset of airflow during the closed phase of vocal fold vibration, which resulted in decreased acoustic energy of voice.

Computer program of High Speed Video Processing (HSVP) program is used as high-speed analysis in this study. It tries to summarize these vibration properties using eight ratio indices, representing glottal area, length, and width.

1. Glottal Area Ratio Index (Fig 2a): It represents the mean glottal area of all vibration cycles in the image. It is calculated as the average of minimum glottal area divided by the maximum glottal area. Theoretically, zero glottal area should be obtained in closed phase while non-zero glottal area should be obtained in open phase. Therefore, the larger the index indicates the larger the glottis in closed phase or closing phase, and vice versa.
2. Left Glottal Area Ratio Index (Fig 2b): It represents the mean left halve of the full glottal area measured from the vertical midline to the left glottis edge. It is calculated as the average of minimum left glottal area divided by the maximum left glottal area. Theoretically, zero left glottal area should be obtained in closed phase while non-zero left glottal area should be obtained in open phase. Therefore, the larger the index indicates the larger the left glottis in closed phase or closing phase, and vice versa.
3. Right Glottal Area Ratio Index (Fig 2c): It represents the mean right halve of the full glottal area measured from the vertical midline to the right glottis edge. It is calculated as the average of minimum right glottal area divided by the maximum right glottal area. Theoretically, zero right glottal area should be obtained in closed phase while non-zero right glottal area should be obtained in open phase. Therefore, the larger the index indicates the larger the right glottis in closed phase or closing phase, and vice versa.
4. Left Glottal Width Ratio Index (Fig 2d): It represents the mean left halve of the full glottal width measured from the vertical midline to the left glottis edge. It is calculated as the

average of minimum left glottal width divided by the maximum left glottal width. Theoretically, zero left glottal width should be obtained in closed phase while non-zero left glottal width should be obtained in open phase. Therefore, the larger the index indicates the wider the left glottis in closed phase or closing phase, and vice versa.

5. Right Glottal Width Ratio Index (Fig 2e): It represents the mean right halve of the full glottal width measured from the vertical midline to the right glottis edge. It is calculated as the average of minimum right glottal width divided by the maximum right glottal width. Theoretically, zero right glottal width should be obtained in closed phase while non-zero right glottal width should be obtained in open phase. Therefore, the larger the index indicates the wider the right glottis in closed phase or closing phase, and vice versa.
6. Anterior Glottal Length Ratio Index (Fig 2f): It represents the mean anterior halve of the full glottal length measured from the horizontal midline to the upper glottis edge. It is calculated as the average of minimum anterior glottal length divided by the maximum anterior glottal length. Theoretically, zero anterior glottal length should be obtained in closed phase while non-zero anterior glottal length should be obtained in open phase. Therefore, the larger the index indicates the longer the anterior glottis in closed phase or closing phase, and vice versa.
7. Posterior Glottal Length Ratio Index (Fig 2g): It represents the mean posterior halve of the full glottal length measured from the horizontal midline to the lowest glottis edge. It is calculated as the average of minimum posterior glottal length divided by the maximum posterior glottal length. Theoretically, zero posterior glottal length should be obtained in closed phase while non-zero posterior glottal length should be obtained in open phase. Therefore, the larger the index indicates the longer the posterior glottis in closed phase or closing phase, and vice versa.



8. Length Width Ratio Index (Fig 2h): It represents the shape of the glottis. It is calculated as the average minimum glottal length-width ratio divided by the maximum glottal length-width ratio.

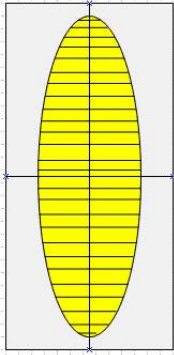


Fig 2a. Total glottal area.

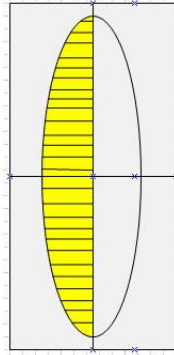


Fig 2b. Left glottal area.

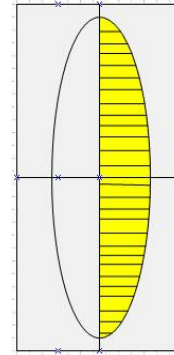


Fig 2c. Right glottal area.

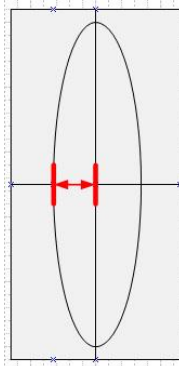


Fig 2d. Left glottal width.

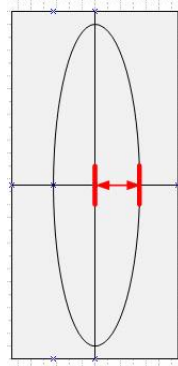


Fig 2e. Right glottal width.

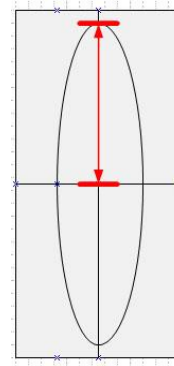


Fig 2f. Anterior glottal length

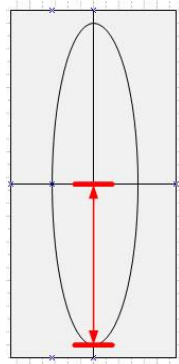


Fig 2g Posterior glottal length

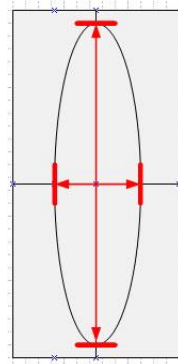


Fig 2h The length-width ratio

Theoretically, asymmetry of the vocal fold vibration can be inferred from significant differences of the indices. For example, one would expect a significant difference between left glottal area ratio index and right glottal area ratio index in a dysphonic patient with left vocal polyps due to incomplete adduction in maximum glottal closure. Also, there would be a significant difference in left glottal width ratio index between people with normal voice and people with left vocal polyps due to pathological changes of voicing tissues.

