<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Quantitative analysis of vocal fold vibration in vocally fatigued voice in high speed laryngoscopic images</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Lo, Ching-yin; 老正賢</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2011</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/192896">http://hdl.handle.net/10722/192896</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>The author retains all proprietary rights, (such as patent rights) and the right to use in future works.</td>
</tr>
</tbody>
</table>
Quantitative analysis of vocal fold vibration in vocally fatigued voice in high speed laryngoscopic images

Lo Ching Yin, Andy

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2011.
Quantitative analysis of vocal fold vibration in vocally fatigued voice in high speed laryngoscopic images

Lo Ching Yin, Andy

Abstract

This study examined the vocal vibration pattern of fatigued voice by high speed laryngoscopic imaging. A computer program, High Speed Video Processing (HSVP) program, was used to quantify the laryngoscopic images using indices measuring glottal area, anterior-posterior length of the glottis and the width of the glottis. Twenty participants aged from 18 to 23 years (mean = 21.2 years, s.d. = 1.3 years) with normal voice were recruited to participate in a singing task. Vocal fatigue was induced through prolonged singing. High speed laryngoscopic imaging was taken before and after the singing task. Images of /i/ phonation were analyzed using the HSVP program. Significant changes were found in the posterior glottal length ratio index and glottal length to width ratio index following vocal fatigue. It was hypothesized that vocal fatigue condition would lead to a compensatory hyperactive laryngeal adjustment. This lengthened the vocal folds anterio-posteriorly and made the glottis narrower. The high speed imaging technique using quantitative analysis has the potential for early identification of vocally fatigued voice.
**Introduction**

**Vocal fatigue**

Vocal fatigue is commonly complained by people, such as teachers, salesmen and singers, who use their voice after a prolonged period. It is characterized by an increased effort in voicing, harshness, strained voice quality, dry and sensation of pain in the throat (Gotaas & Starr, 1993; Stemple, Stanley & Lee, 1995). However, researchers have not agreed on a universal definition for vocal fatigue. Some literature considered vocal fatigue as one of the symptoms of voice disorders (Colton, Casper & Leonard, 2006). Others considered it as an isolated phenomenon (Solomon & DiMattia, 2000; Solomon, Glaze, Arnold & van Mersbergen, 2003; Stemple et al., 1995).

Vocal fatigue is defined in this study as vocal tiredness after voice overuse, misuse or abuse (Stemple et al., 1995; Welham & Maclagan, 2003). It could happen in people without any voice problem. Chronic vocal fatigue, however, could be an indicator of subsequent voice disorder (Colton et al., 2006; Stemple et al., 1995).

Investigators have been examining vocal fatigue in a number of ways. Stemple et al. (1995) reported that subjects complained about feeling dry in the throat and effortful speaking after reading aloud for 2 hours. Teachers with vocal fatigue rated their voice with increased harshness, breathiness and strain after a day of teaching (Gotaas & Starr, 1993).

Acoustically, phonation threshold pressure was found to increase after prolonged reading in both women (Solomon & DiMattia, 2000) and men (Solomon et al., 2003). Solomon and DiMattia (2000) found that the phonation threshold pressure increased under low-hydration condition. They argued that the viscosity of the vocal folds increases in a fatigue state. Without sufficient hydration to the vocal folds during prolonged voice use, the stiffness of vocal folds increased and more effort was required to initiate the vibration of vocal folds, leading to an increase in phonation threshold pressure. Stemple et al. (1995) reported a significant increase in the fundamental frequency after two hours of reading. Gelfer,
Andrews and Schmidt (1991) reported that untrained singers demonstrated increased in jitter ratios and reduced in signal-to-noise ratios. Eustace, Stemple and Lee (1996) found an increased airflow rate and reduced maximum phonation time for subjects with chronic laryngeal fatigue, but the fundamental frequency and jitter values were within normal limit. Indeed, the literatures showed mixed results when using acoustic parameters to investigate vocal fatigue. The fact that different methodology was used in different studies might have contributed to these inconsistent results.

**Laryngoscopic view of vocal fatigue**

With the use of video strobeoscopic analysis, anterior glottal chinks (Eustace et al., 1996; Stemple et al., 1995) and abnormal spindle-shaped closure (Eustace et al., 1996; Solomon & DiMattia, 2000) were found in subjects with vocal fatigue. The presence of chink or incomplete closure correlated with the perceptual finding of increased breathiness and airflow. Stemple et al. (1995) hypothesized that the thyroarytenoid muscles became strained and weak in a fatigue state. Such weakness caused bowing of the vocal folds and led to incomplete closure. Further strained contraction of the muscles gave rise to a sensation of pain and effort in voicing. In a study by Mann et al. (1999), a significant increase in vocal fold edema was reported after a 5-day vocal demand training. It was believed that vocal fold tissues were damaged following a long vocal demand task. Gelfer, Andrews and Schmidt (1996) found an increase in amplitude of glottal opening after one hour loud reading. Gelfer et al. suggested that the participants adapted the loud speaking mode even during the endoscopy task.

