

(Published by International Journal of Speech Language Pathology 2010, v.
12(6) p. 520-528)

**A preliminary study of a quantitative analysis method for high speed
laryngoscopic images**

Edwin M-L YIU^{1,2}

Jiangping KONG^{3,4}

Raymond FONG¹

&

Karen M.K.CHAN¹

¹Voice Research Laboratory
Division of Speech and Hearing Sciences
The University of Hong Kong

²Discipline of Speech Pathology
Faculty of Health Sciences
The University of Sydney

³Department of Chinese Language and Literature
Peking University

⁴The Hong Kong Putonghua Education & Assessment Centre
Faculty of Education
The University of Hong Kong

Keywords: voice evaluation, high speed imaging; vocal fold vibration, automatic image signal processing,

Running Head: High speed laryngoscopic study

Corresponding Author: Prof Edwin Yiu (PhD) Voice Research Laboratory, Division of Speech & Hearing Sciences, 5/F Prince Philip Dental Hospital, University of Hong Kong.

Tel: + 852 28590599 Fax: + 852 25590060 Email: eyiu@hku.hk

Abstract**A preliminary study of a quantitative analysis method for high speed laryngoscopic images**

High speed laryngoscopic study provides new information on the vocal fold vibratory patterns which has not been made possible before. Qualitative analysis of high speed laryngoscopic images is commonly reported in the literature. However, quantitative studies are necessary to provide objective measures for research and clinical purposes. This study reports the development of the High Speed Video Processing Program in analyzing high-speed laryngoscopic video images obtained from 15 non-dysphonic speakers. Nine glottal ratio indices that represented the full opened glottal area, glottal width and glottal length were computed using the processing program. Data from two dysphonic participants were included to illustrate the manifestation of these glottal measures in dysphonic voices. Some of these glottal ratio index measures showed high variability across participants. However, significant differences were found in a number of index measures among the creaky, modal and falsetto registers in the non-dysphonic participants. In addition, the glottal area ratio index was found to be specifically sensitive in detecting dysphonia in high speed glottal images.

The proposed semi-automatic High Speed Video Processing Program showed value in the analysis of high speed laryngoscopic images.

A preliminary study of a quantitative analysis method for high speed laryngoscopic images

Visual examination of vocal fold movements is one of the means of assessing dysphonia. Video endostroboscopy is a commonly used clinical tool in studying the vibratory pattern of vocal fold. Studies have found that vibratory pattern (symmetry and amplitude); mucosal wave, vocal fold edge roughness, and glottal closure configuration obtained from stroboscopic images are useful for identifying abnormalities (Colton & Casper, 1996; Colton, Woo, Brewer, Griffin, & Casper, 1995; Hirano & Bless, 1993; Sataloff, Spiegel, & Hawkshaw, 1991). Interpretation of stroboscopy, however, can be difficult when the vocal fold vibration is irregular or aperiodic because stroboscopic images are indeed virtual images made up of frames combined from different phases of vibratory cycles (Eysholdt, Tigges, Wittenberg, & Proschel, 1996). Therefore, some vocal fold movements within each vibratory cycle are not captured by the image recording process and there is no way to determine how this missing information contribute to the physiology of the actual vocal fold vibration. In other words, stroboscopic method does not capture the entire vibratory patterns fully and some information will be lost during the stroboscopic recording. Since pathological voice is often associated with irregular vocal fold vibration (Eysholdt et al., 1996), this limitation of stroboscopy in investigating pathological vocal fold vibratory patterns should be taken into consideration.

The development of high speed laryngoscopy has overcome this specific limitation of stroboscopy in investigating pathological voice production (Hirose, 1988). Current digital imaging technology can process speed as high as 8000 frames per second. High speed laryngoscopy reveals vocal fold vibratory patterns which were not previously available in stroboscopy. Its first use in the study of vocal fold vibration in normal and dysphonic voice

production began the 1980s. One of the early studies of high speed laryngoscopy reported the usefulness of the method in identifying incomplete glottal closure and asymmetrical vocal fold vibratory patterns (Hirose, 1988). In another study, Koster, Marx, Gemmar, Hess and Kunzel (1999) analyzed high speed images with synchronized electroglottographic and acoustic signals. They examined the glottal area and widths under different voice onset conditions. No specific vibratory pattern was found as the movements were highly variability within and between speakers (Koster et al., 1999).