This study presented a quantitative analysis of vocal fold vibration of modal register in normal and dysphonic subjects by using high-speed laryngoscope. Analysis was carried out using new software of High Speed Video Processing (HSVP) program. Eight ratio indices were compared between normal and dysphonic groups, i.e., glottal area ratio index, left glottal area ratio index, right glottal area ratio index, left glottal width ratio index, right glottal width ratio index, anterior glottal length ratio index, posterior glottal length ratio index, and glottal length width ratio index.

## **METHODS**

### **Participants**

Sixteen dysphonic subjects (six males and ten females) with various types of laryngeal pathologies and seventeen non-dysphonic subjects (five males and twelve females) with normal voices participated in this study. All subjects were native Cantonese speakers and were aged from 20 to 35 years (mean = 27.06, SD = 7.30).

None of the dysphonic subjects had received any voice therapy. They were diagnosed by ear-

nose-throat surgeons as having vocal fold pathologies. All normal subjects reported no history of voice disorders. Subjects were excluded from this study if they had previous vocal training, neurological diseases, habits of alcoholic drink and smoking, oro-facial abnormalities, or severe respiratory and allergies problems.

### **Perceptual voice severity rating**

The subjects' voices were evaluated subjectively by qualified speech therapists and the investigator (NC) prior to the laryngoscopic recording. A ten-point scale was used, in which "0" meant normal and towards the right side "10" meant severe.

### **High-speed laryngoscopic evaluation**

Digital high-speed recordings were made at the Division of Speech and Hearing Sciences in the University of Hong Kong. The rigid laryngoscopic process was performed by qualified speech therapist.

A Kay Elemetrics system was used to record sequences of laryngeal images from subjects with normal voice and voice disorders. A rigid endoscope (model Kay 9106) coupled to a camera (Model 9700) and a halogen light source (for illuminating the vocal folds was required. Parallel to the image series, the acoustical signal as well as the EGG was recorded. The microphone was attached to the computer of Windows XP for recording the synchronized sound files using software of Syntrillium's Cool Edit 2000.

During recording, subjects were asked to position properly. With sitting upright, leaning forward,

and extending the chin (Fig 3), the chair was adjusted in height so that the examiner's eyes were at the level of the oral cavity opening of the subject. Latex gloves and gauze were always used for holding subjects' tongues during the laryngoscopic examination. To prevent fogging of the scope lens, an antifog agent was applied. The subject was requested to stick out the tongue so that it could be held steady by the examiner with gauze. Each subject was first instructed to take a deep breath and then produce a sustained /i/ as long as he/she could at his/her most comfortable pitch and loudness. This vocal maneuver induced both the velar and larynx elevation and epiglottis retraction, which provided the best view of the vocal folds. Gag reflex was sometimes triggered when contact was made between the scope and either the soft palate, tongue, posterior pharyngeal wall, or all three. To help reducing this possible response, the scope was placed on the fingers holding the tongue so that the examiner was able to slide the tip into position for viewing without excitatory contact with these sensitive structures. Upon completion the videotape could be viewed and interpreted.

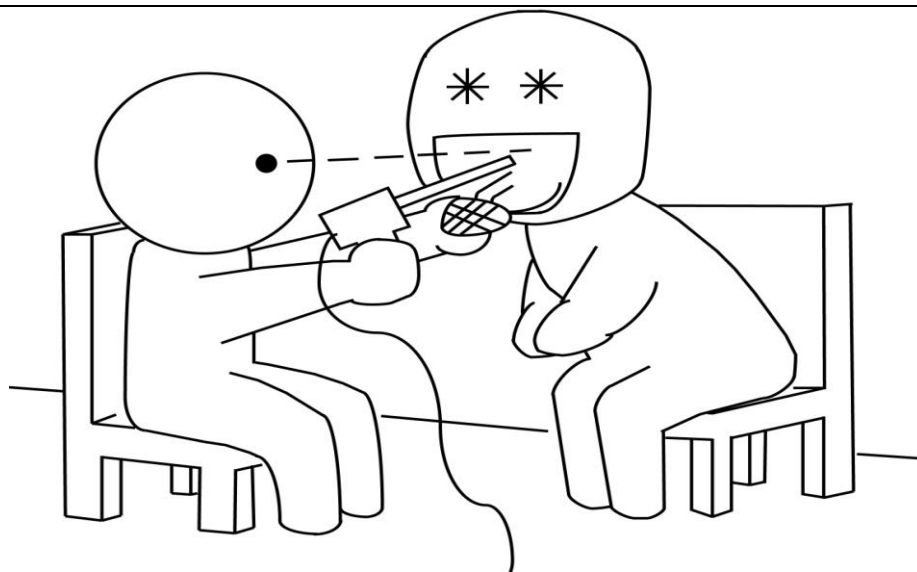


Figure 3. The best position of the subject during laryngoscopic recording.

---

## Data processing

Video images and synchronized sound signals were recorded in each subject. A total of 4368 frames covering a time span of 2.25 seconds were obtained in each subject. The files were then processed by High Speed Video Processing (HSVP) program.

The HSVP program was created by Professor Kong Jianping (from the Beijing University) and Professor Edwin Yiu (from the University of Hong Kong). Analysis included automatic glottal edge detection and calculation of variations in glottal area, glottal length, and glottal width. Four advanced functions were added in 2007 (shown as deep grey button in Fig 4), which increased the accuracy of glottis motion analysis.

