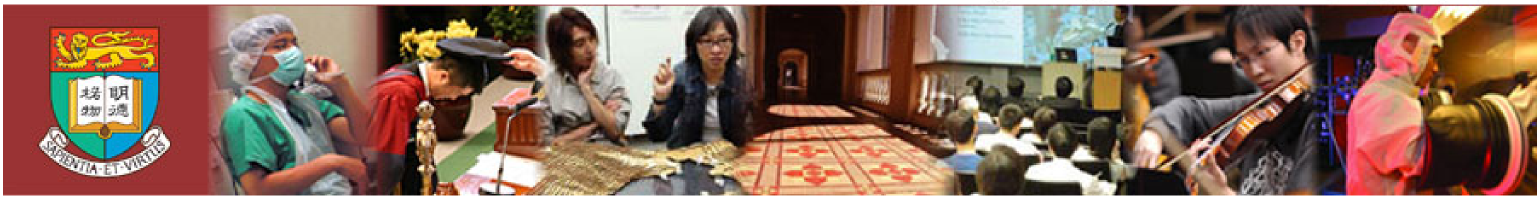


Title	High speed digital imaging the difference of vocal fold vibration between modal, falsetto, vocal fry registers and whisper
Other Contributor(s)	University of Hong Kong.
Author(s)	Wong, Nga-kei; 王雅姬
Citation	
Issued Date	2004
URL	http://hdl.handle.net/10722/48795
Rights	Creative Commons: Attribution 3.0 Hong Kong License

High Speed Digital Imaging: the Difference of Vocal Fold Vibration between Modal, Falsetto, Vocal Fry Registers and Whisper

WONG, Nga-Kei

Division of Speech and Hearing Sciences

The University of Hong Kong

University number: 2000121165

Email: arkiwong@yahoo.com.hk

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Sciences (Speech and Hearing Sciences), The University of Hong Kong, May 7, 2004.

Abstract

High speed digital imaging is the most advanced technology that is available in direct visualization and examination of vocal folds. The purpose of this study was to investigate the vocal fold vibratory patterns of modal, falsetto, vocal fry registers and whisper by using this advanced technique. Six female subjects with normal voices were recruited. Rigid endoscopic images of sustained vowel /i/ were recorded by Kay High Speed Video (HSV) System. Objective measurements of open quotient, frequency, glottal size index and vocal fold size index, and subjective judgement of symmetry, glottal closure and number of sudden openings were carried out. Results revealed significant difference in frequency, but not in other objective measurements across registers, as well as glottal size index and vocal fold size indices within a glottal cycle in modal, falsetto and whisper speaking voices, probably owing to the small sample size. The characteristics of each speaking voices are described qualitatively in this study.

High Speed Digital Imaging: the Difference of Vocal Fold Vibration between Modal, Falsetto, Vocal Fry Registers and Whisper

Different techniques have been utilized in vocal fold viewing and investigation in the literature. These included video stroboscopy (e.g. Li, 2000), high speed photography (e.g. Hollien, Girard & Loleman, 1977), kymography (e.g. Leeuw, Festen & Mahiew, 2001; Švec & Schutte, 1996) and high speed digital imaging (e.g. Eysholdt, Tigges, Wittenberg & Proschel, 1996; Hammarbery, 1995) techniques. High speed digital imaging is the most advanced technique available for visualization of larynx, which allows detailed vibratory movement analysis on vocal folds. The physiology and mechanism of different vocal registers is not fully explained in the literature. Current information of these mechanisms is based on investigations from endoscopic (e.g. strobscopy), aerodynamic (e.g. glottal flow rate), physiologic (e.g. electroglottography, EEG) or acoustic (e.g. fundamental frequency) measures (e.g., Li, 2000; Blomgren, Chen, Ng & Gilbert; 1998). Hence, the relationships between the vocal fold vibratory patterns and the indirect investigations of vocal behaviours, such as aerodynamic and physiologic measures, can be found by using high speed digital imaging techniques.

There are a number of typical vocal registers proposed in the literature. These include modal register, falsetto register, vocal fry register and whisper. They represent a wide range of frequencies, loudness and vocal qualities that can be produced by human.

Different Techniques used in Investigation of Vocal Fold Vibration Patterns

Video Stroboscopy. Video stroboscopy is currently the most common tool in viewing and investigating the vocal fold vibration patterns and supraglottic activities, which provides direct visualization and examination of vocal folds with strobe light (Niimi & Miyaji, 2000; Eysholdt et al., 1996; Švec & Schutte, 1996). This technique assumes that vocal fold vibration is periodic and the images accurately represent the exact vibratory behaviour. However, the images recorded are just virtual ones averaged from many glottal cycles. Thus, it is impossible to have interpretation on the exact cycle-to-cycle vocal fold vibrations and glottal behaviours (Leeuw et al., 2001; Poburka, 1999; Švec & Schutte, 1996; Kiritani, Hirose& Imagawa, 1993). Moreover, only periodic vibrations can be studied, whereas precise laryngeal behaviours in aperiodic and irregular vocal fold vibrations cannot be observed (Larsson, Hertegard, Lindestad & Hammarberg, 2000). Hence, a strobe light is not able to capture the high degree of irregularity of vocal fold vibrations in hoarse voices (Larsson et al., 2000; Tigges, Wittenberg, Mergell, & Eysholdt, 1999; Eysholdt et al., 1996), and incorrect judgement on vocal fold vibratory regularity is obtained.

High Speed Photography. High speed photography technique was first developed in 1940s for visualization of vocal fold vibratory behaviours. Images of vocal fold vibration with good resolution can be obtained and frame-by-frame analysis is possible (Švec & Schutte, 1996; Kiritani et al., 1993). However, this system was massive and expensive. The mechanical noise produced from the high-speed camera prevents simultaneous recording of speech signal and vocal fold vibration, and it is time consuming to have frame-to-frame analysis of the films (Švec & Schutte, 1996; Kiritani et al., 1993). These disadvantages limited its use in experiments and clinical practices.

Videokymography. Videokymography was developed in 1990s. It is used with high speed digital recording system. Each video frame of vocal folds is composed of horizontal lines (perpendicular to glottis). One horizontal line is selected from the whole image of larynx. The successive line images across time are then added together and presented. These line images register the vibratory patterns of a particular part of vocal folds in the selected time interval. Therefore, it permits observation of vocal fold vibratory patterns by comparing the vibration between left and right vocal folds (Leeuw et al., 2001; Larsson et al., 2000; Švec & Schutte, 1996). The limitation is that visualization of anterior-posterior modes of vibration is not allowed. Although it is solved by using multi-plane videokmography, the overall picture of vocal fold movement cannot be captured for examining asymmetry in glottal closure (Tigges et al., 1999).

High Speed Digital Imaging. High speed imaging is a new technology which enables the recording and study of the detailed vocal fold movements (frame rate up to 2000 frames/s) by using slow-motion play back and frame-by-frame analyzing tools (Eysholdt et al, 1996). Furthermore, it allows capturing and studying of the actual aperiodic vocal fold vibrations, as those in hoarse voice. Visualization of voicing start-up and stopping, aperiodic vibrations, very short voicing segments and spasms are allowed. Therefore, this technique provides an excellent tool for investigating the mechanisms of various pathological voice production (e.g. hoarseness and diplophonia), vegetative functions (e.g. coughing and whistling) and vocal registers (e.g. modal, falsetto and vocal fry registers).

A number of studies have attempted to determine the characteristics of hoarse voice production, diplophonia and production of different vowels by using high speed digital imaging and slow motion playback analysis (Niimi & Miyaji, 2000; Eysholdt et al., 1996; Kiritani et al., 1995; Sundberg, 1995). The limitation of high speed digital laryngoscopy is that the image size is reduced when frame rate increases because of technological limitation and it is solved by keeping the sampling frequency so that the image size will not be too small for analysis. The recordings are evaluated by visual examination as there is no standard quantitative analysis available yet (Eysholdt et al., 1996). Therefore, vocal fold vibration is analyzed qualitatively by using the associated image processing program. Nevertheless, there are other limitations. For example, vocal fold vibrations of a two-second duration is the maximum duration obtained, infrequent aperiodicities may not be captured within this short period and recordings of vocal fold vibrations in producing sentences or passages is impossible (Niimi & Miyaji, 2000; Eysholdt et al., 1996; Maurer, Hess & Gross, 1996; Sundberg, 1995; Kiritani et al.,1993).

In summary, high speed digital imaging and slow-motion frame-by-frame analysis are advanced techniques in viewing and analyzing the complex vocal fold vibrations during phonation. The current study took a preliminary look at the vocal fold vibratory patterns in different vocal registers using high speed digital imaging techniques.

Vocal Registers

According to Hollien (1974), vocal register is defined as the perceptual voice quality range that can be produced by the larynx. There are some typical vocal registers. These include modal register (normal speaking frequency range), falsetto or loft register (phonation produced at highest human vocal frequency range), vocal fry, glottal fry or pulse register (phonation produced at lowest human vocal frequency range) and whisper. A number of studies have attempted to define vocal registers by using perceptual, acoustic, physiologic and aerodynamic parameters (e.g. Li, 2000; Blomgren, Chen, Ng, & Gilbert, 1998). This will be discussed in the following sections.