Objective measurement has been developed in analyzing high speed images using quantitative analysis. For example, Maurer, Hess and Gross (1996) investigated glottal opening during vowel production. They determined the amount of opened glottis within a vibratory (opening-opened-closing-closed) cycle. The ratio of the number of frames with an opened glottis to the number of frames within a full vibratory cycle was termed open-quotient by the investigators. They found different vowels showing distinct patterns of open-quotients. It should be noted that the open-quotients derived from the high speed images showed no simple correlations with the open-quotients derived from electroglottographic data (Maurer et al., 1996). Other more sophisticated quantitative measurements of high speed laryngoscopic images are also available. Example are kymography (Svec & Schutte, 1996), phonovibrography (Lohscheller & Eysholdt, 2008; Lohscheller, Eysholdt, Toy, & Dollinger, 2008) and glottal area waveform (Yan, Ahmad, Kunduk, & Bless, 2005; Yan, Chen, & Bless, 2006). However, these analyses use some kind of transformation analyses which make the interpretation less intuitive when compared to those that examined simply the glottal measurements (Koster et al., 1999; Maurer et al., 1996). As it is anticipated that high speed laryngoscopy will eventually become clinically affordable, an intuitive quantitative measurement protocol that inform investigators or clinicians the glottal configuration and movement will be needed.

This paper proposes a semi-automatic analysis method to quantify the glottal measurements in high speed images. These quantitative measures are expected to be intuitive in nature. That is, they will provide objective measures that substantiate visual examination of the vocal fold movement in the high speed images. The analysis method was used in investigating modal, falsetto and creaky register productions. These three registers cover the low, mid and high frequency of the human voice (Colton & Casper, 1996) and represent three perceptually salient vocal qualities (Laver, 1991). The study of modal phonation, which is used by the majority of people (Laver, 1991), and non-modal phonations (creaky and falsetto registers) in healthy speakers provides information on how vocal folds vibrate differently under different voicing conditions when compared to dysphonic speakers.

The objective of the study was to determine if a set of quantitative glottal measurements (glottal area, length and width) were sensitive enough to identify the different vocal register productions (modal, falsetto and creaky) in high speed laryngoscopic recordings. These quantitative measures will be useful in studying abnormal vocal fold vibration.

Method

Participants

Fifteen non-dysphonic participants (two males and 13 females, mean age=21.3 years, standard deviation=1.2 years, range=19-23 years) and two female dysphonic participants (both age=21 years) took part in this study. All participants were non-smokers, non-drinkers, and were all university undergraduate students. All participants were able to produce modal, falsetto and creaky registers using the vowel /i/ following the demonstrations by the third author (RF). The non-dysphonic participants had neither history of voice disorder nor vocal pathologies which was confirmed with rigid endo-laryngoscopy carried out prior to the high

speed laryngoscopy. One dysphonic participant was diagnosed by a laryngologist to have unilateral vocal polyp and the other had a unilateral vocal cyst.

Equipment

Kay's High Speed Video System (Model 7000) with a 70 degree rigid laryngoscope (KAY 9106) was used to capture laryngoscopic images at 2000 frames per second.

Synchronized audio signals were recorded simultaneously using a microphone (Sanken COS 11BP) attached to the camera head which was placed approximately 5 cm from the participant's lips during the laryngoscopic procedure. Syntrillium's Cool Edit version 2000 (Syntrillium, 2000) was used to digitize and store the audio signals.

Procedures

Prior to the final recording, each participant was asked to practice producing modal, falsetto and creaky phonations using the vowel /i/ with the tongue protruded to simulate the rigid endoscopic procedure. Once the participant's ability to produce the three registers was confirmed and was consistent as determined by the third author (RF), laryngoscopic procedure using a rigid endoscope was carried out to record the /i/ phonations in modal register, followed by falsetto and creaky registers. Samples were saved for the final analyses only if the phonation types produced by the participants were considered by all four authors (three speech pathologists and one phonetician) during the endoscopic procedure to be appropriate. All 15 participants were able to allow clear image recordings of modal and falsetto /i/ register phonations but only seven participants produced clear recordings of creaky /i/ register phonations.