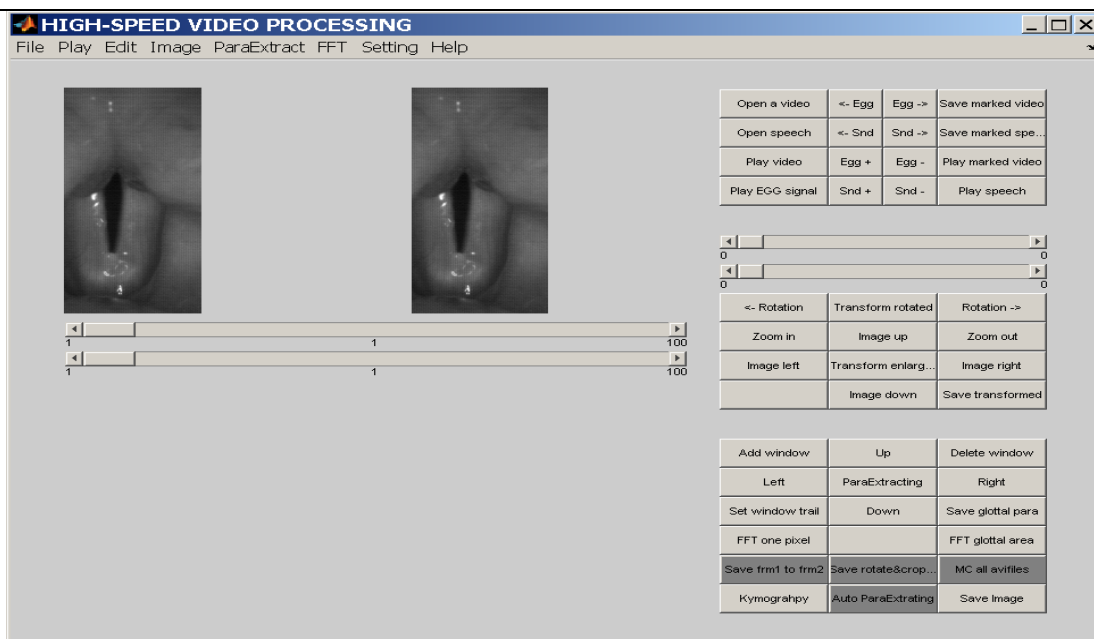


Figure 4. The interface of the HSVP program. Four advanced functions buttons were shown in deep grey colour on right bottom, i.e., “save frm1 to frm2”, “save rotate & crop”, “auto paraextracting”, and “MC all avifiles”.

General step sequences were done as follow:

1. Cropping (Fig 5a): About 1000 to 1500 frames with minimum of one hundred oscillations of vocal fold abduction/adduction cycles were required for later analysis. The most critical selection criterion was that the most stable one hundred oscillations with the full view of glottis were chosen. Clicking button “save frm1 to frm2” for saving the cropped frames.
2. Adjusting contrast of the images (Fig 5b): Contrast was adjusted such that the glottis and vocal folds were clearly identified for later analysis. This step was important because the program grasped the size of glottis by the contrast difference.
3. Rotating the images (Fig 5c): Vertical alignment of glottis was obtained for adding the analysis window in later analysis. Clicking button “save rotate and crop” for saving the adjustments.
4. Compensating drifted vocal fold oscillations: This step greatly increased the accuracy of analysis. Clicking button “MC all avifiles”, which meant motion compensation. It eliminated any unnecessary motion fluctuation of glottis during rigid laryngoscopic recording.
5. Marking glottis area for computing the eight parameters (Fig 5d): From the cropped 1500 frames, window trail was set on the frame with the biggest glottis. An analysis window, which was a rectangular box, was then added with the size of just fitted the glottis. Clicking button “set window trail” to confirm the placement and size of the analysis window, followed by clicking button “Auto ParaExtracting”. The eight ratio indices would then be computed. Clicking button “save glottal para”. The results would be saved in a new Excel file.

Ratio index was necessary in computation of the eight parameters. For the reasons of large

individual differences in laryngeal structures (e.g., length of vocal tract, size of vocal folds etc.) and unknown magnification factors (e.g., bigger laryngeal dimensions in men etc.), comparisons were made between minimum mean value and maximum mean values in each glottal dimension.

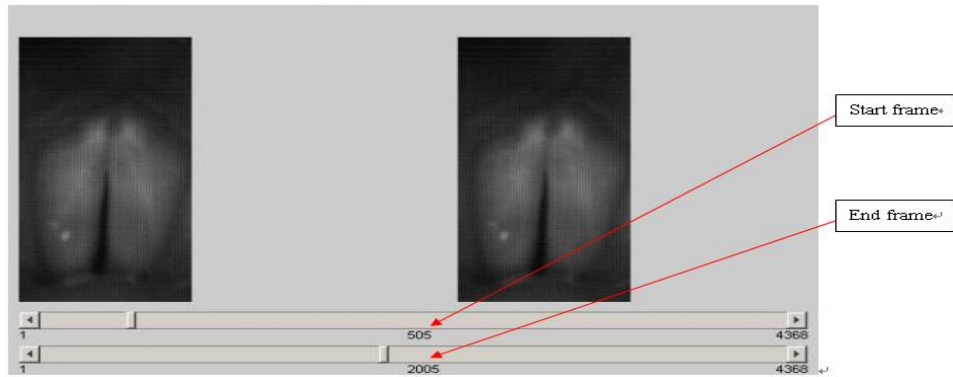


Figure 5a. Cropping: selected the most stable 1500 frames.