None of these investigations provided direct view on the intra-cycle vocal fold vibration pattern over time in fatigued voice. Videostroboscopy provides superior view of the vocal fold and allows investigation on the vibration pattern. However, it cannot provide images that capture a complete vibratory cycle due to its limited recording rate at around 24-30 frames per second. Videostroboscopy produces images of vibratory motion by sequencing different frames from different phases across many glottal periods (Deliyski et al.,
This means the videostroboscopy display does not represent actual glottal cycles.

**High-speed laryngoscopy**

Current technology allows high speed laryngoscopic imaging system to capture up to 8000 frames per second. With such advanced temporal resolution, it is possible to examine the complete cycle-by-cycle vocal vibration pattern recorded by digital image recording (Deliyski et al., 2008). Any irregular vibration cycle or phase symmetry could be identified in detail using the technique. Patel, Dailey and Bless (2008) compared the capability of high speed imaging and videostroboscopy in identifying the vibratory features in dysphonic voice. The videostroboscopy identified 37% of the dysphonic subjects correctly, compared with 100% identification rate when high speed imaging was used. Their results indicated that the videostroboscopy was not effective in analyzing aperiodic signals or pathological voices. The use of high speed laryngoscopic imaging system in investigating physiological change of fatigued vocal fold should yield a more precise finding.

High speed imaging technique has been used more frequently over the decade due to reduced cost and improved resolution of the equipment. A number of researchers utilized this technique to examine normal phonatory physiology and pathological voice. The high speed data can be analyzed qualitatively and quantitatively. For example, Inwald, Dollonger, Schuster, Eysholdt and Bohr (in press) studied individuals with organic and nonorganic voice disorder by means of subjective (visual) and objective evaluation. The subjective visual evaluation included rating on the degree of mucosal wave, glottal closure, asymmetry, phonovibrogram and mucus deposit. The objective parameters included degree of asymmetry and perturbation measures derived from the high speed images. Inwald et al. (in press) found that the combined-method evaluation were able to differentiate dysphonic voice from normal voice. Larsson, Hertegard, Lindestad and Hammarberg (2000) applied the high speed line scanning technique, kymography, with acoustic measurements to investigate the vibration pattern of diplophonic phonation. The kymographic results revealed a specific pattern of
glottal closure. Attempts have been made to make the analysis more objective. In a recent study, Mehta, Deliyski, Quatieri and Hillman (2011) quantified the left-right displacement waveforms on kymography into indices of pixel in order to examine the vocal fold vibratory asymmetry in comfortable and pressed phonations. No significant difference in asymmetry was found between these two types of phonations. Quantitative measurement was also used in examining the voice onset on normal phonation. Koster, Marx, Gemmar, Hess and Kunzel (1999) quantified and analyzed the change in glottal area and glottal width during different modes of voice onset. No significant differences were found among the different modes of voice onset. Furthermore, they reported a high intra-subject variability. In another study, Yiu, Kong, Fong and Chan (2010) developed a digital signal processing software – High Speed Video Processing (HSVP) program - to quantify these types of measurement: glottal area, glottal width and glottal length. They used the concept of index for the area, length and width to study normal and dysphonic phonation. Yiu et al. found the glottal ratio indices were able to discriminate the dysphonic voice from normal voice. Hess and Gross (1996) used an open quotient (i.e. ratio of the open phase to the glottal period) derived from the high speed imaging frames to examine the vocal vibration pattern for different vowels. Association between vocal fold movement and vowel identity was found (Hess & Gross, 1996).

Qualitative measurement always has the issue of reliability as it is a subjective rating process. Inter-rater reliability in evaluating the images was found to range from 70% to 78% in the study by Patel et al. (2008). This moderate correlation suggests that the subjective rating may be dependent on the experience of the judges. Quantitative measurement (e.g. quantified glottal area in pixels) should provide a more objective and less variable method. It allows data summarization and grouping among individual data for comparison.

The High Speed Video Processing program

This study used the HSVP program (Kong & Yiu, 2010) to provide quantitative measurement of high speed laryngoscopic images to analyze the vocal fold vibratory pattern
of experimental-induced vocal fatigue (see Figure 1). Eight measurements associated with glottal area, anterior-posterior length of the glottis, width of the glottis and three temporal measurements for vocal fold vibration waveform were taken.