Modal Register. Modal register refers to phonation with normal speaking frequency range (i.e. neutral voice), associated with a middle pitch and a wide range of loudness (Hollien, 1974; Zemlin, 1998). In modal register, the length and thickness of the vocal folds are in between those of vocal fry and falsetto phonation (Hollien, 1974). The vibratory cycle is characterized by a long open phase and short closed phase in EEG waveform (Blomberg et al., 1998).

Falsetto Register. Falsetto or loft register refers to the phonation produced at highest human vocal frequency range, associated with soft loudness, high pitch and some breathiness (Hollien, 1974; Zemlin, 1998). In falsetto voice, the vocal folds are lengthened and thin (Hollien, 1974) with vibration along tensed and bow-shaped borders (Zemlin, 1998). The vibratory cycle is characterized by a gradual open phase with a rapid or no closed phase (Hollien, 1974). Posterior chink configuration was revealed at glottal closure when the vocal folds are under extreme tension, resulting in the breathy quality perceived in falsetto voice.

Vocal Fry Register. Vocal fry or pulse register refers to the phonation produced at the lowest human vocal frequency range, associated with soft loudness, low pitch and rough quality (Hollien, 1974; Zemlin, 1998). In vocal fry register, vocal folds are short and thick during vibration (Hollien, 1974). A flaccid vibrating margin is resulted from the low tension of the vocal muscle, while strong medial compression and tensed lateral portion of vocal folds are noted. The glottal cycle is characterized by a long closed phase and then sharp open phase in rapid succession with closely approximated vocal folds (Zemlin, 1998). Doublet or triplet openings within a single glottal cycle were revealed in EEG study (Blomberg et al., 1998). It was proposed that such vibratory pattern was attributed to the thick vocal folds and false vocal fold compression.

Whisper. Solomon et al. (1989) as reported by Colton & Casper (1996) found that in whisper voices, vocal folds are vibrating with straight margin, parallel configuration, and a toeing in of the vocal processes. Two main glottal configurations were reported, the first was a straight and parallel vocal-fold edge with no adduction of vocal folds, and the second was some approximation or contact of vocal folds occurred, but just at a portion of vocal folds.

Objectives of Study

The objectives of this study was to make use of high speed digital imaging techniques to examine the vocal fold vibratory patterns in four types of speaking voices, including modal register, falsetto register, vocal fry register and whisper, and to develop a physiological definition for these different vocal registers using the analysis from the high speed images.

Method

Subjects

Six female subjects, with a mean age of 21.5 (range = 19 – 22) were selected from a pool of forty at Division of Speech and Hearing Sciences of the University of Hong Kong. Potential subjects were screened for their ability to tolerate the rigid endoscope without triggering of a gag reflex, by putting a tongue depressor into the buccal cavity and moving it towards the posterior pharyngeal wall. Only those who were able to pass this screening will be participated in a trial session by using a rigid endoscope, a week before data collection. In order to be included in the present study, speakers should be able to sustain a vowel for five seconds with the endoscope in place for clear laryngeal image recording. A total of ten subjects were selected based on the screening and four subjects were subsequently excluded from the study owing to their inability to tolerate the rigid endoscope in vowel prolongation.

All subjects were native speaker of Cantonese, non-smokers, non-professional voice users. They were able to phonate modal, falsetto, vocal fry and whisper speaking voices upon demonstration by the investigator. All of them had no vocal complaints or voice disorders, pre-existing chronic medical condition, as well as previous laryngeal or pharyngeal surgery.

Materials

Kay Model 9106 rigid 70 degree endoscope, was connected to Kay High-speed Video (HSV) system - Model 9700, to capture and record the vocal fold vibrations. A 20mm lens adapter was used to connect the rigid endoscope to the camera head. The light source was a 300-watt xenon lamp. In the present system, the resolution was selected at 256 (horizontal) x 120 (vertical) pixels with grey scale, at frame rate of 2000 frame/s. The total number of frames recorded was 4368 at frame rate of 2000 frames/s, while that number was 2183 at frame rate of 1000 frames/s, i.e. a duration of 2.18 seconds, which are the default numbers of frames recorded by the instrument.

Task Utterance and Categorization of Voice Types

The subjects were required to sustain vowel /i/ with the following voice types: modal, falsetto, vocal fry and whisper voices, so as to obtain a clear view of larynx. Sustained vowel /i/ was selected because of its elevated laryngeal position for a better view of vocal folds (Colton & Casper, 1996). Identification of different voice types was carried out by an expert in voice pathology and the investigator. Both are native Cantonese speakers with normal hearing. It took few minutes to save each recording and subjects were asked to practice the type of voice production that would be recorded next during this period. The next recording was started until the participant was able to phonate that particular type of voice.

Procedure

Each subject was explained with the objectives and procedures of the study. They gave consent to the investigator for using their information and data collected in this project. Then, the subjects practiced modal, falsetto, vocal fry and whisper voices before the recording. Demonstrations were given by the investigator to ensure that these different types of voices were phonated correctly by each subject. After the subjects were able to produce the target voice types, endoscopic procedure was conducted by using a rigid endoscope. The limitation of Kay HSV is that the scan area decreases with the increase of frame rate. Therefore, recordings were made at frame rate of both 1000 frames/s and 2000 frames/s as the whole length of vocal folds could not be captured at frame rate of 2000 frames/s in most registers. The images were then stored temporarily at 492 Mb/s into very high-speech video RAM. After that, it took four to eight minutes to save the video into computer memory as AVI files at frame rate of 1000 frames/s and 2000 frames/s respectively.

Rigid Endoscopic Procedure. The rigid endoscopic procedure was carried out by experienced examiner using Kay model 9106 rigid endoscope. Each subject was required to phonate with modal register first, as it was the easiest one to be produced. The other three speaking voices, including falsetto, vocal fry and whisper voices, were randomized across subjects by using all possible order.

Data Analysis

Kay Image Processing System (KIPS) Version 1.11 is an image processing program for analysing images of larynx, which are recorded by stroboscopy and Kay HSV. A number of tools are available in KIPS for image analysis, including play tools, annotation, montage creation, edge detection, glottal area waveform (GAW) with numerical report, FFT point and area analysis with numerical report, kymographic analysis and measurement tools.

The images recorded by the endoscope were analyzed using the slow motion play back function of Kay HSV system and projected at frame rate of 1, 5, 30 frames/s, whereas frame-by-frame analysis was carried out at rate of 1 frame/s by using KIPS. The investigator had used analyzing tools of play, montage, length measurement and GAW analysis in current study. Consecutive frames could be captured from AVI files by montage, which is a time series of image sequence. Hence, dynamic and static analysis of vocal fold vibration was allowed. After montage creation, GAW analysis, with area and opening of glottis, as well as area rate and opening rate of glottis were generated and plotted in the same figure (see Appendix A) for analysis. The data of GAW analysis would also be listed numerically. Glottal cycles and opening were identified in GAW to assist data analysis.

Both objective measurements and subjective judgement were carried out in the selected frames. Objective measurements, including open quotient, glottal size index, vocal fold size indices, were computed. Subjective judgement included symmetry between two vocal folds, glottal closure and number of sudden openings during closed phase.

All objective measurements, including open quotient, frequency, glottal size index and vocal fold size indices, in a particular glottal cycle or in a particular frame were repeated by the investigator until the figures of two consecutive calculations were within 10 per cent of each other. The answer of the last calculation of each measurement was used in data analysis. This reflected that the intra-judge reliability was less than 10% difference. In order to obtain the inter-judge reliability, objective measurements and subjective judgement on half of selected images were repeated by another judge who studied Speech and Hearing Sciences. When the measurements were within 10 per cent of each other, they were regarded as "agree".

Objective Measurements

Open quotient. Open quotient refers to length of open phase in relation to duration of a glottal cycle. It was the ratio of number of frames during open phase to total number of frames in a vibratory cycle. Vibratory cycles were first identified from the projected frames. Two phases, open phase and closed phase, were then identified in the vibratory cycles. Open phase refers to any time of the cycle where there is a glottal opening, whether the vocal folds are opening (moving laterally) or closing (moving medially) (Woo, 1996). Closed phase refers to any time that the glottis is closed (Woo, 1996). It was computed by using the images that recorded at frame rate of 2000 frames/s.

Frequency. Frequency refers to the rate of vocal fold vibration in a second. It was calculated with the formula of "2000 / number of frames per glottal cycle", which was computed by using the images that recorded at frame rate of 2000 frames/s.

Index Computation. Vibratory cycles were identified from the endoscopic images recorded and high speed digital imaging allowed the observation of change during the cycles. Glottal size index and vocal fold size index were computed for each frame in the whole vibratory cycle. For index computation, basic points were set up manually according to the five anatomical landmarks. According to Zemlin (1998) and Li (2000), five anatomical landmarks, including left and right vocal process, posterior midglottic wall, anterior commissure of vocal folds and prominences of arytenoids cartilages, should be identified to compute the indices. These five landmarks of laryngeal structures are shown in Appendix B.

After review of all the recordings, it was found that the five anatomical landmarks mentioned before were not always identified in the images because of the small size of scan area limited by the frame rate or blockage by supraglottic compressions. Therefore, only those images with steady view and clear anatomical landmarks were used to calculate the index. Frames from modal and falsetto registers recorded at frame rate of 1000 frames/s of two subjects, as well as one of their whisper frames, could fulfil these requirements and were used in index computation. It was observed that there were four frames per glottal cycle in modal register, while only two frames per glottal cycle found in falsetto register. As no glottal cycles could be identified in whisper images, readings of four consecutive frames were used for index computation and analysis.