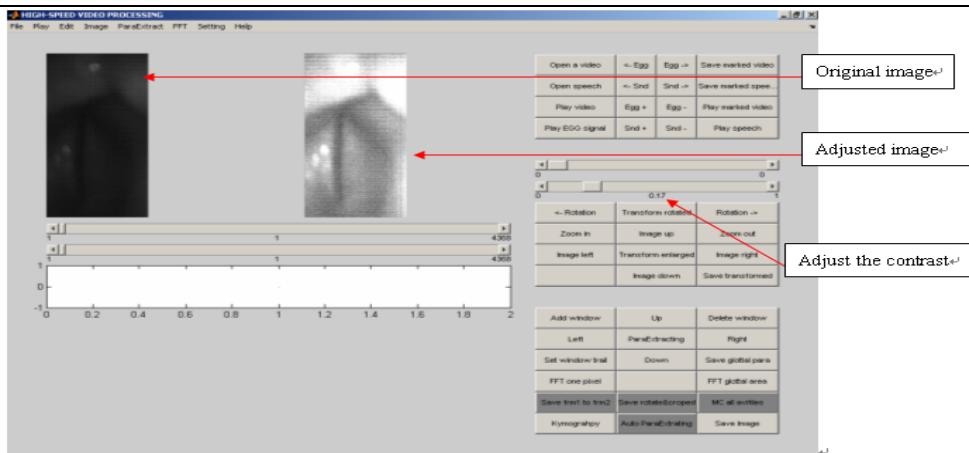


Figure 5b. Adjusting contrast.

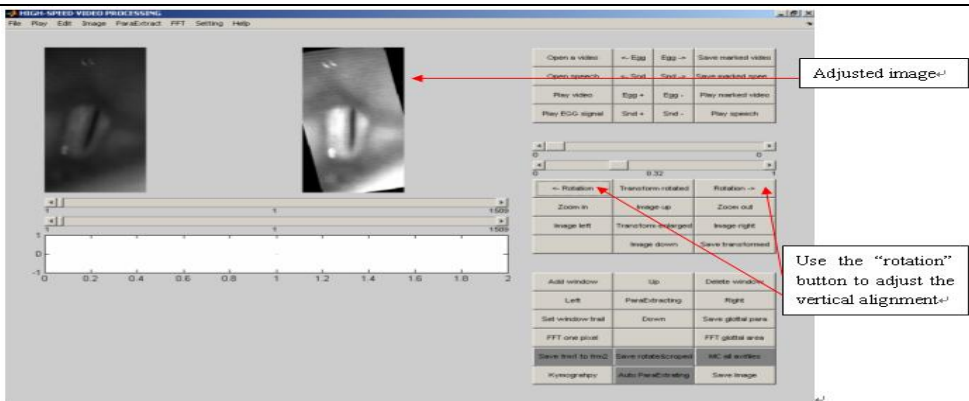


Figure 5c. Rotating the glottis into vertical position.

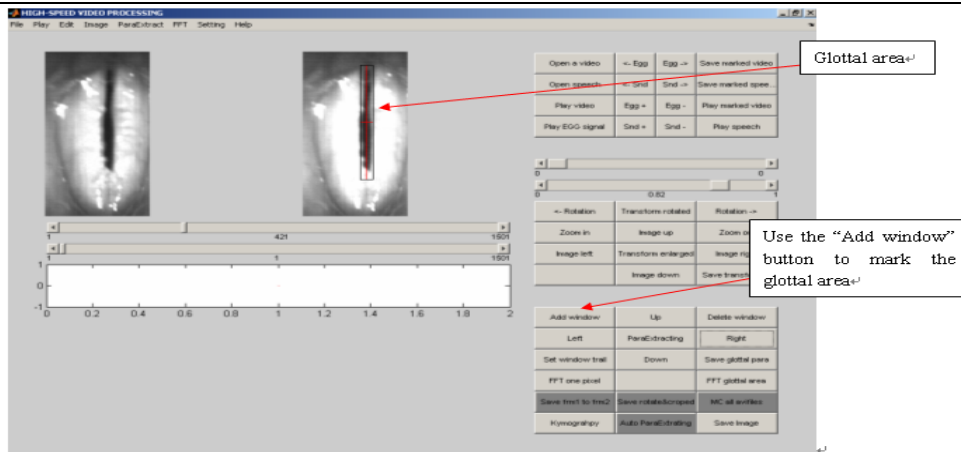


Figure 5d. Adding an analysis window and adjusting its size and position.

### Inter-rater and intra-rater reliability

The five steps of data processing were critical in the whole analytical process using the HSVP program. Reliability testing was designed by re-analyzing 20% (seven avi files) of the laryngoscopic data from step one to step five, stated above.

Inter-rater reliability test was administered by inviting another person, who was a speech and hearing sciences student, to re-analyze the randomly selected 20% data. Intra-rater reliability test was administered by the investigator (NC) to re-analyze the randomly selected 20% data after two weeks.

## RESULTS

The analysis was done for a portion of phonation /i/ at which a clear and stable image of the vocal fold was obtained. Such phonation portion was analyzed using HSVP program. Eight parameters (concerning the glottal shape) are in interest to see if there is significant difference between people with normal voice and people with dysphonia.



## Perceptual voice rating

The dysphonic group represents moderate to severe dysphonia, which have significantly more severe voice quality than the normal group. That is, dysphonic subjects have overall perceptual severity rating from 5.0 to 9.0 in a ten-point scale, while the normal subjects have rating from 0 to 1.0.

## Reliability tests of data analysis using HSVP program

Table 1 demonstrates the intrarater and interrater reliability of the data processing. All eight glottal shape indices presents good correlation with intrarater computation and interrater computation ( $P \leq 0.05$ ), which attains at least a correlation coefficient of 0.70. That is, intrarater correlation coefficient ranges from 0.76 to 0.97 and interrater correlation coefficient ranges from 0.77 to 0.97.