Each recording covered a duration of 2.184 seconds of video images in AVI format with a resolution of 120 x 256 pixels, synchronized with a stereo audio signal, and stored in

WAV format at a sampling rate of 10000 Hz and 16 bit. For the phonation, loudness level was monitored across all participants by asking each of them to produce the phonation at a comfortable loudness level with the Cool Edit Program recording level display to be maintained between -4 to 0 dB. A total of 4183 video frames were available for analysis for each video recording. The video and audio signals were then processed by a digital signal processing software – High Speed Video Processing (HSVP) program (beta version) - designed by the first (EY) and second (JK) authors. The program was written in MATLAB 6 (MathWorks, 2001) by the second author (JK).

Data analysis

Stage 1. Raw image processing - Each video recording file consisted of 4183 frames covering a span of 2.184 seconds. The third author (RF) inspected 15 video recording files and found that extraction of approximately 0.5 seconds (1000 frames) in each image file was generally practical with an unobscured view of the anterior tip, posterior tip and the lateral edges of the vocal folds. Among these 1000 frames, 100 vibratory cycles were selected for analysis. If there were less than 100 cycles among these 1000 frames, which could possibly happen in some male modal and creaky production as a result of low fundamental frequency, 1500 frames were selected. The following steps were used to process the images and to select the frames that contained the 100 vibration cycles using the HSVP program:

1. The contrast and brightness of the entire image recordings were adjusted to allow the vocal folds and the surrounding anatomical landmarks to be visually identified.
2. From the 4183 frames, images that did not have the four landmarks (anterior tip, posterior tip and the lateral edges of the vocal folds) clearly visible were excluded.

3. Frames with the vocal folds that were more than 5mm away from the midline of the frame were excluded. This was to ensure that the vocal folds and the surrounding structures were all within the visual field.
4. The middle 1000 frames of the remaining frames after conducting steps 1 to 3 above were segmented from the original recording file. These 1000 frames were further examined to extract the middle 100 cycles which contained 100 frames of complete glottal closure and 100 frames of maximum glottal opening.
5. The images were rotated, if necessary, so that the longitudinal axis of the vocal folds is aligned with the vertical axis (see Figure 1).
6. An Analysis Window was placed on the image to delineate the area surrounding the glottis (see Figure 1b). The edges of the analysis window were placed on four points on the glottis: anterior tip, posterior tip, the left lateral edge, and the right lateral edge.

Stage 2. Quantitative image analysis – A number of measurements were carried out using the Analysis Window of the HSVP program. The Analysis Window divided the glottis into four quadrants using the horizontal midline and longitudinal midline (see Figure 2). Calculations were done for:

1. The area of the glottis (measured in pixel-square) – total glottal area (GA), left (L-GA) and right glottal area (R-GA). The L-GA and R-GA were the areas to the left and right side of the longitudinal midline inside the Analysis Window respectively (see Figure 2).
2. The longitudinal and horizontal dimensions of the glottis, which included anterior length (A-Len), posterior length (P-Len), left width (L-Width) and right width (R-Width) of the glottis. The A-Len and P-Len of the glottis refer to the longitudinal dimension of the glottis measured from the mid point of the Analysis Window to the anterior and posterior ends of the glottis respectively along the longitudinal midline. The L-Width and R-Width

of the glottis refer to the width of the glottis measured from mid point of the Analysis Window to the left and right lateral margin of the glottis along the horizontal midline (see Figure 2). The unit for these four measurements was pixels.

After the Analysis Window was placed on the first image frame which contained a fully opened glottis, the program automatically adjusted the brightness and contrast, traced the position of the glottis in each subsequent image frame, and performed the measurements automatically. The program was able to trace the glottis even if it changed position across image frames. Details of the automatic algorithms are beyond the scope of this paper. This can be found, however, in a paper by Kong and Yiu (2010). In brief, the program first expanded the brightness of the images to cover the full grey scale range (0-255); then rotated the images to ensure the vertical axis of the glottis is aligned with the centre of the frame by using the bi-cubic algorithm (Kong & Yiu, 2010); compensated for the motion of the recording camera by first tracking all the centers of the glottal images (Deliyski, 2005); binarized the pixel of the glottis into 0 or 1 so that the area of the glottis was calculated based on the black (1) pixel; and finally extracted the parameters as described below.