Glottal size index. Glottal size index refers to the “configuration of glottis in terms of length and width” (Li, 2000, p. 11). It was the ratio of glottal length to width. Identification of four landmarks was required, including left and right vocal process, posterior midglottic wall and anterior commissure of vocal folds.

Vocal fold size index. Vocal fold size index refers to the “configuration of vocal fold in terms of length and width” (Li, 2000, p. 11). It was the ratio of length to width of vocal fold. Vocal fold size index was computed for each vocal fold. Identification of five landmarks was required, including the left and right vocal process, anterior commissure of vocal folds and prominences of arytenoids cartilages.

Subjective Judgment

A form (Appendix C) was filled in by both the investigator and another judge on different occasions to rate the laryngeal images for subjective judgement. Multiple choice questions were given and the most appropriate and representative answer were selected from the choices provided. The images that recorded at frame rate of 2000 frames/s were rated.

Symmetry. Symmetry refers to “the extent of which the left and right vocal folds are mirror image of each other” (Cutler & Cleveland, 2002, pp. 464). The percentage of time with left and right vocal folds being symmetric was emphasized.

Glottal Closure. Glottal closure refers to the “degree that the vocal folds approximate when they are in closed phase” (Cutler & Cleveland, 2002, pp. 464). Figures of four choices (complete, posterior, anterior, incomplete and others) were provided (see Appendix B), so as to decrease the task difficulty and enhance the subjective judgment.

Number of Openings during Closed Phase. Number of openings during closed phase referred to sudden openings of glottis during the closed phase in a vibratory cycle. The start of closed phase is marked by the first frame with a closed glottis, which is right following the frame with a slight glottal opening, within a glottal cycle.

Results

Some subjects were unable to tolerate the rigid endoscope as it became too hot after few seconds, especially in vocal fry and whisper voices. Trigger of gag reflex prevented the recordings, and caused data elimination in both objective measurements and subjective judgement. Since the view of vocal folds was obscured by other structures, such as false vocal folds and epiglottis, as well as small scan area owing to technology limitation, some landmarks could not be identified in frames. Hence, only images with clear and stable view of larynx were used for index calculation. Such obscuration and limited scan area led to further data elimination in glottal size index, and vocal fold size indices. Table 1 listed the number of subjects with recordings in phonation of different speaking voices. Table 2 listed the number of subjects analysed in both objective measurements and subjective judgement.

Reliability on Objective Measurements of vocal fold vibration in the four vocal registers

Half of the selected images were re-analyzed by the investigator and another judge on separate occasions to obtain the intra- and inter-judge reliability (see Appendix D). The percentage of agreement of intra-judge reliability of all objective measurements and subjective judgement was higher than 90%, while that of inter-judge reliability of all measures, except left and right vocal fold size index, was higher than 85%.

Table 1.

Number of subjects with image recordings during production of different speaking voices

		Vocal Registers							
	Total number	Modal		Falsetto		Vocal fry		Whisper	
Frame rate		1000/s	2000/s	1000/s	2000/s	1000/s	2000/s	1000/s	2000/s
N	6	6	6	6	6	6	4	6	3

Table 2.

Number of subjects demonstrated clear view of larynx, who was analyzed in both objective measurements and subjective judgement

	Vocal Registers			
	Modal	Falsetto	Vocal fry	Whisper
Open Quotient and Frequency				
Frame rate	2000 frames/s	2000 frames/s	2000 frames/s	2000 frames/s
N	6	6	3	2
Glottal size Index and Vocal fold size Index				
Frame rate	1000 frames/s	1000 frames/s	1000 frames/s	1000 frames/s
N	2	2	0	1
Symmetry, Glottal Closure and Number of openings during closed phase				
Frame rate	2000 frames/s	2000 frames/s	2000 frames/s	2000 frames/s
N	6	6	3	2

Objective Measurements

Table 3 lists the number of cases, group means (and standard deviation) associated with objective measurements of endoscopic data, including open quotient and frequency. Table 4 lists the number of cases, group means (and standard deviation) associated with glottal size index, and left and right vocal fold size indices in a glottal cycle of modal, falsetto and whisper voices.

Table 3.

The number of cases, group means (and standard deviation) associated with open quotient and frequency based on frames recorded at frame rate of 2000 frames/s.

	Vocal Registers			
Measure	Modal	Falsetto	Vocal fry	Whisper
		Open quotient		
N	6	6	3	2
Mean (SD)	0.52 (0.06)	0.76 (0.22)	0.21 (0.05)	1.00 (0.00)
		Frequency		
N	6	6	3	2
Mean (SD)	263.22Hz (26.64)	483.33Hz (40.82)	128.10Hz (9.06)	–

Key: SD = Standard Deviation

Table 4.

The number of cases, group means (and standard deviation) associated with glottal size index, and left and right vocal fold size index within a glottal cycle based on frames recorded with frame rate of 1000 frames/s.

	Frame number within a glottal cycle			
Measure	Frame 1	Frame 2	Frame 3	Frame 4
		Glottal size index		
Modal register N = 2				
Mean (SD)	0.00 (0.00)	31.74 (3.41)	10.76 (0.90)	7.75 (1.09)
Falsetto register N = 2				
Mean (SD)	74.17 (8.25)	14.66 (8.59)	–	–
Whisper register N = 1				
	13.54	13.54	13.54	13.54
		Left vocal fold size index		
Modal register N = 2				
Mean (SD)	3.35 (0.92)	3.42 (1.01)	3.51 (1.06)	3.45 (1.00)
Falsetto register N = 2				
Mean (SD)	5.38 (0.33)	5.27 (0.23)	–	–
Whisper register N = 1				

	6.32	6.32	6.32	6.32
	Right vocal fold size index			
Modal register N = 2				
Mean (SD)	3.28 (0.85)	3.29 (0.82)	3.37 (0.91)	3.17 (0.69)
Falsetto register N = 2				
Mean (SD)	5.31 (0.13)	5.33 (0.10)	–	–
Whisper register N =1				
	6.14	6.14	6.14	6.14

Key: SD = Standard Deviation

Open Quotient. The result of Wilcoxon Signed Ranks test revealed no statistically significant difference of open quotient across Modal and Falsetto registers [*T* (5) = -1.75; p = 0.08]. The mean of open quotient for vocal fry register is 0.21, which is much less than that of modal and falsetto registers (Table 3), and open quotient for whisper is 1.00, which is much larger that that of modal and falsetto registers, but no statistical tests could be carried out as there were not enough data for analysis.

Frequency. The result of Wilcoxon Signed Ranks test revealed a statistically significant difference of frequency across Modal and Falsetto registers [*T* (5) = -2.23; p = 0.03]. The mean of frequency for vocal fry register is 128.10 Hz, which is much less than that of modal and falsetto registers (Table 3), but no statistical tests could be done because of the small sample size. No frequency could be calculated for whisper voices as the number of frames per glottal cycle could not be counted in frame-by-frame analysis.

Glottal Size Index. The result of Freidman test revealed no statistically significant difference of glottal size index between the first two frames across Modal and Falsetto registers in a glottal cycle [*F* (3, 2) = 6.00; p = 0.11]. Figure 1 showed the change of glottal size index across time in modal, falsetto and whisper registers.

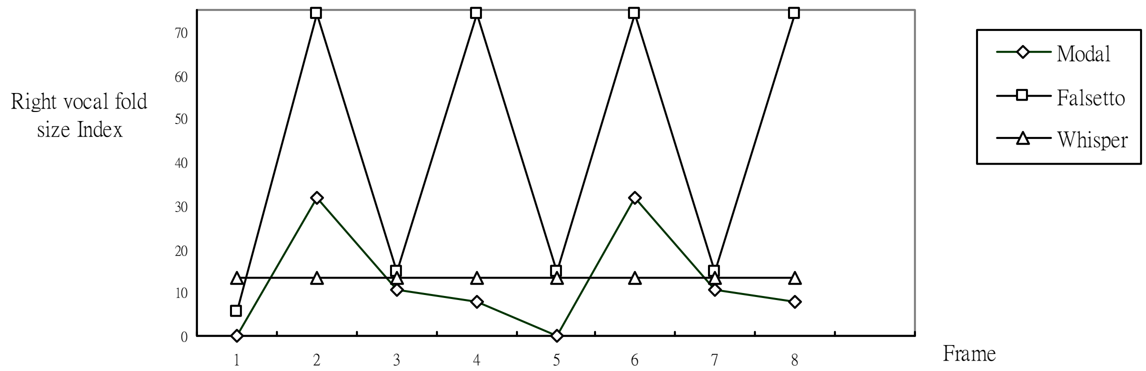


Figure 1. *Glottal size Index in Modal, Falsetto and Whisper speaking voices in terms of time (represented by frame number), which is recorded at frame rate of 1000 frames/s.*

Left Vocal Fold Size Index. The result of Freidman test revealed no statistically significant difference of left vocal fold size index within a glottal cycle in Modal register [F (3, 2) = 5.84; p = 0.12], and in Falsetto register [F (1, 2) = 2.00; p = 0.16]. Since there were not enough cases for data processing in whisper, no statistical test could be carried out. However, the left vocal fold size index within a glottal cycle in whisper voices were identical.