	Intrarater Reliability	Interrater Reliability
Glottal Area Ratio Index	*0.76*	*0.93*
Left Glottal Area Ratio Index	*0.97*	*0.92*
Right Glottal Area Ratio Index	*0.89*	*0.77*
Left Glottal Width Ratio Index	*0.88*	*0.82*
Right Glottal Width Ratio Index	*0.79*	*0.87*
Anterior Glottal Length Ratio Index	*0.91*	*0.97*
Posterior Glottal Length Ratio Index	*0.82*	*0.86*
Glottal Length-Width Ratio Index	*0.93*	*0.77*

\*  $P \leq 0.05$  (2 tailed) \*

Table 1. Spearman's Rho: The intrarater and interrater reliability test of data analysis process.

### HSVP data analysis between dysphonic group and normal group

Table 2 presents the mean of eight glottal ratio indices. Mann-Whitney *U* test was used to compare the mean differences between dysphonic group and normal group. The dysphonic group demonstrated significantly greater mean total glottal area, left glottal area, right glottal area, right glottal width, and posterior glottal length than the normal group for the phonation /i/ in most comfortable pitch and loudness ( $P \leq 0.05$ ).

There are two reasons of choosing Mann-Whitney *U* test as the non-parametric analysis to compare the mean rank (or median) between groups. The first reason is that the variables in this data set are measured on nominal and ordinal. The second reason is that the sample size is small (< 40 subjects).

	Dysphonic Group (N=16) Mean (SD)	Normal Group (N=17) Mean (SD)	Z	p
Glottal Area Ratio Index	0.44 (0.21)	0.25 (0.14)	-2.449	*0.014*
Left Glottal Area Ratio Index	0.44 (0.21)	0.27 (0.15)	-2.053	*0.040*
Right Glottal Area Ratio Index	0.43 (0.21)	0.24 (0.14)	-2.558	*0.011*
Left Glottal Width Ratio Index	0.18 (0.28)	0.02 (0.09)	-1.881	0.060
Right Glottal Width Ratio Index	0.19 (0.30)	0.01 (0.05)	-2.482	*0.013*
Anterior Glottal Length Ratio Index	0.47 (0.18)	0.36 (0.21)	-1.405	0.160
Posterior Glottal Length Ratio Index	0.34 (0.30)	0.04 (0.08)	-3.229	*0.001*
Glottal Length-Width Ratio Index	3.78 (2.44)	3.32 (2.38)	-0.648	0.517

\*  $P \leq 0.05$  (2 tailed) \*

Table 2. Wilcoxon-Mann-Whitney test: The eight glottal ratio indices between dysphonic group and normal group.

	Descriptive Statistics		Wilcoxon-Mann-Whitney test	
	Left Glottal	Right Glottal	Z	p
	Area	Area		
	Mean (SD)	Mean (SD)		
Normal Group (N=17)	0.27 (0.15)	0.24 (0.14)	-1.627	0.104
Dysphonic Group (N=16)	0.44 (0.21)	0.43 (0.21)	-1.548	0.122
	Left Glottal	Right Glottal		
	Width	Width		
	Mean (SD)	Mean (SD)		
Normal Group (N=17)	0.02 (0.09)	0.01 (0.05)	-1.667	0.095
Dysphonic Group (N=16)	0.18 (0.28)	0.19 (0.30)	-0.859	0.390
	Anterior	Posterior		
	Glottal	Glottal		
	Length	Length		
	Mean (SD)	Mean (SD)		
Normal Group (N=17)	0.36 (0.21)	0.04 (0.08)	-1.563	0.118
Dysphonic Group (N=16)	0.47 (0.18)	0.34 (0.30)	-2.355	*0.019*

\*  $P \leq 0.05$  (2 tailed) \*

Table 3. Wilcoxon-Mann-Whitney test: The symmetry of glottis (left-right comparison of glottal area and width, anterior-posterior comparison of glottal length).

### **HSVP data analysis about symmetry of glottis**

Table 3 presents mean comparison between left-right glottal area, left-right glottal width, and anterior-posterior length of normal group and dysphonic group respectively. The Wilcoxon-Mann-Whitney test was used to compare the mean differences of glottal shape between left and right, anterior and posterior respectively. The mean of left glottal area ratio index revealed no significant difference with the mean of right glottal area ratio index. The mean of left glottal width ratio index revealed no significant difference with the mean of right glottal width ratio index. However, the mean of anterior glottal length ratio index was significantly greater than the mean of posterior glottal length ratio index ( $P \leq 0.05$ ).

## **DISCUSSION**

Voice is the result of airflow pulses variation through different variations in vocal tract configuration. The combination effect of vocal fold tissue vibratory characteristics, subglottic pressure, vocal fold tension, and vibration amplitude gives variation in perceptual pitch and loudness as well as voice qualities. High-speed digital imaging in current research is trying to investigate these dynamic characteristics using a new analysis system of HSVP program.

### **High-speed data processing**

The eight ratio indices are computed through measuring area of glottis opening. The accuracy of analysis depends on extraction of the vocal fold opening from images of vibrating vocal folds. Efficient segmentation algorithm is essential to accurately delineate and calculate the glottis area on a frame-by-frame basis. A two-step algorithm (Deliyski, 2005) is applied in the HSVP

program to increase the accuracy of automatic vocal fold edge detection and compensation of fluctuation during recording, i.e., threshold selection and erosion operation.

Gray-level thresholding method is selected in HSVP program for image segmentation. The purpose of segmentation is to automatically separate the image into regions, that are in interest, from the whole picture. For the aim of HSVP program, it requires the extraction of glottis opening (region in interest) from the whole image of vocal folds. The first step is selecting a gray level value, which represents a dark cavity (glottis), based on first 200 frames in each recording. This value is estimated in the second step of data processing (contrast-adjustment). After an appropriate threshold value is determined, the whole laryngeal image sequence is segmented on a frame-by-frame basis. The second step is erosion operation. It removes isolated, mislabeled regions and hence improves the reliability of the segmentation. A clear glottis opening region is then generated automatically. Figure 6 demonstrates the result of image segmentation.