Only measurements from the frames with the minimum glottal opening and maximum glottal opening were used to obtain the ratio indices to be described below to give an indication of how large the glottis was opened at the maximum opening phase compared to the minimum opening (or closed) phase. The use of ratio indices was necessary as absolute measurements of the glottal area and dimension were not possible due to unknown magnification factor.

The Ratio Index for each measurement was calculated as follows:

$$\frac{\text{Mean measurement from the 100 minimum glottal opening frames}}{\text{Mean measurement from the 100 maximum glottal opening frames}}$$

In theory, the image frames containing the closed glottis should contain zero glottal area or length. However, in practice, an image with a closed glottis would still contain residual digital noise, i.e. hot pixel, that contributes to some positive pixel value. Therefore the pixels in the closed glottis were not equal to zero although they should be approaching zero. Therefore, as the measurement of the minimum glottal opening became large, a smaller index was obtained. Thus, a larger glottal opening in each vibratory cycle would be represented by a small ratio index.

Nine ratio indices were calculated for each video image set:

1. Full glottal area ratio index: This index reflected the glottal area of the maximum opening frames in relation to the corresponding minimum opening frames.
2. Left and right glottal area ratio indices: These two indices represented the left and right halves of the full glottal area measured from the longitudinal midline to the left and right lateral edges of the glottis respectively. Asymmetry of the vocal folds vibration can be inferred from the difference between the left and right indices.
3. Full glottal opening width ratio index: This index represented the width of the glottis of the maximum opening frames measured along the horizontal midline of the Analysis Window in relation to the corresponding minimum opening frames.
4. Left and right glottal width ratio indices: This index represented the left and right halves of the full glottal width measured from the longitudinal midline to the left and right lateral edges of the glottis respectively. Asymmetry of the vocal folds vibration can be inferred from difference between the left and right width indices.
5. Full antero-posterior glottal length ratio index: This index represented the full length of the glottis of the maximum opening frames measured along the longitudinal midline of the Analysis Window in relation to the corresponding minimum opening frames.

6. Anterior-half and posterior-half glottal length ratio indices: These two indices represented the anterior half and posterior half of the full glottal length measured along the longitudinal midline from the midpoint of the Analysis Window. Difference between the anterior-half and posterior-half glottal length indices suggests asymmetry between the anterior and posterior parts of the vocal folds during the maximum glottal opening.

Agreement, inter/intra-rater reliability

The calculation of the ratio indices depended on the reliability in extracting analyzable frames and placing the Analysis Window using the HSVP program. Therefore, 15% (8) of the video images were re-analyzed two weeks later after the first analysis to obtain intra-rater reliability. Inter-rater reliability was also determined by re-analyzing 15% of the video images by another independent rater. For the analyzable frame selection, the agreement was calculated by dividing the number of identical frames selected on both occasions (intra-rater) or by both raters (inter-rater) over the total number of frames for analysis (200, as described above in step 4 during Stage 1 raw image processing). For the reliability in placing the Analysis Window, the indices obtained within and between raters were examined using intra-class correlations (ICC) with a two-way mixed model and Bland-Altman analysis (Bland & Altman, 1986) by plotting the differences against the average of the indices.

Results

Agreement and reliability measures

For the selection of frames, intra-rater exact agreement was found to be high at 0.87, while inter-rater exact agreement was moderate (0.62). For the intra-rater reliability in placing the Analysis Window, ICC was high (0.82, $F=10.116$, $p<0.0001$). The Bland-Altman analysis

showed that more than 95% of the measures were within the mean \pm 2 standard deviations limits with two outliers (see Figure 3). Narrow distribution was shown for measures below 0.60 and a wider spread from 0.60 to 1.00 (see Figure 3). The ICC for the inter-rater agreement was moderate (0.70, $F=5.68$, $p<0.0001$) but the Bland-Altman analysis also showed more than 95% of the measures were within the mean \pm 2 standard deviations limits with three outliers (see Figure 4).