The eight ratio indices were full glottal area (GA), left GA, right GA, left glottal width, right glottal width, anterior glottal length, posterior glottal length and glottal length to width ratio indices (see Figure 2). Since the absolute glottal area, length and width could not be measured due to the unknown magnification factors (i.e. the distance between the vocal folds and the laryngoscope), ratio indices were used for the area, length and width of glottis. They were based on dividing the mean measurement (i.e. number of pixels) from 100 vibratory cycles of the minimum glottal opening (i.e. close phase) by the mean measurement from 100 vibratory cycles of the maximum glottal opening (i.e. open phase). The definitions of the eight ratio indices adapted from Kong and Yiu (2010) were as follows:

1. Full glottal area ratio index: It represents the mean total glottal area in the minimum opening over the maximum opening. A decrease in the index will mean a larger glottal opening in the open phase or a smaller glottal opening (i.e. increase in glottal closure) in close phase, and vice versa.

2. Left and right glottal area ratio indices: They represent the left and right halves of the glottal area measured from the longitudinal midline to the left and right glottal edge. Difference between these two indices might reflect asymmetry of the vocal fold vibration.

3. Left and right glottal width ratio indices: They represent the left and right halves of the glottal width measured from the longitudinal midline to the left and right glottal edge. Difference between these two indices might reflect asymmetry of the vocal fold vibration.

4. Anterior and posterior glottal length ratio indices: They represent the anterior and posterior halves of the glottal length measured from the horizontal midline to the anterior and posterior glottal edge. A decrease in the index will mean a longer glottal length in the open phase, and vice versa.
5. Glottal length to width ratio index: It represents the mean glottal length to width ratio in the minimum opening over the maximum opening. A decrease in the index will mean the shape of the glottis become a longer and narrower in the open phase, and vice versa.

Figure 1. The High Speed Video Processing (HSVP) program.

Area ACBD = full glottal area
Area ACBO = left glottal area
Area ADBO = right glottal area
CO = left glottal width
DO = right glottal width
AO = anterior glottal length
BO = posterior glottal length
AB/CD = glottal length to width ratio

Figure 2. Dimension of the glottis as measured by the High Speed Video Processing program.

In theory, a complete glottal closure should give a zero index. In practice, it is not uncommon to find digital noises, which are represented by hot pixels that give rise to positive value of the indices (Yiu et al., 2010). Taken glottal area ratio index as example, a decrease in that index will mean a larger glottal opening in the open phase or a stronger glottal closure in the close phase. While an increase in the index means a smaller glottal opening in the open phase.
phase or reduced glottal closure in the close phase (i.e. incomplete closure).

The other three temporal measurements used were fundamental frequency (F0), open quotient (OQ) and speed quotient (SQ) (see Figure 3). The three measures were based on the average timing of the glottal period, open phase and close phase of each vibratory cycle. The HSVP program transformed the number of frames into a time function by an algorithm of sampling rate at 4000 frames per second. For example, if 200 completed vibratory cycles were identified within a 4000-frame video sample, it will be equivalent to 200 cycles per second (i.e. fundamental frequency is equal to 200Hz). The definitions of the three temporal measurements adapted from Kong and Yiu (2010) were as follows:

1. **Fundamental frequency**: It is the inverse of the glottal period. It represents the number of vibratory cycles per second. It is calculated through analyzing the vibratory cycles from the imaging data instead of the acoustic speech signal. An increase in fundamental frequency will mean a faster vocal fold oscillation in a given time, and vice versa.

2. **Open quotient**: It is calculated by dividing the duration of the open phase by the glottal period. It represents the ratio of the glottal opening over one vibratory cycle. An increase in open quotient will mean a longer glottal opening in a given cycle, and vice versa.

3. **Speed quotient**: It is calculated by dividing the duration of opening by the duration of the closing within the open phase. It represents the symmetry between the open phase and the close phase. An increase in speed quotient will mean a longer glottal opening and shorter glottal closing in a given cycle, and vice versa.

\[
F0 = \frac{1}{D - A} \\
OQ = \frac{C - A}{D - A} \\
SQ = \frac{B - A}{C - B}
\]

Figure 3. Definition of fundamental frequency (F0), open quotient (OQ) and speed quotient (SQ) using the vocal fold vibration images
Study objective

The objective of this study was to examine the vocal vibration pattern of fatigued voice using quantitative analysis of high speed laryngoscopic images. It was hypothesized that fatigued voice would demonstrate a different vibratory pattern from normal pattern due to physiological change in vocal fold. The present study ‘induced’ vocal fatigue in the participants by using prolonged singing. Inducing vocal fatigue allowed observation of the fatiguing effects on normal voice (Welham & Maclagan, 2003). Any observable vocal changes between normal and fatigued voice in the same participant could be attributed to vocal fatigue. The fatigue inducing procedure was based on the procedure used in the study by Yiu and Chan (2003). They employed a singing task to determine the amount of time that was required to induce vocal fatigue. They found the mean singing time for a group of amateur singers to report vocal fatigue was 84.48 minutes when no hydration was given.