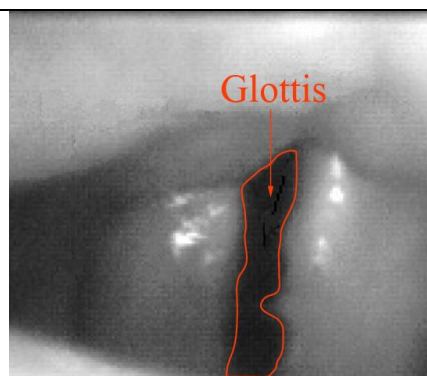


Figure 6. The resulted vocal fold diagram from image segmentation.

---

Motion compensation (MC) is another innovative improvement in HSVP program, which greatly

increases the accuracy of computation. The endoscopic motion (it means examiner's endoscope-holding instability and subjects' body movement in this paper) and changes in glottis during phonation have different dynamics. Such simulated motion creates a negative effect on reliability of computation of glottal region. Technically, MC function addressed this issue by improved detection accuracy of glottis edge while removed the distorting endoscopic motion. Figure 7 demonstrates the motion compensation function of HSVP program.

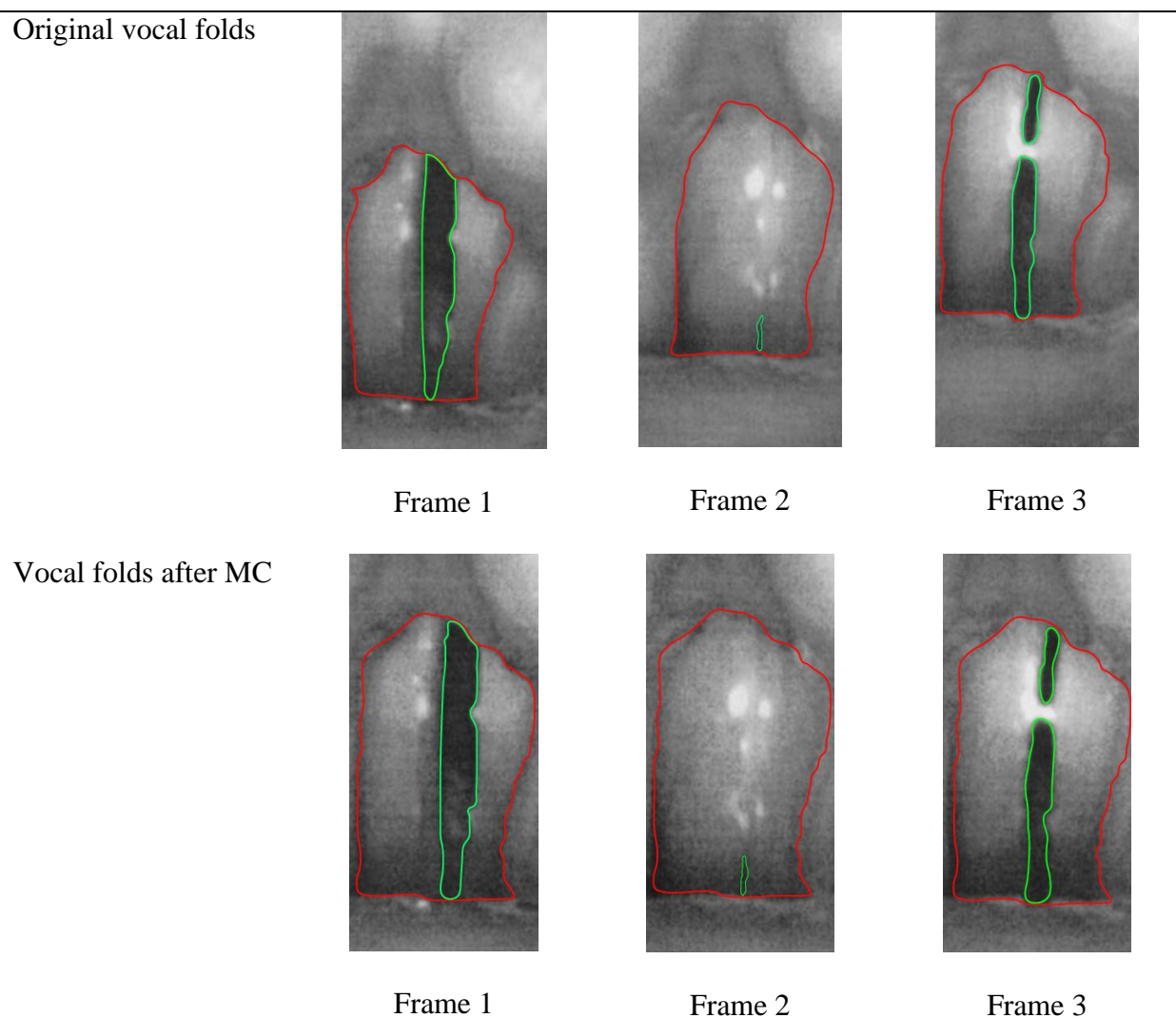


Figure 7. The function of motion compensation in HSVP program (from Deliyski, 2005).

---

### **Reliability on perceptual severity rating**

Throughout the analysis, categorical data between dysphonic group and normal group was the main interest. Therefore, both groups were straightly defined in two aspects, i.e., medical and perceptual. For medical aspect, all dysphonic subjects had pathological diagnosis from ear-nose-throat physicians. For perceptual voice quality aspect, all dysphonic subjects suffered from moderate and severe dysphonia. Moderately rated subjects were recruited as they exhibited significant difference of voice quality from normal population, which could be judged by trained and untrained listeners. A ten-point scale of overall perceptual severity rating was used and categorized into normal (range from 0.0 to 0.9), mild (range from 1.0 to 3.9), moderate (range from 4.0 to 6.9), and severe (range from 7.0 to 10.0). All subjects in dysphonic group are perceived to score with at least 5.0 in a ten-point scale of overall perceptual severity, which means that their voice are considered to be abnormal by both trained and untrained listeners.