Put Figure 3 and 4 here

Glottal ratio indices of normal participants

Table 1 lists the mean glottal ratio indices of the three vocal register productions. The modal register demonstrated relatively smaller values while the creaky register showed relatively larger values. This indicated that the modal register produced relatively larger glottal area with larger and wider measurements. The creaky register produced the smallest measurements, and the falsetto register fell in between the modal and creaky registers.

Non-parametric Wilcoxon signed rank tests were used and showed that the modal register demonstrated a significantly smaller anterior-half glottal length ratio index when compared to the falsetto register. No significant differences were found between modal and falsetto register in all other indices. All nine ratio indices of the creaky register demonstrated significantly larger values than those of the modal register (see Table 1). Six of the ratio indices of the creaky register also demonstrated significantly larger values than those of the falsetto register (see also Table 1).

The modal register showed a significant difference between the left and right glottal width ratio indices (See Table 2). No significant difference, however, was found between the left and right glottal width ratio indices in the falsetto and creaky registers, nor in the glottal

area ratio indices in all three registers. No significant difference was found in the anterior-posterior symmetry of the glottis in all three vocal registers (See Table 2).

Put Tables 1 & 2 here

Comparison of the dysphonic participants with the normal participant group

The sample size of the dysphonic participants was small (N=2). Therefore, the data were compared with the 95% confidence interval level of the nine glottal ratio indices obtained from the normal participants (Table 3). It should be noted that three indices were unanalyzable for some of the samples due to the system limitation. For the samples with analyzable indices, 12 out of 21 indices were outside the 95% confidence interval of the normal participants for Dysphonic Participant 1 and 12/26 for Dysphonic Participant 2 (see Table 4).

Put Tables 3 & 4 here

Discussion

The aim of the study was to investigate the usefulness of a quantitative method for measuring high speed laryngoscopic images. Glottal ratio indices were calculated to describe the three registers in participants with normal voices. The indices were able to distinguish creaky register from modal and falsetto registers.

Comparison among the three vocal registers

Falsetto voice production is often described as demonstrating medial compression of the vocal folds which results in a glottal appearance that is thin in the antero-posterior

dimension when compared to the modal production (Laver, 1991). Interestingly, the indices in the present study were, however, not able to discriminate between modal and falsetto productions (see Table 1). Indeed, the significant anterior-half glottal length ratio index demonstrated that the modal register showed larger (smaller index) dimension than the falsetto register. The glottal ratio indices demonstrated by the participants in the present study indeed showed relatively high variability (as indicated by the standard deviation) in both the modal and falsetto production (see Table 1). One possible reason could be attributed to the inadequate speed of the Kay's High Speed Video System (Model 7000) in capturing enough frames of falsetto phonation recordings. In using a speed of 2000 fps to assess falsetto phonation, which could be up to 400-500 Hz, only four to five frames covered each cycle of falsetto vibration. Therefore, the information used in calculating the indices would be limited. Unfortunately, that was the limitation of the available high speed system at the time of the study. More recent commercial systems can provide recording speed of up to 4000 fps for color images and 8000 fps for monochrome images.

In creaky production, the vocal folds are often characterized by a strong medial compression accompanied with inward movement of the ventricular folds (Laver, 1991). Thus resulting in relatively smaller glottal area, glottal width and length dimensions compared to the falsetto (in six indices) and modal (in all nine indices) productions. It should also be noted that only seven participants had the creaky production recorded successfully. The difficulty in recording a visible glottis in creaky production was due to the lowered position of the larynx during creaky phonation, and also the supraglottic constriction. The supraglottic constriction blocked the view of the glottis in some participants. Therefore, the apparent "glottal" measurements obtained in creaky phonation might reflect some form of ventricular involvement and not the glottis *per se*. This is indeed quite a limitation of the endoscopic

procedure in studying creaky phonation. Therefore, it is necessary to compensate for this missing data by recruiting more participants if possible in future studies.

Table 1 shows that modal register demonstrated relatively smaller values while creaky register had relatively larger values. The sample size could have contributed to the lack of significant difference between the modal and falsetto registers. However, the creaky register clearly showed significant larger values. Therefore, this quantitative procedure indeed demonstrates a potential in differentiating the three registers.