Method

Participants

Ten males and ten females aged from 18 to 23 years (mean = 21.2 years, s.d. = 1.3 years) were recruited from the social circles of the investigator (AL). All of the participants were reported to have no prior voice training, voice disorder, neurological disease, smoking and drinking habit. All of them were Cantonese speakers and had tertiary educational level. Participants were excluded if they reported to have respiratory disease such as sore throat and flu one day prior to the examination. All of the participants were rated perceptually by the investigator (AL) to have normal voice quality before the experiment.

Procedure

Singing task

All participants were asked to sing karaoke for at least 95 minutes without rest and
without hydration in order to achieve the vocal fatigue condition. The singing time used in the present study was based on the mean singing time reported in the study by Yiu and Chan (2003) plus two standard deviations. The loudness level of the background music was around 60dB for all participants, and the participants were required to sing over 80dB measured with a sound pressure meter (TES 1530A) at a distance of 30cm from their mouth.

After singing 95 minutes, all the participants were asked if they felt tried. They were asked to continue singing until they feel vocal tiredness if they reported that they were not yet tired. This was used to ensure all the participants achieved a vocal fatigue condition after at least 95 minutes of singing. The final mean singing time was 103.8 minutes (s.d. = 7.2 minutes, range = 95-115 minutes).

**Participant’s self rating of vocal fatigue**

All participants were asked to rate their own vocal conditions on the level of discomfort in the throat, dryness and effort used in voicing, each on an 11-point rating scale (0=normal, 10=most severe) before and after the singing task (see Appendix A).

**High speed imaging**

Richard Wolf GmbH (Knittlingen, Germany) digital high speed imaging system was used to record the laryngoscopic images through a rigid endoscope (Model 5562). This was performed by a qualified speech therapist before and after the singing task. Synchronized sound signals were recorded by a microphone attached to the endoscope. Participants were asked to sustain /i/ as long as possible at their most comfortable pitch and loudness. Two seconds in the middle of the sustained /i/ was captured. Vocal onset and offset were excluded. The system also simultaneously recorded the acoustic signal for fundamental frequency and the intensity level within the two seconds. A total of 8192 frames were available for the two seconds time span for each video recording. Three recordings were taken each before and after the singing task. That means in total there were six high speed video recordings (in AVI format) and six corresponding sound signals (in WAV format) for each participant.
Data analysis

Pre-singing and post-singing high speed imaging data was quantified using the High Speed Video Processing (HSVP) program. Eight ratio indices and three temporal measurements were extracted.

*Extraction of the eight ratio indices*

Eight procedural steps were carried out in order to identify the glottal area, length and width and extract the corresponding eight ratio indices. Figure 4 shows the step 3 to 8.

1. Selection of a pair of pre and post data for each participant. Since it was not possible for the participants to produce their pitch and loudness of the /i/ exactly the same between pre and post endoscopy recording, the most pitch-matched and loudness-matched pairs of the video recording were chosen for analysis. For every participant, one pre-task video was selected among three pre-trial samples, and one post-task video was selected among three pos-trial samples. The pitch and the loudness of each acoustic signal for every pre and post video recording were judged by the investigator (AL) and two best matched samples was selected. This was to minimize the confounding effect on the difference in pitch and loudness between the pre and post video recording. Table 1 showed that the mean fundamental frequency and the intensity of the selected pre-singing and post-singing speech signal recording were not significantly different.

<table>
<thead>
<tr>
<th>Table 1. Mean, standard deviation and Wilcoxon signed rank test for the fundamental frequency and intensity level from the acoustic signal of the selected pre- and post-video data (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean and Standard deviation</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Fundamental frequency (Hz)</td>
</tr>
<tr>
<td>Intensity level (dB)</td>
</tr>
</tbody>
</table>
2. Extraction of the frames. One hundred vibratory cycles would be required from each video recording for analysis. Since the fundamental frequency for each participant was different, around 1000 to 2000 stable frames that contained 100 cycles were extracted from the 8192 frames of the raw video recording according to the participant’s frequency. Frames that did not show a full glottis would not be selected for extraction.