The result of Freidman test revealed no statistically significant difference of left vocal fold size index between the first two frames across Modal and Falsetto registers in a glottal cycle [F (3, 2) = 5.84; p = 0.12]. Figure 2 showed the left vocal fold size index across time in modal, falsetto and whisper registers.

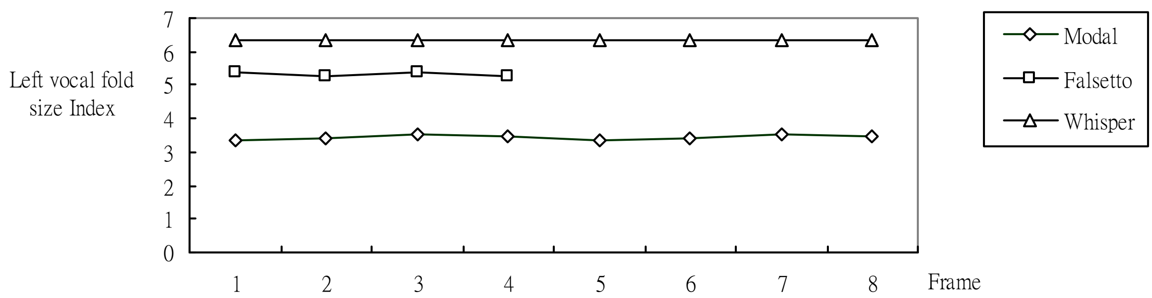


Figure 2. *Left vocal fold size Index in Modal, Falsetto and Whisper speaking voices in terms of time (represented by frame number), which is recorded at frame rate of 1000 frames/s.*

Right Vocal Fold Size Index. The result of Freidman test revealed no statistically significant difference of right vocal fold size index within a glottal cycle in Modal register [F (3, 2) = 4.20; p = 0.24], and in Falsetto register [F (3, 2) = 1.00; p = 0.32]. Since there were not enough cases for data processing in whisper voice, no statistical test could be done. However, the right vocal fold size index within a glottal cycle of whisper voice were identical.

The result of Freidman test revealed no statistically significant difference of right vocal fold size index between the first two frames across modal and falsetto registers in a glottal cycle [F (3, 2) = 5.21; p = 0.16]. Figure 3 showed the right vocal fold size index across time in modal, falsetto and whisper registers.

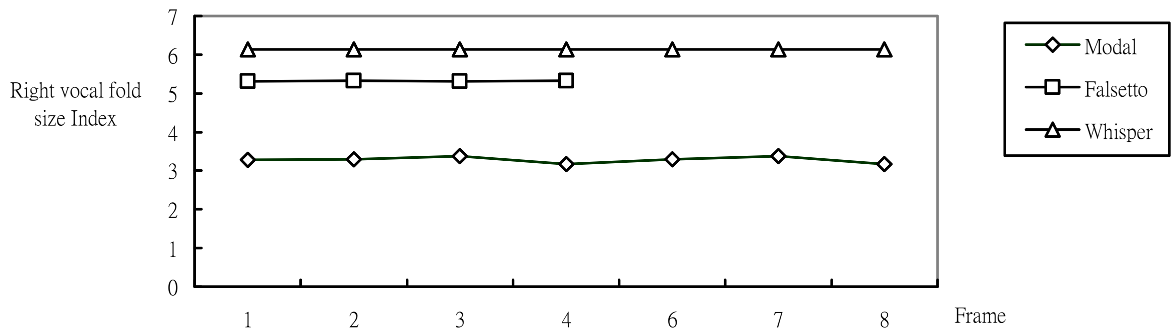


Figure 3. *Right vocal fold size Index in Modal, Falsetto and Whisper speaking voices in terms of time (represented by frame number), which is recorded at frame rate of 1000 frames/s.*

Glottal Area Waveform (GAW) Analysis. Only one appropriate GAW plot could be showed by using the associated processing program, KIPS. It was the modal voice of one of the subjects in index computation, recorded at frame rate of 1000 frames/s (see Fig. 5).

Subjective Judgement

Table 5 lists the results of each subject associated with subjective judgement of the images, including symmetry, glottal closure and number of openings during closed phase.

Symmetry. All subjects produced different speaking voices, including modal, falsetto, vocal fry and whisper, with symmetrical left and right vocal fold movement with 100% of the time during a glottal cycle.

Glottal Closure. The results showed that four subjects had complete glottal closure, while two subjects had posterior chink in glottal closure in modal speaking voices. For falsetto register, there were four subjects who had posterior gap glottal configuration, whereas two subjects had incomplete glottal closure (Table 5). All three subjects with recordings of vocal fry register had complete glottal closure during phonation. The only two subjects with clear view of vocal folds in whisper production had incomplete glottal closure during phonation.

Number of Openings during Closed Phase. There were no sudden openings during the closed phase in these four speaking voices (i.e. modal, falsetto, vocal fry and whisper) in all subjects.

Table 5.

The results of Subjective Judgement (Symmetry, Glottal Closure and Number of openings during closed phase) of each subject by examining images recorded at frame rate of 2000 frames/s.

Subjective Judgement	Vocal Registers			
	Modal	Falsetto	Vocal fry	Whisper
		Symmetry		
CCT	100%	100%	–	100%
CYF	100%	100%	100%	100%
LTM	100%	100%	–	–
MS	100%	100%	100%	–
SWL	100%	100%	–	–
WHY	100%	100%	100%	–
		Glottal Closure		
CCT	Complete	Posterior	–	Incomplete
CYF	Posterior	Incomplete	Complete	Incomplete
LTM	Posterior	Incomplete	–	–
MS	Complete	Posterior	Complete	–
SWL	Complete	Posterior	–	–
WHY	Complete	Posterior	Complete	–

	Number of openings during closed phase			
CCT	0	0	–	0
CYF	0	0	0	0
LTM	0	0	–	–
MS	0	0	0	–
SWL	0	0	–	–
WHY	0	0	0	–

Discussion

The purposes of this study was to examine the vibratory patterns of vocal folds in four speaking voices, including modal register, falsetto register, vocal fry register and whisper, by using high speed digital imaging techniques, and to develop a physiological definition for these registers. Both objective measurement and subjective judgement were used to evaluate the vocal fold vibrations in the four different speaking voices.

Vocal Fold Vibratory Patterns in Different Vocal Registers

No statistically difference across speaking voices was obtained in all the objective measurements (including open quotient, glottal size index, left and right vocal fold size index), except in frequency. The results of glottal size index, and left and right vocal fold size index were also statistically insignificant within modal and falsetto registers. Concerning the small sample size, it was difficult to obtain statistically significant difference across and within different speaking voices. Hence, the investigator will compare the means for a descriptive analysis in the following sections.

Modal Register. In modal register, the mean open quotient was 0.52 (Table 3), reflecting that the open phase was about half of the glottal cycle, which was comparable to the findings of Woo (1996) that the open phase was half of a glottal cycle. The mean frequency of modal speaking voices was 263.22 Hz (Table 3).

Glottal size index (see Fig. 1), frames (see Fig. 4) and glottal area waveform (GAW) within a glottal cycle (see Fig. 5) revealed that maximum opening occurred at about 3ms of a glottal cycle. As the glottis was opening with small glottal width at 2 ms (frame 2 of a glottal cycle), the glottal size index reached its maximum value. The widening of glottal opening resulted in a decrease in glottal size index and reaching a maximum point in GAW analysis.

The smallest vocal fold size index (from 3.17 to 3.51 of left and right vocal folds) (see Table 4) indicated that thickness of vocal fold in modal register was relatively larger, while the length of vocal fold was relatively shorter, compared with falsetto and whisper voices. Such thickness and length were mentioned without significant change during a glottal cycle.

Vocal folds vibrated symmetrically in falsetto voices (Table 5). Posterior gap configuration existed in glottal closure (see Table 5 & Fig. 6) in two subjects, supporting the finding of Titze (1994, pp. 252-253) that “females always use the mixture of modal and falsetto registers in their speaking voices”, as posterior chink is found in falsetto registers.

Frame 439 440 441 442 443 444 445 446

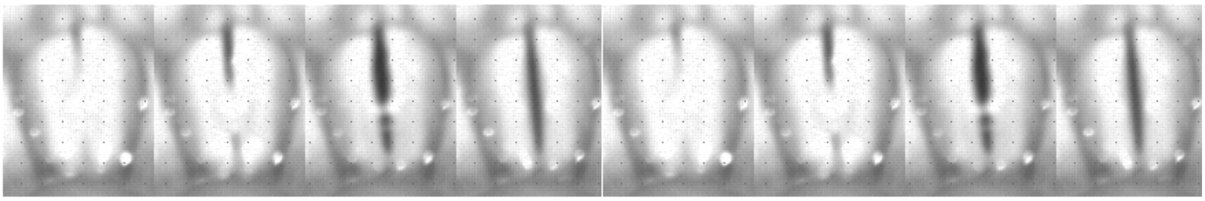


Figure 4. *Frames of Modal register recorded by using Kay HSV at frame rate of 1000 frames/s, with two glottal cycles presented*