Glottal symmetry

In normal vocal fold vibratory patterns, the measurements for left and right halves of the glottis should not be different. Any difference in the left and right indices suggested either an asymmetry between the left the right vocal fold vibratory movements or a tilting of the tip of the endoscope which results in image distortion. None of the left and right ratio indices obtained from the normal participants showed any significant difference in each of the three registers. This indicated that this procedure was adequate to capture the symmetry of the vocal fold vibration.

It should also be noted that the sample included more females (N=13) than males (N=2). There are anatomical differences between male and female larynges and therefore, whether these glottal area ratios differ between genders is an unanswered question. Further studies with larger sample size that explore possible differences between male and female speakers would be needed to answer this question.

Use of the glottal measurement indices in dysphonic participants

More than half of the analyzable measurement indices (24/47: 12/21 for Dysphonic Participant 1 and 12/26 for Dysphonic Participant 2; see Table 4) demonstrated by the two

dysphonic participants were outside the 95% confidence interval of normal participants. It should be noted that *all* the *abnormal* ratio indices obtained from Dysphonic Participant 1 were greater than the upper boundary of the 95% confidence interval demonstrated by the normal participants while all but one of the “abnormal” indices (the posterior glottal length ratio index) obtained from Dysphonic Participant 2 were below the lower boundary of the 95% confidence interval of the normal participants. This indicates that dysphonic participants do not necessarily demonstrate definite larger or smaller index values when compared to the normal participants. The larger index values shown by Dysphonic Participant 1 suggested that the maximum glottal opening was relatively small. The smaller glottal opening might have been caused by a higher medial compression of the vocal folds. The high medial compression may increase aperiodic vibration which subsequently resulted in perceptual roughness. Whether a small glottal opening will cause aperiodic vibration or perceptual roughness would require further investigation. The relatively smaller indices demonstrated by Dysphonic Participant 2 indicated possible incomplete glottal closure at the closed phase, i.e. an incomplete closed adduction. Incomplete adduction of the vocal folds during the closed phase would lead to perceptual correlate of breathiness. This would of course require further validation with the inclusion of more dysphonic participants.

High speed laryngoscopy is showing potential clinical benefit in which vocal fold vibratory patterns could be revealed more precisely without the limitations of stroboscopy (Lindestad, Sodersten, Merker, & Granqvist, 2001; Verdonck-de Leeuw, Festen, & Mahieu, 2001). Quantitative analysis conducted by another independent study also showed that the vibratory patterns are variable (Koster et al., 1999). The preliminary findings from the two dysphonic participants concurred with that of Koster et al (1999) and showed that the abnormal vocal fold vibratory patterns could be variable. Nevertheless, the present study showed that the HSVP measurements were useful in revealing abnormalities when

compared to the normal participants. The automatic computation feature of the HSVP is a useful function which facilitates a more accurate analysis.

The findings from the present study serve as a preliminary effort to quantify the glottal opening in high speed images. More participants are, however, needed to form a larger database so that the unique nature of the different register productions could be better investigated.

The sample size included in the present study was biased in gender. The majority of the participants were female and only two males were recruited. Further study will be necessary to determine if gender difference exists in the normal and dysphonic participants.

The intra-rater reliability of the analysis procedure was relatively high ($ICC=0.82$) but only moderate ($ICC=0.70$) for the inter-rater reliability. The analysis procedure involved subjective selection of 1000 stable frames, then extraction of the middle 200 frames, and the precision of adding the Analysis Window manually. Given the 1000 frames involved only one second, the selection of a different portion of one second (i.e. different 1000 frames) would have little effect on the outcome of the measurements unless the vocal fold vibratory pattern changed drastically within the two-second recording duration.

The placement of the Analysis Window was more critical for the analysis procedure. Subjective judgment was required to locate the anterior, posterior and later edges of the glottis. A zooming function to enlarge the image would have helped to allow placement of the window more precisely. The HSVP program is still under further technical refinement and we are still in a process of refining the algorithms to improve the validity and reliability of the process.