3. Re-sizing and rotation (180 degree) of the images. Due to the current limitation of the HSVP program, the size of the raw video had to crop into 120x256 resolutions. In addition, all of the raw video had to rotate at 180 degree so as to fit into the analysis window in the HSVP program. The anterior edge of the vocal fold should point upward while the posterior edge of the vocal fold should point downward, as shown in Figure 2. These could be done by using the video processing software VirtualDub (a freeware available from http://www.virtualdub.org).

4. Converting the video into grayscale and image rotation. The HSVP program converted the color image data into grayscale automatically as processing of color images was not possible as this moment. In addition, the images could be further rotated so that the longitudinal axis of the glottis aligned with the vertical midline.

5. Image zooming. Zooming allowed enlarging the images for clearer interpretation.

6. Brightness and contrast adjustment. Brightness and contrast of the extracted images were adjusted manually so that the structures of the vocal fold were clearly displayed.

7. Image motion compensation. The HSVP program was able to track the dynamic movement of the images due to the movement of the endoscope and made corresponding adjustment. The glottis remained at the same position across the frames.

8. Delineation of glottis by adding laryngeal landmarks in the images. An analysis window was added to the images, which covered the left, right, anterior and posterior glottal edge. The cross of the window was placed on the midline of the glottis.
After the above procedures, the HSVP program would automatically binarized the pixels in the area within the analysis window into black and white pixels. The total number of the black pixels would represent the area of the glottis. The absolute value (in pixels) of glottal area, length and width would be extracted (see Figure 5). Finally, the eight ratio indices would be calculated automatically.

From top to bottom:
1. Full glottal area;
2. Left glottal area (upper part) and right glottal area (lower part);
3. Left glottal width (upper part) and right glottal width (lower part);
4. Anterior glottal length (upper part) and posterior glottal length (lower part);
5. Glottal length to width ratio

Figure 5. Extraction of the eight parameters in terms of pixels in graphic form.

**Extraction of the three temporal measurements**

Extraction of the F0, OQ and SQ depended on the markings of the exact glottal period of the open and close phase (see Figure 3). It was observed that 70% (28/40) of the pre and post-singing data demonstrated different degree of posterior glottal chink (see Appendix B).
Due to the limitation of the HSVP program, the presence of glottal chink (i.e. presence of pixels) would prevent the marking of the close phase period in a vibratory cycle (i.e. point C and D in Figure 3), which affected the extraction of the three temporal measures.

Therefore, adjustment was made in order to eliminate any glottal chinks for the analysis. Firstly, the extraction followed the same procedures from step 1 to 7 as stated above. In the next step, however, any glottal chinks were excluded in the delineation of the glottis. The analysis window covered the glottis except the glottal chinks (see Figure 6). This allowed the markings as shown in Figure 3. F0, OQ and SQ would then be generated automatically.

![Analysis window](image)

**Figure 6. Exclusion of the glottal chink from the analysis window**

**Inter-rater and intra-rater reliability**

The measurements extraction depended on the subjective process in the procedures stated above. Twenty percent of the randomly selected video data (i.e. eight AVI files) was re-analyzed to determine the reliability on frames selection and the 11 parameters. Intra-rater reliability was carried out by the investigator (AL) two weeks later. Inter-rater reliability was carried out by a final year undergraduate of the Division of Speech and Hearing Science.

**Results**

**Reliability on the frames selection and the 11 parameters**

Table 2 presents the inter-rater and intra-rater reliability measurement on the extraction of frames from the raw high speed imaging videos (8192 frames). Seventy-five percent of the re-analyzed videos had the frame extraction well within 500 frame segment range.
administrated by the investigator (AL), while 50% of the videos had the frame extraction within 1000 frame segment range administrated by the inter-rater.

Table 3 presents the inter-rater and intra-rater reliability on the eight ratio indices. The agreements of the indices were compared using Wilcoxon signed rank test. No significant differences were found in all indices for both intra-rater and inter-rater reliability.