### **Reliability on data processing**

The whole HSVP analysis was divided into five main steps, i.e., cropping, adjusting contrast, rotating, motion compensation, and marking the glottis. All steps were critical. Therefore, both intrarater reliability and interrater reliability were done by re-analyzing 20% of the total number of subjects from step one to step five. That was, the original 2.25 seconds videotape (containing 4368 frames) was being re-analyzed. The means of final eight ratio indices were compared.

Steps of cropping, contrast-adjustment, rotation, and marking the glottis are done manually by

subjective judgment of the analyzer. Also, the instrumental steps of motion compensation and ratio index computation are dependent on the previous step. In the step one of cropping, one thousand frames, which considered to be the most stable and with the full view of glottis, is selected from the 4368 frames and cropped for remaining steps. Frames chosen may differ within and across raters. In the step of contrast-adjustment, the brightness of the images is adjusted manually such that the glottis and vocal folds can be visually identified. This step is crucial because the HSVP program determined glottis shape by the contrast difference in the first two hundreds cropped frames (estimation of threshold value). The glottis is a dark area in the frame during the open phase. The third step of glottis-rotation is done in order to obtain a vertically aligned glottis for later step of analysis window placement. The default analysis window is a vertical rectangular box. It is divided into four equal parts by anterior-posterior axis. Although the size and placement can be adjusted, the best glottis shape is traced with the full inclusion of the glottis inside the analysis window. The last step of data processing is marking the analysis window on the glottis. The best size is considered to be just fit the glottis with minimal free edged margin. The area marked will be instrumentally computed to the eight ratio indices.

Together with the manual operations throughout the data analysis, the HSVP program seemed to work well. Satisfactorily, both intrarater and interrater comparisons of the eight ratio indices were statistically correlated.

### **High-speed quantitative analysis between dysphonic group and normal group**

The 2.25-second videotape of vocal fold vibration was recorded using high-speed laryngoscope. A total number of 4368 frames were collected for each subject in modal voice. Modal register,



also called chest register, covered the middle range of fundamental frequency range used in normal conversation (Hollien, 1974).

In this study, all subjects are young adults (mean age = 27.06, SD = 7.30). According to Colton and Casper (1996), human voice change continuously throughout life. Fundamental frequency is determined by age, gender, and race, in relation to physiological difference of vocal folds. Chinese young adults with age from twenty to thirty-five are recruited in this study. Both males and females became stable after age of twenty in terms of growth changes within the vocal tract that affects vocal tract resonance, rather than the vibratory pattern of vocal folds. Age-related changes (e.g., increased stiffness of respiratory and laryngeal structures, muscle and nervous system degeneration, reduced speed of neural transmission, and reduction in vocal tract mucous glands etc.) may occur from age of thirty-five.

Eight glottal ratio indices are computed for each subject and compared between the dysphonic group and normal group. Five mean ranks are found to be significantly greater in dysphonic group than normal group, i.e., glottal area ratio index, left glottal area ratio index, right glottal area ratio index, right glottal width ratio index, and posterior glottal length ratio index.

In dysphonic group, the three glottal area ratio indices are significantly greater. It indicated incomplete glottal closure in closed phase, overabduction in closing phase, or excessively short closure duration. The vocal pathologies may cause a certain extent of it, in which the outgrowing mass of vocal folds (e.g., vocal nodules) prevents the vocal folds from complete adduction. Adduction problem is suggested in dysphonic subjects. Therefore, both the left and right glottal

areas are greater in subjects with dysphonia than that with normal voice, and hence further contribute to the greater overall glottal area.

Another explanation is that the closed phase is much shorter in dysphonic group than that in normal subjects with regular vibration. Comfortable pitch and loudness were recorded in all subjects, at which the duration of glottal closure is supposed to be ideally performed in normal group. There is a significantly larger ratio index of total glottal area, left glottal, and right glottal area in dysphonic group. Besides adduction deficit, excessively short closure duration is also suggested to give such large value of glottal area in dysphonic group.

There is a significantly greater value of right glottal width ratio index in dysphonic group. It contributes one of the functions of vibration amplitude, which is determined as the distance from the glottal midline when maximally lateralized during oscillation. Amplitudes of the vocal fold vibrations are usually larger at low-pitch phonation. And in clinical practices, a large proportion of dysphonic population complained their voice were low-pitched. It explains the significantly greater width in dysphonic group. Another possibility is the incomplete adduction of vocal folds, stated in previous paragraph. The third possibility is that the unevenness mass of vocal fold delimits the glottal width during open phase. The structures are not moving at a regular rate. In other words, aperiodic vibration of vocal fold was demonstrated. According to Sundberg (1995), one possible reason for that would be a difference in vibration phase and velocity between the upper and lower parts of the glottal mucosa.

The posterior glottal length ratio index is statistically greater in dysphonic group than that in

normal group. According to the description of vocal fold length from Hirano and Diane (Hirano & Diane, 1993), it is the distance between the insertion at the thyroid cartilage and the processes vocalis of the arytenoids cartilage. It is determined during the open phase of the vibration cycle, when the medial tip of the processes vocalis is detectable. Such finding indicates that the closure of the posterior glottis is imperfect in dysphonic subjects during adduction, which leads to excessively short closed phase or excessively long opening and open phase.