References

- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *8*, 307–310.
- Colton, R. H., & Casper, J. (1996). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. New York: Williams & Wilkins.
- Colton, R. H., Woo, P., Brewer, D. W., Griffin, B., & Casper, J. (1995). Stroboscopic signs associated with benign lesions of the vocal folds. *Journal of Voice*, *9*(3), 312-325.
- Deliyski, D. D. (2005). Endoscope motion compensation for laryngeal high-speed videoendoscopy. *Journal of Voice*, *19*(3), 485-496.
- Eysholdt, U., Tigges, M., Wittenberg, T., & Proschel, U. (1996). Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia Phoniatica et Logopaedica*, *48*(4), 163-170.
- Hirano, M., & Bless, D. M. (1993). *Videostroboscopic evaluation of the larynx*. San Diego, CA: Singular Publishing.
- Hirose, H. (1988). High-speed digital imaging of vocal fold vibration. *Acta Oto-Laryngologica - Supplement*, *458*, 151-153.
- Kong, J. P., & Yiu, E., M.-L. (2010). Quantitative high speed video laryngoscopy. In E. Ma, P.-M. & E. Yiu, M.-L. (Eds.), *Handbook of clinical voice assessment*. San Diego, CA: Plural Publisher.
- 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.
- Laver, J. (1991). *The gift of speech. Papers in the analysis of speech and voice*. Edinburgh: Edinburgh University Press.
- Lindestad, P. A., 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.
- Lohscheller, J., & Eysholdt, U. (2008). Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope*, *118*(4), 753-758.
- Lohscheller, J., Eysholdt, U., Toy, H., & Dollinger, M. (2008). Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics. *IEEE Transactions on Medical Imaging*,

27(3), 300-309.

MathWorks. (2001). Matlab (Version 6). Natick, MA: MathWorks Inc.

Maurer, D., Hess, M., & Gross, M. (1996). High-speed imaging of vocal fold vibrations and larynx movements within vocalizations of different vowels. *Annals of Otolology, Rhinology and Laryngology*, 105(12), 975-981.

Sataloff, R. T., Spiegel, J. R., & Hawkshaw, M. J. (1991). Stroboscovideolaryngoscopy: Results and clinical value. *Annals of Otolology, Rhinology and Laryngology*, 100(9 Pt 1), 725-727.

Svec, J. G., & Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10(2), 201-205.

Syntrillium. (2000). Cool Edit (Version 2000). Pheonix, AZ: Syntrillium.

Verdonck-de Leeuw, I. M., Festen, J. M., & Mahieu, H. F. (2001). Deviant vocal fold vibration as observed during videokymography: The effect on voice quality. *Journal of Voice*, 15(3), 313-322.

Yan, Y., Ahmad, K., Kunduk, M., & Bless, D. (2005). Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology. *Journal of Voice*, 19(2), 161-175.

Yan, Y., Chen, X., & Bless, D. (2006). Automatic tracing of vocal-fold motion from high-speed digital images. *IEEE Transactions on Biomedical Engineering*, 53(7), 1394-1400.

Acknowledgements

This study was funded in parts by a grant from the University of Hong Kong (Seed Grant 2005-06) and the Faculty of Education Research Fund (2004-2005) awarded to the first author. The authors would like to thank Elaine Kwong and Cecilia Fung for their assistance in participant recruitment. The participation of the volunteers are gratefully acknowledged. The first author would also like to acknowledge that the major revision of this manuscript was written while taking his sabbatical at The University of Sydney, from which new ideas and stimulating discussions with colleagues have given new thoughts to this manuscript.

Figure legends

Figure 1. High Speed Video Processing analysis program page illustrating image rotation and placement of the Analysis Window

Figure 2. Analysis window measurements

Figure 3. Bland and Altman Analysis of intra-rater agreement

Figure 4. Bland and Altman Analysis of inter-rater agreement

Table 1. Mean, standard deviation (SD) and Wilcoxon signed ranked tests for the nine glottal ratio indices of normal participants