Table 2. The inter-rater and intra-rater reliability on the extraction of frames from the raw high speed imaging videos (n=8)

<table>
<thead>
<tr>
<th></th>
<th>± 250 frames</th>
<th>± 500 frames</th>
<th>± 1000 frames</th>
<th>± 2000 frames</th>
<th>± 4000 frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-rater</td>
<td>37.5% (3/8)</td>
<td>75% (6/8)</td>
<td>87.5% (7/8)</td>
<td>87.5% (7/8)</td>
<td>100% (8/8)</td>
</tr>
<tr>
<td>Inter-rater</td>
<td>0% (0/8)</td>
<td>37.5% (3/8)</td>
<td>50% (4/8)</td>
<td>50% (4/8)</td>
<td>100% (8/8)</td>
</tr>
</tbody>
</table>

Table 3. The inter-rater and intra-rater reliability on the 11 parameters using Wilcoxon signed rank test (n=8)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Intra-rater reliability</th>
<th>Inter-rater reliability</th>
<th>Z</th>
<th>p</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full glottal area ratio index</td>
<td>-1.40</td>
<td>-1.12</td>
<td>0.899</td>
<td>0.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left glottal area ratio index</td>
<td>-0.28</td>
<td>-1.26</td>
<td>0.779</td>
<td>0.208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right glottal area ratio index</td>
<td>-0.42</td>
<td>-1.12</td>
<td>0.674</td>
<td>0.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left glottal width ratio index</td>
<td>-0.51</td>
<td>-0.94</td>
<td>0.612</td>
<td>0.345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right glottal width ratio index</td>
<td>-0.51</td>
<td>-0.67</td>
<td>0.612</td>
<td>0.500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anterior glottal length ratio index</td>
<td>-1.01</td>
<td>-0.14</td>
<td>0.310</td>
<td>0.893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior glottal length ratio index</td>
<td>-1.54</td>
<td>-0.84</td>
<td>0.123</td>
<td>0.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glottal length to width ratio index</td>
<td>-1.54</td>
<td>-1.26</td>
<td>0.123</td>
<td>0.208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundamental frequency (Hz)</td>
<td>-0.34</td>
<td>-0.84</td>
<td>0.732</td>
<td>0.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open quotient (%)</td>
<td>-0.28</td>
<td>-0.56</td>
<td>0.779</td>
<td>0.575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed quotient (%)</td>
<td>-1.47</td>
<td>-1.40</td>
<td>0.141</td>
<td>0.161</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Self rating on vocal condition before and after singing task

Table 4 presents the participants’ self judgment on their vocal conditions before and after the singing task. The non-parametric Wilcoxon signed rank test was used to compare the mean difference of the three rating between pre-singing task and post-singing. Significant differences were found for all three conditions: discomfort in the throat ($Z = -3.94, p < 0.05$), dryness ($Z = -3.94, p < 0.05$) and effort in voicing ($Z = -3.94, p < 0.05$). The non-parametric Wilcoxon signed rank test was used because of the small sample size (< 30 participants).

Besides the rating of these three vocal conditions, five out of the 20 participants reported an increase in harshness of their voice after the singing task, and four reported an increased number of pitch breaks.

Table 4. Mean, standard deviation and Wilcoxon signed rank test for self rating on vocal conditions using an 11-point rating scale (0-10) in pre- and post-task

<table>
<thead>
<tr>
<th>Vocal conditions:</th>
<th>Mean and Standard deviation</th>
<th>Wilcoxon signed rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discomfort in the throat</td>
<td>Pre (n=20) 0.90 (0.91)</td>
<td>Post (n=20) 6.80 (1.74)</td>
</tr>
<tr>
<td>Dryness in the throat</td>
<td>0.95 (1.00)</td>
<td>7.70 (1.78)</td>
</tr>
<tr>
<td>Effort in voicing</td>
<td>0.55 (1.05)</td>
<td>6.95 (1.99)</td>
</tr>
</tbody>
</table>

* $p < 0.05$ (2-tailed)

Ratio indices of the eight parameters before and after singing task

Table 5 presents the eight ratio indices in pre-singing and post-singing tasks. There was a general decrease trend in the mean value of all ratio indices after the singing task except glottal width and anterior glottal length ratio indices. Wilcoxon signed rank test was used to compare the mean difference of the eight ratio indices between the pre-singing and post-singing tasks. Significant differences were found in the posterior glottal length ratio index ($Z = -2.95, p = 0.003$) and glottal length to width ratio index ($Z = -2.65, p = 0.008$).
Table 5. *Mean, standard deviation and Wilcoxon signed rank test for eight indices of pre- and post-task data from normal participants (n=20)*