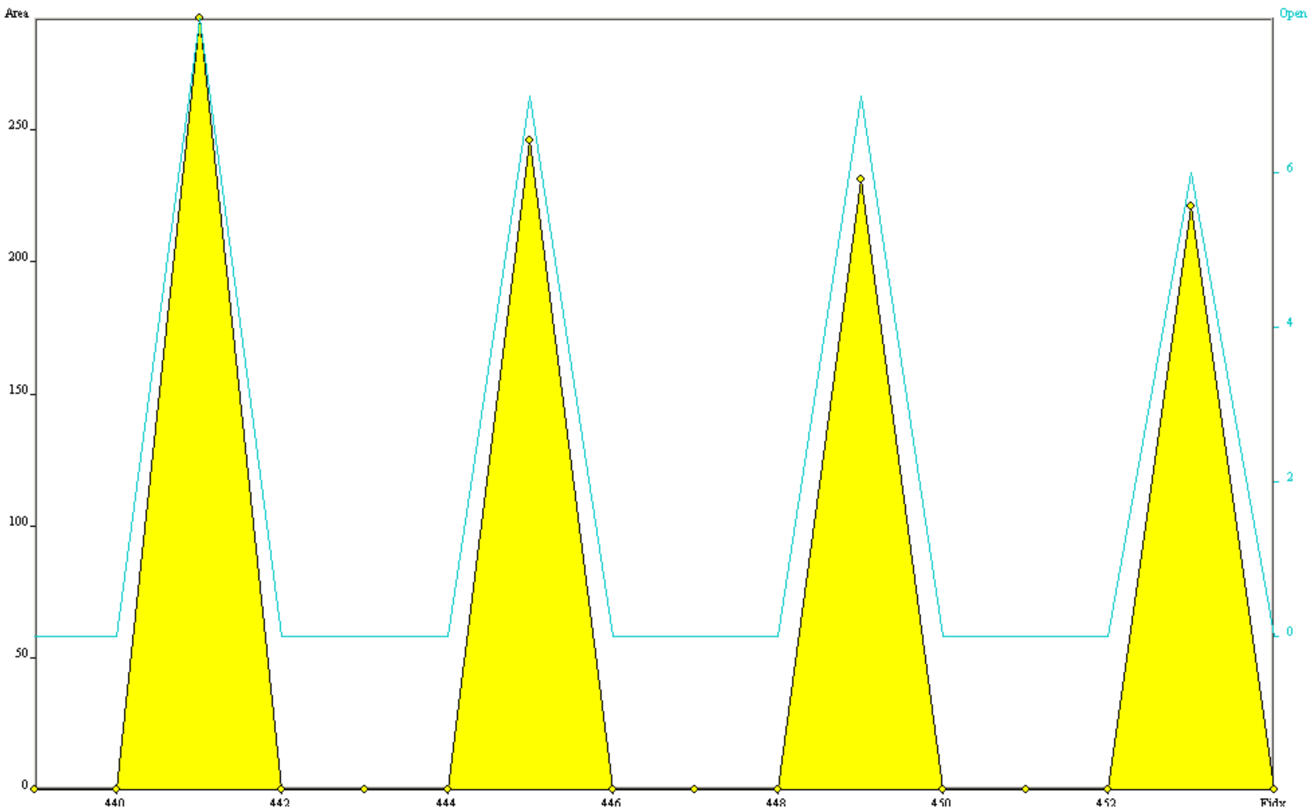


Figure 5. *Glottal area waveform of Modal register in corresponding frames of Figure 5, with 4 glottal cycles presented. (Key: Yellow portion – GAW, Blue line – opening).*

Falsetto Register. Falsetto voice had a mean open quotient of 0.76 (Table 3), reflected that a long open phase in females. Two subjects even showed no closed phase, which was comparable to the finding in previous study (Hollien, 1974) that rapid or no closed phase was noted (Hollien, 1974). However, the open phase observed in current study is longer than two-third of a glottal cycle found by Woo (1996). The mean frequency was 483.33 Hz (Table 3).

Glottal size index within a glottal cycle revealed no complete closure of glottis occurred in falsetto speaking voice (Table 4), but complete closure of glottis were rated subjectively in the images recorded at frame rate of 2000 frames/s, from the two subjects included in index computation (see Appendix E). Two reasons are suggested to account for this phenomenon. The first was that frame rate of 1000 frames/s (see Fig. 7 & Fig. 10) was not fast enough to capture the short closed phase in falsetto speaking voices as only two frames were obtained for each glottal cycle; and the second was that posterior chink existed in closed phase instead of complete glottal closure.

The medium vocal fold size index (from 5.27 to 5.38 in left and right vocal folds) (Table 4) reflected that thickness and length of vocal folds of falsetto register were medium compared with modal and whisper voices. Such thickness and length did not change during the whole glottal cycle.

An interesting vocal fold vibratory pattern was found, in which, four subjects had complete glottal closure in modal phonation, then a posterior chink was observed in falsetto speaking voice. Two subjects were observed to have posterior gap in modal phonation, and incomplete glottal closure was found during falsetto production (Table 5).

Vocal Fry Register. In vocal fry register, the mean open quotient was 0.21 (Table 3), reflecting very long closed phase and a rapid open phase during a glottal cycle (about one-fifth of a glottal cycle). This vibratory pattern was comparable to findings in stroboscopic and EEG studies in the literature (Blomgren et al., 1998; Zemlin 1988). The mean frequency was 128.10 Hz (Table 3).

Although no glottal size and vocal fold size indices could be computed owing to obscuration of landmarks, thick and short vocal folds were observed in vocal fry register (see Fig. 9). A flaccid vibrating margin and strong medial compression of vocal folds were simultaneously noted in investigating the frames (see Fig. 9).

Vocal folds vibrated symmetrically (Table 5). Complete glottal closure was observed, with only one opening found in each glottal cycle, but no doublet or triplet of openings in a single glottal cycle (see Table 5 & Fig. 9). This result contradicted with the findings of EEG studies that doublet or triplet openings were found in a single glottal cycle in the literature (Blomberg et al., 1998). Supraglottic compression (both anterior-posterior and false vocal fold compression) existed in one out the four subjects recorded. Such supraglottic activity nearly obscured the whole vocal folds.

Whisper. Whisper had a mean open quotient of 1.00 (Table 3), reflected no closed phase and the glottis kept open during the whole glottal cycle. The lengthening of glottis and widening of glottal opening led to a relatively small glottal size index (13.54) in eight consecutive frames (see Fig. 1), compared with modal and falsetto speaking voices. The unchanged glottal size index (Table 4) indicated that the length and width of glottis had no significant change during phonation. The largest vocal fold size index (6.14 and 6.32 of left and right vocal folds) (Table 4) reflected that thickness of vocal fold in whisper speaking voice was the smallest, while length of vocal fold was the longest, compared with modal and falsetto registers. Such thickness and length had no significant change during the whole glottal cycle.

Incomplete glottal closure was observed, without any contact or approximation of left and right vocal folds in whisper speaking voices (see Fig. 8), which is the same as one configuration suggested by Solomon et al. (1989) as reported by Colton & Casper (1996).

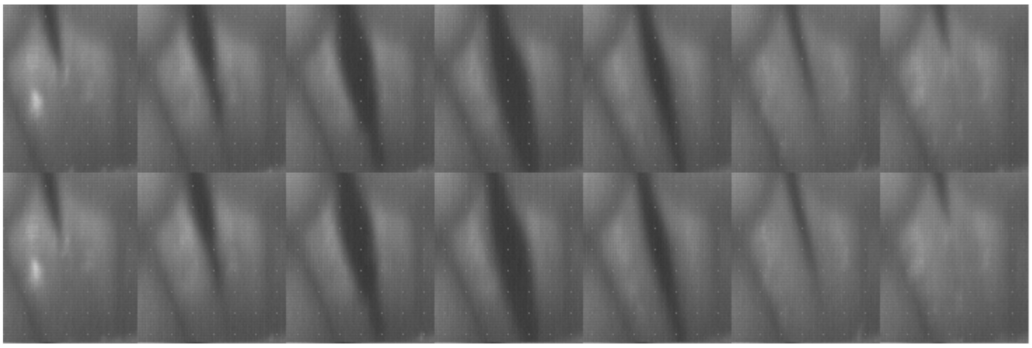


Figure 6. *Frames of Modal register recorded by using Kay HSV at frame rate of 2000 frames/s, with two glottal cycles presented.*

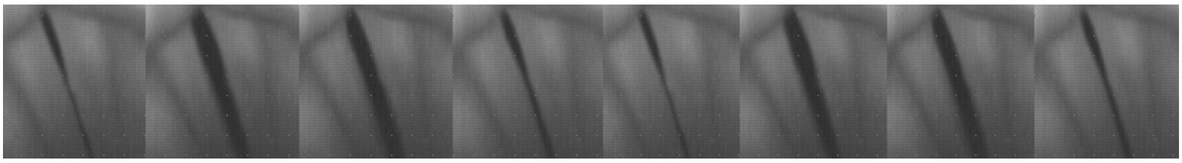


Figure 7. *Frames of Falsetto register recorded by using Kay HSV at frame rate of 2000*

frames/s, with two glottal cycles presented.