### **High-speed quantitative analysis on symmetry of glottis**

The vocal fold vibratory pattern should be symmetrical in people with normal voice. Asymmetry may indicate disturbances in any voicing parameters, e.g., vibratory phase synchrony, glottal competency, mass and tension variability during phonation etc. Both horizontal symmetry (left-right vocal fold comparison) and vertical symmetry (ventral-dorsal vocal fold comparison) are studied.

The results revealed that there is not any statistically significant difference in left-right symmetry of glottis, but anterior-posterior symmetry of glottis. That was, the mean of anterior glottal length ratio index is significantly smaller than the mean of posterior glottal length ratio index. However, there should be expected asymmetry between left and right vocal folds as a consequence of different vibrating masses in dysphonic group with pathological tissue changes, which is due to inhomogeneities of the collision forces during glottic closure (a kind of compensation by increased subglottic pressure). It is resulted in left and right desynchrony and so different fundamental frequencies on both folds, consequently turned out in different dysphonic voice qualities (e.g., roughness). However, the result does not show this horizontal asymmetry

significantly. It may be due to the interacting and complex vocal fold vibratory dynamics in a wide range of vocal pathologies. In other words, such expected left-right desynchrony may be compensated by other means (e.g., muscular tension in the whole vocal tract, amplitude differences, shift of glottal axis during closure etc.). For example, the frequency differences may transform into phase differences (e.g., the healthy fold vibrates faster and so reaching its maximum opening earlier than the pathological fold). The phase differences are, in turn, accompanied with axis shift during closure (e.g., the glottal axis moves from side of slower fold toward faster fold). Therefore, the resulting ratio value is a combination of a number of interacting factors. Further investigation in future research on this aspect may give a more detailed explanation.

The significant vertical asymmetry demonstrates an inhomogeneous distribution of the muscular tension over the vocal fold length. Consequently, vocal folds vibrate with different frequencies between anterior and posterior regions. This irregularity results in unfavorable resonance with the vocal tract and produces a dysphonic voice quality. Results indicate that the anterior portion of the vocal fold is abducting while the posterior portion is adducting during the opening phase. Another indication is imperfect adduction of the posterior glottis.

However, glottal adduction is sometimes accounted as a compound variable of glottal resistance, subglottal pressure, and the ratio between average glottal flow and subglottal pressure (Sundberg, 1995). The eight glottal ratio indices mentioned in this high-speed analysis reflect vocal fold vibratory disturbances in one or more of the parameters stated above. Drawing conclusions of underlying vocal fold physiology from only glottal adduction may be too general. Therefore,

common clinical practice of voice evaluation adopted a multiparametric approach. That is, perceptual rating of voice quality together with instrumental measures (e.g., high-speed digital processing, acoustic analysis, electromyography etc.). It provides both subjective and objective ratings of voice quality to improve the accuracy.

## **CONCLUSION**

A new quantitative analysis system on high-speed laryngoscopic images of vocal fold vibration is established. A multidimensional study of phonatory mechanism (e.g., horizontal and vertical symmetries) can be achieved, which is found to be distinctive contributions of dysphonia diagnosis. It is applicable to study aperiodic phonation as well. Tentative conclusion is made about statistically significant difference of eight glottal ratio indices between dysphonic group and normal group. The result suggests that five ratio indices are representative in differentiating people with dysphonia from normal population.

At present, high-speed recordings are made in black and white. Innovations on improving image resolution and colour imaging in future may give an even detailed examination of phonatory mechanisms.

## REFERENCES

- Colton, R., & Casper, J. (1996). *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. New York: Williams & Wilkins.
- Deliyski, D. (2005). Endoscope Motion Compensation for Laryngeal High-Speed Videoendoscopy. *Journal of Voice*, *19*(3), 485-96.
- Fong, R. (submitted). *High speed laryngoscopic study of coal fold vibratory patterns in normal and dysphonic subjects*.
- Granqvist, S., Hertegard, S., Larsson, H., & Sundberg, J. (2003). Simultaneous Analysis of Vocal Fold Vibration and Transglottal Airflow; A New Experimental Set-up. *Journal of Voice*, *17*(1), 319-30.
- Hertegard, S, Larsson, H., & Wittenberg, T. (2003). High-Speed Imaging: Applications and Development. *Logopedics, Phoniatrics, Vocology*, *28*(1), 133-39.
- Hirano, M., & Diane, M. (1993). *Videostroboscopic Examination of the Larynx*. San Diego, California: Singular Publishing Group, Inc.
- Hollien, H. (1974). On vocal registers. *Journal of Phoniatrics*, *2*(2), 25-43.
- Hong, K.H., Kim, H.K., & Niimi, S. (2002). Laryngeal Gestures during Stop Production using High-speed Digital Images. *Journal of Voice*, *16*(1), 207-14.
- Hoppe, U., Rosanowski, F, & Dollinger, M. (2003). Glissando: Laryngeal Motorics and Acoustics. *Journal of Voice*, *17*(1), 370-76.
- Koster, O., Marx, B., Gemmar, P., Hess, M.M., & Kunzel, H.J. (1999). Qualitative and quantitative analysis of voice onset by means of a multidimensional voice analysis system (MVAS) using high-speed imaging. *Journal of Voice*, *13*(3). 355-374.

- Lindestad, P., Sodersten, M., Merker, B., & Granqvist, S. (2001). Voice source characteristics in Mongolian “throat singing” studied with high-speed imaging technique, acoustic spectra, and inverse filtering. *Journal of Voice*, 15(1), 78-85.
- Schuberth, S. (2002). High-precision Measurement of the Vocal Fold Length and Vibratory Amplitudes. *Laryngoscope*, 112(17), 1043-49.
- Sundberg, J. (1995). Vocal fold vibration patterns and modes of phonation. *Folia Phoniatica et Logopedica*, 47(4), 218-228.