<table>
<thead>
<tr>
<th>Ratio indices</th>
<th>Mean and Standard deviation</th>
<th>Wilcoxon signed rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre (n=20)</td>
<td>Post (n=20)</td>
</tr>
<tr>
<td>Full glottal area ratio index</td>
<td>0.21 (0.07)</td>
<td>0.18 (0.09)</td>
</tr>
<tr>
<td>Left glottal area ratio index</td>
<td>0.22 (0.08)</td>
<td>0.18 (0.08)</td>
</tr>
<tr>
<td>Right glottal area ratio index</td>
<td>0.20 (0.07)</td>
<td>0.17 (0.10)</td>
</tr>
<tr>
<td>Left glottal width ratio index</td>
<td>0.05 (0.08)</td>
<td>0.08 (0.09)</td>
</tr>
<tr>
<td>Right glottal width ratio index</td>
<td>0.05 (0.07)</td>
<td>0.06 (0.07)</td>
</tr>
<tr>
<td>Anterior glottal length ratio index</td>
<td>0.05 (0.08)</td>
<td>0.11 (0.14)</td>
</tr>
<tr>
<td>Posterior glottal length ratio index</td>
<td>0.40 (0.14)</td>
<td>0.24 (0.15)</td>
</tr>
<tr>
<td>Glottal length to width ratio index</td>
<td>2.79 (1.12)</td>
<td>1.68 (1.52)</td>
</tr>
</tbody>
</table>

* p<0.05 (2-tailed)

Symmetry measurement of the left-right comparison of the glottal area and glottal width

Tables 6 and 7 present the comparison between the left and right glottal area (Table 6) and glottal width (Table 7) in pre and post task conditions using Wilcoxon signed rank test.

No significant differences were revealed in left and right glottal area ratio indices in both pre and post task conditions. No significant differences were revealed in left and right glottal width ratio indices in the two conditions.

Table 6. *Wilcoxon signed rank test for left and right glottal area ratio indices*

<table>
<thead>
<tr>
<th>Ratio indices</th>
<th>Mean and Standard deviation</th>
<th>Wilcoxon signed rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left glottal area ratio index</td>
<td>Right glottal area ratio index</td>
</tr>
<tr>
<td>Pre-singing task</td>
<td>0.22 (0.08)</td>
<td>0.20 (0.07)</td>
</tr>
<tr>
<td>Post-singing task</td>
<td>0.18 (0.09)</td>
<td>0.17 (0.10)</td>
</tr>
</tbody>
</table>
Table 7. Wilcoxon signed rank test for left and right glottal width ratio indices

<table>
<thead>
<tr>
<th>Ratio indices</th>
<th>Mean and Standard deviation</th>
<th>Wilcoxon signed rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left glottal width ratio index</td>
<td>Right glottal width ratio index</td>
</tr>
<tr>
<td>Pre-singing task</td>
<td>0.05 (0.08)</td>
<td>0.05 (0.07)</td>
</tr>
<tr>
<td>Post-singing task</td>
<td>0.08 (0.09)</td>
<td>0.06 (0.07)</td>
</tr>
</tbody>
</table>

The three temporal measurements of vocal fold vibration before and after singing task

Table 8 presents the mean of the fundamental frequency, open quotient and speed quotient in pre-singing task and post-singing task. Wilcoxon signed rank test was used to compare the mean difference between pre-singing task and post-singing. No significant differences were found in all the temporal measurements: F0, OQ and SQ.

Table 9 presents the comparison between the two fundamental frequency generated from the acoustic signal and from the high speed images using the HSVP program. Result revealed no statistically significant difference between them.

Table 8. Mean, standard deviation and Wilcoxon signed rank test for fundamental frequency, open quotient and speed quotient

<table>
<thead>
<tr>
<th></th>
<th>Mean and Standard deviation</th>
<th>Wilcoxon signed rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre (n=20)</td>
<td>Post (n=20)</td>
</tr>
<tr>
<td>Fundamental frequency (Hz)</td>
<td>239.9 (69.7)</td>
<td>242.9 (72.1)</td>
</tr>
<tr>
<td>Open quotient (%)</td>
<td>69.5 (9.6)</td>
<td>67.1 (11.9)</td>
</tr>
<tr>
<td>Speed quotient (%)</td>
<td>106.4 (20.5)</td>
<td>106.6 (29.4)</td>
</tr>
</tbody>
</table>
Table 9. Mean, standard deviation and Wilcoxon signed rank test for the fundamental frequency derived from the acoustic signal and from HSVP program of the pre- and post-video data (n=20)

<table>
<thead>
<tr>
<th>Fundamental frequency (Hz)</th>
<th>Mean and Standard deviation</th>
<th>Wilcoxon signed rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acoustic signal</td>
<td>HSVP program</td>
</tr>
<tr>
<td>Pre-singing task</td>
<td>240.3 (71.1)</td>
<td>239.9 (69.7)</td>
</tr>
<tr>
<td>Post-singing task</td>
<td>241.2 (72.0)</td>
<td>242.9 (72.1)</td>
</tr>
</tbody>
</table>

**Discussion**

The aim of this study was to examine the vocal vibration pattern of the fatigued voice using high speed laryngoscopic imaging. Vocal fatigue was induced after prolonged singing. The vibration pattern was quantified into ratio indices and time based quotients using the HSVP program.