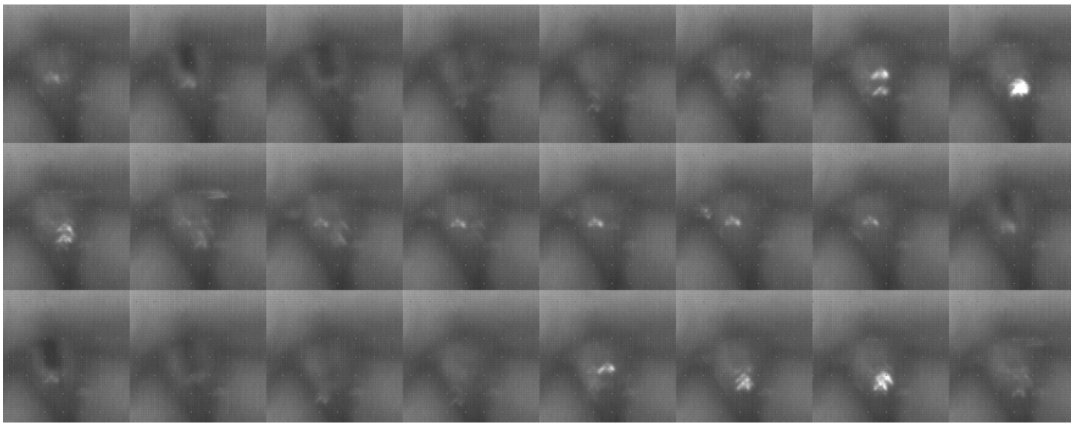


Figure 9. *Frames of Vocal Fry register recorded by using Kay HSV at frame rate of 2000 frames/s, with one glottal cycle presented.*

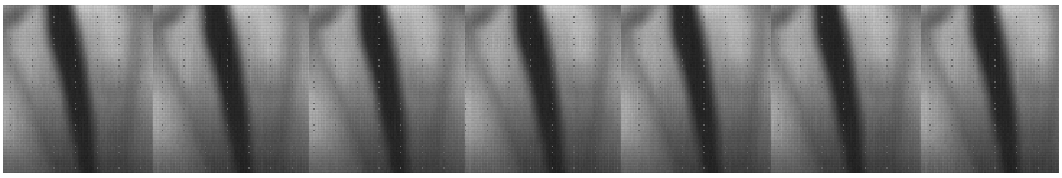


Figure 8. *Frames of Whisper voice recorded by using Kay HSV at frame rate of 2000 frames/s.*

Limitations of Kay Imaging Processing System (KIPS) Version 1.11

Kay Image Processing System (KIPS) Version 1.11 is an image processing program for analysing images of larynx recorded by stroboscopy and high speed video system. A number of tools are available in KIPS for image analysis, including play tools, annotation, montage creation, edge detection, glottal area waveform plot, FFT point and area analysis, kymographic analysis and measurement tools. Its limitations will be discussed in order to improve the accuracy of measurements in future studies.

The first limitation was that the GAW and opening of glottis could not be compared directly across subjects or across registers of the same subject. Different magnifications would be resulted from various lens-to-glottis distances during recording. Hence, relative values, but not absolute values, should be used for comparison across subjects, across registers of the same subject or even across frames within the same register. This problem could be solved by using glottal size index as that used in current study for comparison of glottal opening across registers, subjects or frames since the magnification factor owing to different lens-to-glottis distances, could be ruled out.

The second limitation was the mismatch between the frames and the GAW analysis figure. Figure 4 revealed that a fully open glottis was shown in frame 446 in modal voice production, but the reading in GAW figure was zero (see Fig. 5). Due to the automatic edge detection, in which area which is darker than a particular value, would be regarded as glottis, and then used for GAW analysis. Thus, if the glottal area that was not dark enough for edge detection, zero reading would be resulted for both the glottal area and width of glottal opening (see Fig.10). On the other hand, if the image for analysis was too dark owing to insufficient light source (see Fig.11), automatic edge detection would include areas other than the glottis, such as left and right vocal processes, as the “glottal opening”, and plotted on GAW. This would largely increase the readings in the glottal area waveform and the reading would not be reaching zero even the glottis is closed at frames 118 to 120 in Figure 11. Hence, this analyzing tool was inappropriate to be used if the light source was insufficient.

Limitations and Further Studies

The insertion of rigid endoscope might elicit the protective mechanism of laryngeal structures, which resulted in decrease of the naturalness of voice production. Only speech task of sustained vowel /i/ could be conducted, but not other vowels or even connective speech, which was not a representative phonatory pattern of daily voice use. Some subjects were unable to tolerate the rigid endoscope for too long since the camera became too hot after few seconds. This resulted in missing images and data in vocal fry and whisper speaking voices since it required putting the endoscope more inside towards the vocal folds. In addition, this also reduced the sample size because four subjects were unable to tolerate the hot endoscope by triggering of gag reflex and subsequently eliminated after the trial session.

Concerning images selection for analysis, high incidence of data (four subjects in modal and falsetto, all in vocal fry and five in whisper voices) was excluded in index computation owing to obscuration of more than one landmark by supraglottic activity or the small scan area resulted from technological limitation. In addition, the long saving time owing to technological constrain led to only one images recorded and saved for each vocal register at frame rate of 1000 frames/s and 2000 frames/s. These problems limited the data available for calculation of glottal size index and vocal fold size index. Furthermore, the small sample size reduced the generalization of the results to the populations.

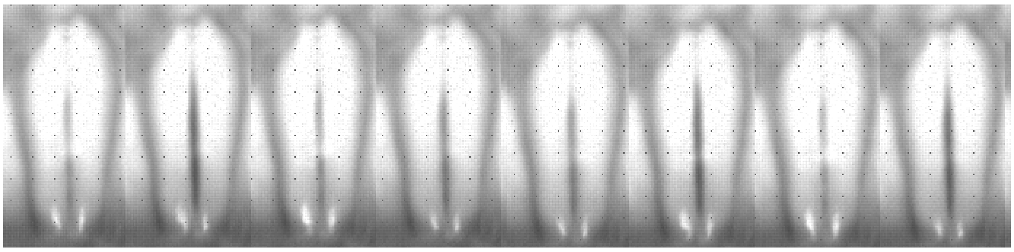
No male subjects were included in this study. As reported by Li (2000), gender difference was found in endoscopic measurements between modal, falsetto and vocal fry speaking voices. Further study should evaluate if gender difference exists and investigate such difference in details by using high speed digital imaging techniques.

The objective and subjective measurements in current study were based on those utilized in stroboscopic studies in previous literature (Li, 2000; Poburka, 1999), as there is still no standard quantitative measurement available. However, these measures could not be completely replicated in high speed imaging studies as they were restricted to analyse static frames, rather than reflecting the detailed transition of vocal fold or glottal configurations during phonation. It is proposed that further study may focus on creating measures that are specific to quantify and qualify the detailed transition of vocal fold vibrations in images recorded by using high speed imaging techniques.

Conclusion

The vocal folds vibratory patterns had been investigated by using Kay High Speed Video system and Kay Image Processing System (KIPS) and described qualitatively. It was observed that different vibratory patterns were associated with different vocal registers. Further study would be needed to create more objective measurements specific to be used for high speed digital images.

Frames 261 262 263 264 265 266 267 268



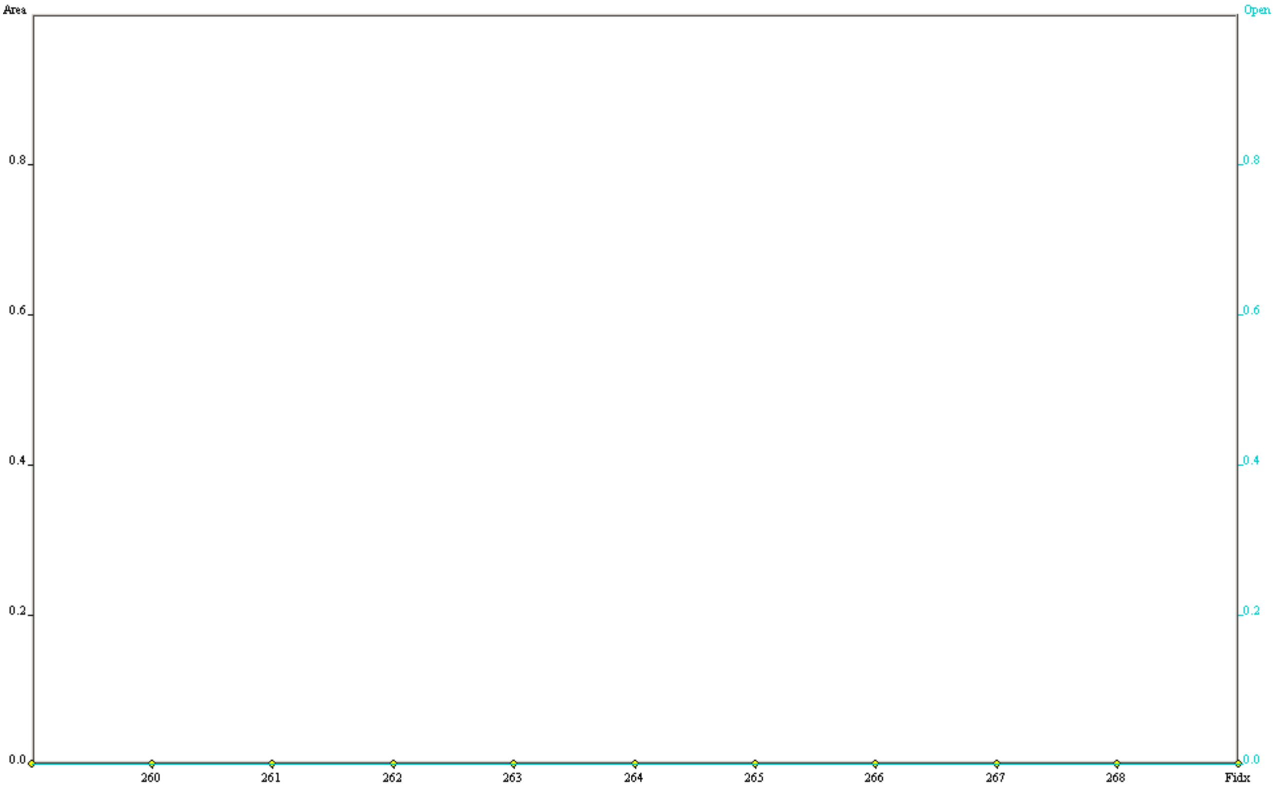


Figure 10. *Frames of Falsetto register recorded by using Kay HSV at frame rate of 2000 frames/s and the glottal area waveform (GAW) in corresponding frames, with four glottal cycles presented.*