	Wilcoxon signed rank test								
	Modal (n=15)	Falsetto (n=15)	Creaky (n=7)	Modal Versus Falsetto (n=15)		Modal Versus Creaky (n=7)		Falsetto Versus Creaky (n=7)	
	Mean (SD)	Mean (SD)	Mean (SD)	Z	p	Z	p	Z	p
Full glottal area ratio index	0.27 (0.15)	0.36 (0.16)	0.57 (0.21)	-1.42	0.16	-2.37	0.02*	-2.20	0.03*
Left glottal area ratio index	0.27 (0.15)	0.36 (0.17)	0.57 (0.24)	-1.31	0.19	-2.37	0.02*	-2.20	0.03*
Right glottal area ratio index	0.27 (0.18)	0.35 (0.18)	0.60 (0.20)	-1.59	0.11	-2.37	0.02*	-2.37	0.02*
Full glottal width ratio index	0.27 (0.26)	0.33 (0.20)	0.64 (0.36)	-1.08	0.28	-2.20	0.03*	-2.03	0.04*
Left glottal width ratio index	0.30 (0.27)	0.31 (0.23)	0.44 (0.45)	-0.40	0.69	-1.35	0.18*	-1.01	0.31
Right glottal width ratio index	0.25 (0.31)	0.32 (0.22)	0.86 (0.16)	-1.31	0.19	-2.20	0.03*	-2.20	0.03*
Full antero-posterior glottal length ratio index	0.36 (0.16)	0.41 (0.16)	0.63 (0.23)	-1.82	0.07	-2.37	0.02*	-1.86	0.06
Anterior-half glottal length ratio index	0.33 (0.16)	0.50 (0.22)	0.73 (0.24)	-1.99	0.05*	-2.37	0.02*	-2.20	0.03*
Posterior-half glottal length ratio index	0.40 (0.30)	0.44 (0.24)	0.59 (0.28)	-0.63	0.53	-2.03	0.04*	-1.86	0.06

*p<0.05 (2-tailed)

Table 2. Wilcoxon signed rank test of left-right and antero-posterior indices

	Modal		Falsetto		Creaky	
	Z	P	Z	P	Z	P
Area (Left-right)	-0.11	0.91	-0.62	0.53	-0.51	0.61
Width (Left-right)	-2.27	0.02*	-0.47	0.64	-1.57	0.12
Length (Antero-posterior)	-0.85	0.39	-0.74	0.46	-1.18	0.24

*p<0.05 (2-tailed)

Table 3. 95% confidence interval level of the nine glottal ratio indices from normal participants

	Modal		Falsetto		Creaky	
	Lower Boundary	Upper Boundary	Lower Boundary	Upper Boundary	Lower Boundary	Upper Boundary
Full glottal area ratio index	0.12	0.41	0.17	0.34	0.33	0.80
Left glottal area ratio index	0.15	0.36	0.17	0.31	0.29	0.85
Right glottal area ratio index	0.06	0.48	0.11	0.45	0.37	0.82
Full glottal width ratio index	0.01	0.31	0.05	0.36	0.21	1.01
Left glottal width ratio index	0.07	0.36	-0.01	0.33	0.04	0.99
Right glottal width ratio index	-0.16	0.58	0.02	0.39	0.70	1.03
Full antero-posterior glottal length ratio index	0.13	0.53	0.24	0.47	0.37	0.90
Anterior-half glottal length ratio index	0.17	0.54	0.23	0.73	0.44	0.98
Posterior-half glottal length ratio index	0.05	0.60	0.14	0.33	0.33	0.93

Table 4. Glottal ratio indices demonstrated by the two dysphonic participants

Measurement Indices	Dysphonic Participant 1			Dysphonic Participant 2		
	Modal	Falsetto	Creaky	Modal	Falsetto	Creaky
Full glottal opening area ratio index	0.42*	0.35*	0.91*	0.16	0.17	0.25**
Left glottal opening area ratio index	0.37*	0.44*	0.86*	0.16	0.10**	0.24**
Right glottal opening area ratio index	0.51*	0.25	0.92*	0.16	0.26	0.27**
Full glottal opening width ratio index	N/A	N/A	0.98	0.04	0.02**	0.01**
Left glottal opening width ratio index	N/A	N/A	1.05*	0.04	0.01	0.02**
Right glottal opening width ratio index	N/A	N/A	0.96	0.04	0.03	N/A
Full antero-posterior glottal opening length ratio index	0.64*	0.30	0.91*	0.33	0.39	0.21**
Anterior-half glottal opening length ratio index	0.49	0.37	0.95	0.37	0.18**	0.25**
Posterior-half glottal opening length ratio index	0.81*	0.26	0.88	0.26	0.59*	0.17**

*Exceed the upper boundary of the 95% confidence interval of normal participants

**Below the lower boundary of the 95% confidence interval of normal participants

N/A = The measurements could not be computed by the HSVP program