Frame 115 116 117 118 119 120 121 122

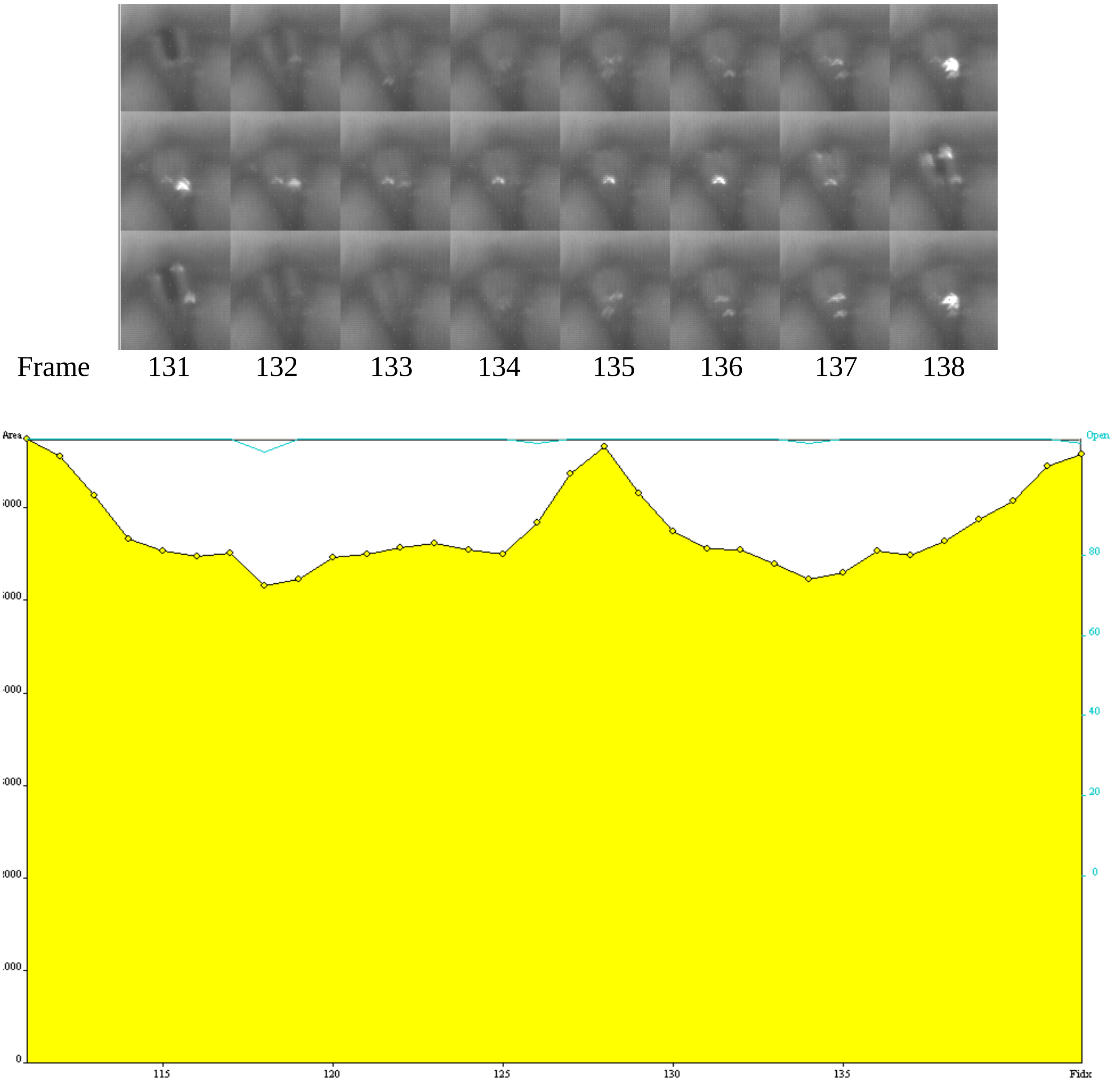


Figure 11. *Frames of Vocal Fry register recorded by using Kay HSV at frame rate of 2000 frames/s and glottal area waveform (GAW) in corresponding frames, with one_glottal cycle presented.*

References

Baken, R.J. (1992). Electroglottography. *Journal of Voice*, *6* (2); 98-110.

Blomgren, M., Chen, Y., Ng, ML. & Gilbert, HR., (1998). Acoustic, aerodynamic, physiologic and perceptual properties of modal and vocal fry registers. *Journal of Acoustic Society of America*, *103*, 2649-2658.

Colton, R. H. & Casper, J. K. (1996). *Understanding voice problems: a physiological perspective for diagnosis and treatment* (2nd ed.). Baltimore: Williams & Wilkins.

Cutler, J. L. & Cleveland, T. (2002). The clinical usefulness of laryngeal videostroboscopy and the role of high-speed cinematography in laryngeal evaluation. *Current Opinion in Otolaryngology & Head & Neck Surgery, 10*(6), 462-466.

Eysholdt, U., Tigges, M., Wittenberg, T., & Proschel, U. (1996). Direct evaluation of high-speed recordings of vocal folds vibrations. *Folio Phoniatrica et Logopaedics*, *48* (4), 163-170.

Hammarbery, B. (1995). High speed observations of diplophonia phonation. In: Osamu, F., Minoru, H. et al. (Eds.) *Vocal Fold Physiology: Voice Quality Control*. San Diego, Calif: Singular Publishing Group. (pp. 343-345).

Hollien, H. (1974). On the vocal registers. *Journal of Phonetics, 2,* 125-142.

Larsson, H., Hertegard, S., Lindestad, P. A., & Hammarberg, B. (2000). Vocal fold vibrations: high speed imaging, kymography, and acoustic: A preliminary report. *Laryngoscope*, *110* (12), 2117-2122.

Li, N. Y. K. (2000). *Physiological characteristics of different types of voice production.* Bachelor of science dissertation, The University of Hong Kong, Hong Kong.

Maurer, D., Hess, M. & Gross, M. (1996). High-speed imaging of vocal fold vibrations and larynx movements within vocalizations of different vowels. *Annals of Octology Rhinology and Laryngology*, *102* (7), 502-507.

Niimi, S. & Miyaji, M. (2000). Vocal fold vibration and voice quality. *Folio Phoniatrica et Logopaedics*, *52,* 32-38.

Poburka, B.J. (1999). A new stroboscopy rating form. *Journal of Voice, 13* (3), 103-113.

Sunberg, J. (1995). Vocal fold vibration patterns and modes of phonation. *Folia Phoniatrica et Logopedica*, *47* (4), 218-228.

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

Tigges, M., Wittenberg, P., Mergell, P. & Eysholdt, U. (1999). Imaging of vocal fold vibration by digital multi-plane kymography. *Computerized Medical Imaging and Graphics*, *23,* 323-330.

Titze, I.R. (1994). *Principles of Voice Production.* New Jersey: Prentice-Hall.

Woo, P. (1996). Quantification of videostrobolaryngoscopic findings - measurements of normal glottal cycle. The American Layngological, Rhinological & Otalogical Society, Inc. pp. 1-27

Zemlin, W. R. (1998). *Speech and Hearing Sciences: Anatomy and Physiology* (4^{th} ed.). Englewood Cliffs: Ally and Bacon.

Acknowledgments

I would like to express my gratitude to Dr. Edwin Yiu for his assistance in rigid endoscopic procedures and the valuable comments throughout the whole project. I would also like to thank Miss Brenda Wun for assisting in data collection and Miss Karen Chan for her fruitful suggestions. Thanks are also expressed to all volunteered subjects for their kind assistance, and to my friends and family for their encouragement and support.

Appendix A

Glottal area waveform (GAW) presented in Kay Image Processing System (KIPS) Version 1.11

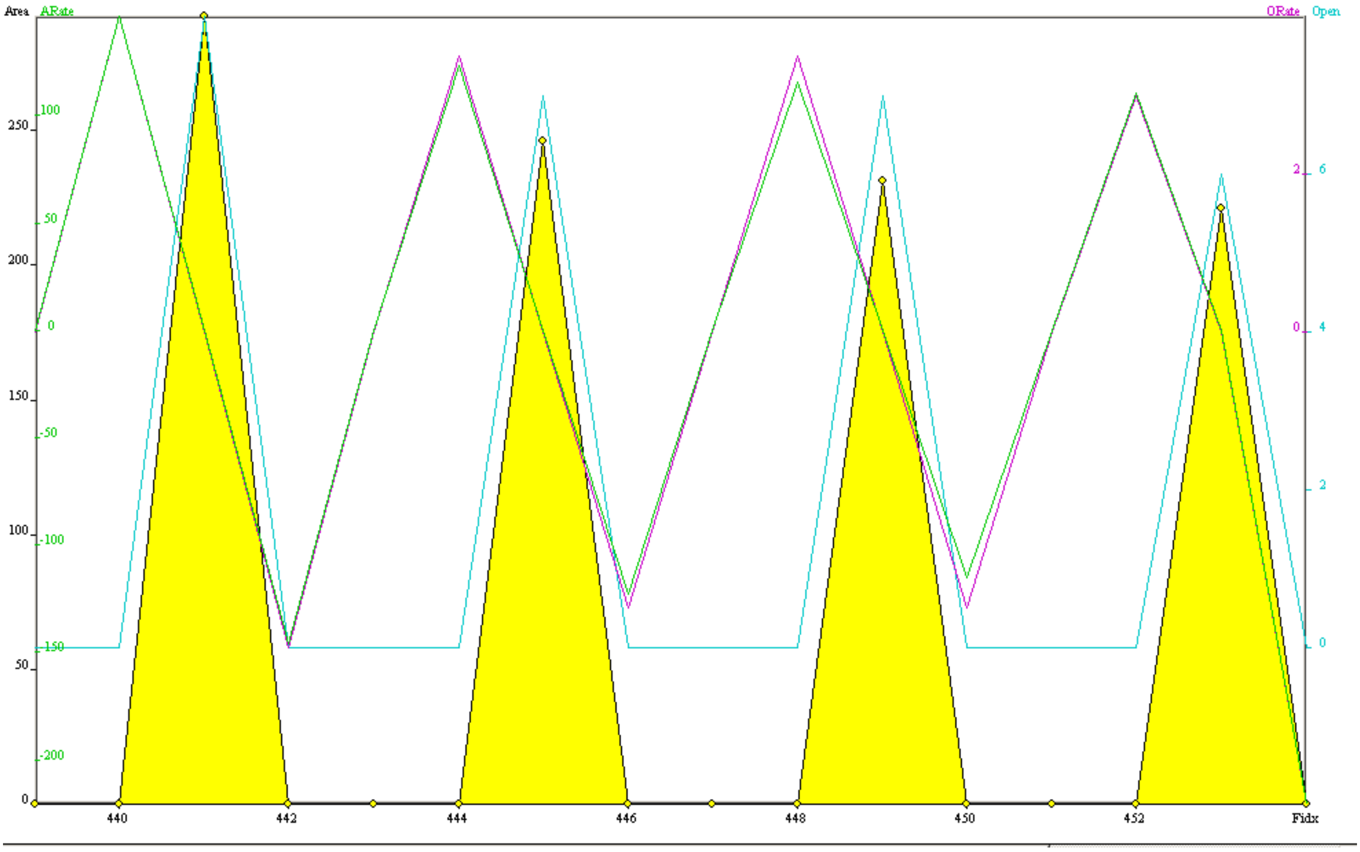
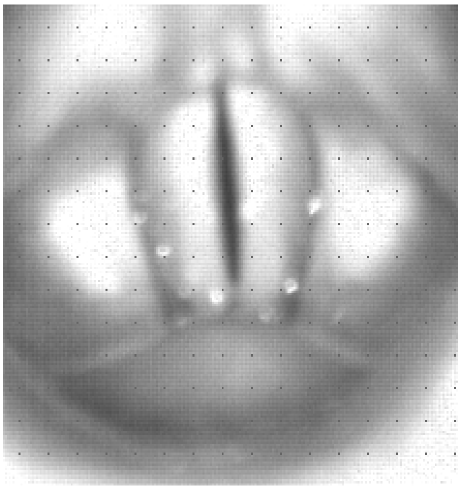
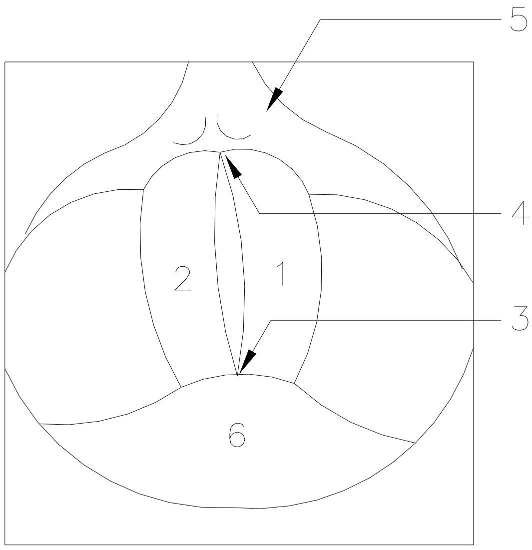


Figure 2. *Glottal area waveform presented in Kay Image Processing System (KIPS) Version 1.11, showing glottal area (yellow portion), area rate (green line), opening (blue line) and opening rate (pink line) of vocal fold vibrations.*

Appendix B

Identification of Anatomical Landmarks for Index Computation





1 – Left vocal process

2 – Right vocal process

3 – Anterior commissure of vocal folds

4 – Posterior midglottic wall

5 – Prominences of arytenoids cartilages

6 – Epiglottis

Figure 1. *Identification of Laryngeal Structures and Anatomical Landmarks for Index Computation.*

Appendix C

High Speed Video (HSV) Subjective Judgement Sheet

Name of rater: ____________________ File name: ____________________

Definitions:

Cutler, J. L. & Cleveland, T. (2002). The clinical usefulness of laryngeal videostroboscopy and the role of high-speed cinematography in laryngeal evaluation. *Current Opinion in Otolaryngology & Head & Neck Surgery*. 10(6), 462-466.

Poburka, B.J. (1999). A new stroboscopy rating form. *Journal of Voice*, 13 (3), 103-413.

- Open phase – any time of the cycle where there is a glottal space, whether the vocal folds are opening (moving laterally) or closing (moving medially)
- Closed phase – any time the glottis is closed

Subjective Judgement (Please circle the most representative one)

1. Symmetry between two vocal folds

Definition – phase symmetry between two vocal folds during vibration

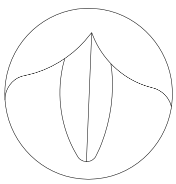
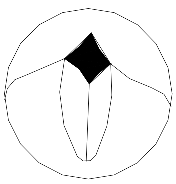
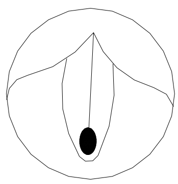
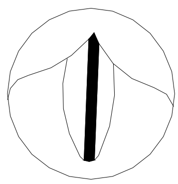
Percentage of time with left and right vocal folds being symmetric is ___________

100% 80% 60% 40% 20% 0%

2. Glottal Closure

Definition – phase symmetry between anterior and posterior portion of vocal folds

Glottal Closure is ________ (Complete Posterior Anterior Incomplete Others)

Complete Posterior Anterior Incomplete

3. Number of openings during closed phase

Definition – a sudden opening of glottis during the closed in a vibratory cycle

No. of openings during closed phase is _________ 0 1 2 3 4 5

Appendix D

Intra- and Inter-judge Reliability of Objective Measurement and Subjective Judgement on half of Selected Images

Table D1.

Intra- and inter-judge reliability of objective measurement and subjective judgement on half of selected images

		Percentage of agreement
	Objective Measurement	
Open quotient	Intra-judge	100.0%
	Inter-judge	88.2%
Frequency	Intra-judge	100.0%
	Inter-judge	88.2%
Glottal size Index	Intra-judge	93.8%
	Inter-judge	87.5%
Left vocal-fold size Index	Intra-judge	93.8%
	Inter-judge	81.3%
Right vocal-fold size Index	Intra-judge	93.8%
	Inter-judge	81.3%
	Subjective Judgement	
Symmetry	Intra-judge	100.0%
	Inter-judge	100.0%
Glottal Closure	Intra-judge	100.0%
	Inter-judge	100.0%
Number of openings during closed phase	Intra-judge	100.0%
	Inter-judge	100.0%

Appendix E

Frames of Falsetto Register recorded at Frame Rate of 2000 frames/s

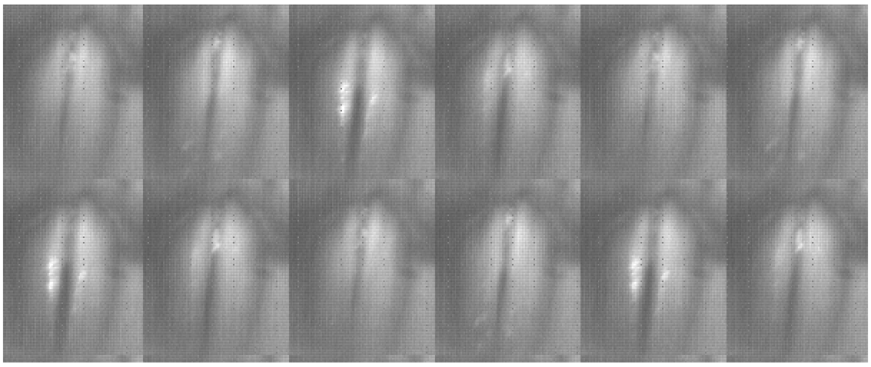


Figure 3. *Frames of Falsetto register recorded by using Kay HSV at frame rate of 2000 frames/s, with 3 glottal cycles presented.*