**Reliability of the data processing**

The accuracy of the measurements analyzed by HSVP program depended on the precision of the extraction procedures by the examiner. Results from Table 2 indicated that there would be discrepancy in judging and selecting the frames containing the most stable 100 vibratory cycles among the 8000 frames raw video between different raters. Statistical results in comparing the final extracted measurements revealed that there was no significant difference (see Table 3) despite of the discrepancy in frame selection.

Apart from frame selection, other potential source of errors included adjustment of contrast, zooming, rotation and delineation of the glottis. Table 3 again indicated that the results of the extracted indices conducted by the two raters were not significantly different. Nevertheless, it was recommended to develop a standard analysis protocol for the HSVP program in the future, which should contain a standard reference and manual for each procedure including the recommended zooming size, contrast and landmarks of the glottis.
Self rating on vocal condition in vocal fatigue state

After a mean singing time of 103.8 minutes (s.d. = 7.2 minutes, range = 95-115 minutes), all of the participants were reported to be vocally fatigue. The participants self rated their vocal condition before and after the singing task using an 11-point rating scale (0-10) on level of discomfort in the throat, dryness and voicing effort. The participants rated the three dimensions of their vocal condition significantly higher (i.e. more severe) after the singing task (see Table 4). Some of the participants reported an increase in harshness and pitch breaks in addition. It revealed that the participants in this study demonstrated the common signs of vocal fatigue as reported in Hunter and Titze (2009) and Stemple et al. (1995). Hunter and Titze suggested that laryngeal tissue fatigue would cause tissue changes such as increase in viscosity. The tissue changes would result sensation of pain and discomfort, change the vibratory pattern and shift the phonation threshold pressure. More efforts would be needed to sustain voicing, and voice breaks would be easily resulted. These changes occurred in the participants in this study as reflected by their self rating. The singing task was found to be effective in inducing vocal fatigue. The observable changes in vocal fold vibration recorded by high speed imaging could be attributed to the vocal fatigue state.

The results also agreed with the study by Yiu and Chan (2003) who found that prolong vocal loading task without hydration would lead to negative vocal function changes. One of the participants was found to develop vocal polyp after the singing task (see Appendix B; Subject 13). These highlighted the importance for the role of hydration and vocal rest in voice protection as mentioned in Yiu and Chan (2003).

Quantitative analysis of high speed imaging data between pre and post singing task

*Ratio indices of the area, length and width of glottis*

In the present study, the vibration pattern of the vocal fold before and after a singing task was recorded using high speed imaging technique. The data was quantified into eight
ratio indices using HSVP program for comparison. Results indicated that the posterior glottal length ratio index and the glottal length to width ratio index significantly decreased after the prolonged singing task (see Table 5). The posterior glottal length ratio index represented the posterior halves of the glottal length measured from the horizontal midline of the glottis to the posterior glottal edge. As it was a ratio of glottal length in the minimum opening over the maximum opening, a decrease in the index should reflect a longer posterior glottal length in the open phase. It suggested that the posterior part of the vocal folds was stretched longer and the tension increased in the fatigued state. However, stretching of the vocal folds in fatigued voice was somewhat contradictory to the findings of Stemple et al. (1995). Stemple et al. revealed an increase in presence of incomplete glottal closure after two hours loud reading. They hypothesized that the incomplete glottal closure was due to thyroarytenoid muscle weakness, causing bowing at the edge of the vocal folds. In the present study, glottal chink was noticed in the participants in both pre and post-singing data (see Appendix B). However, the quantitative data did not demonstrate any change in incomplete glottal closure (i.e. glottal area in terms of pixels) after the singing task as the mean difference of the pre and post glottal area ratio index was statistically insignificant (see Table 5). The overall quantitative result reflected a stretching instead of weakening of the vocal folds.

It was believed that such hyperfunctional adjustment in vocal folds was a compensatory behavior secondary to the vocal fatigue state. In the post-singing endoscopy task of the present study, the participants were encouraged to produce vowel /i/ with the pitch and loudness level matched with the pre-task trials. At this point the vocal fold muscle was being weak and fatigued. The participants might need to employ more effort in voicing in order to maintain a pitch-matched and loudness-matched phonation. Increasing the medial compression of the glottis would require elongation of the fatigued vocal folds. In a single case study by Boucher and Ayad (2010), the muscular activities of lateral cricoarytenoid, thyroarytenoid and cricothyroid muscles were monitored along a loud reading task using
electromyographic and acoustic measurements. Boucher and Ayad reported that the activities of lateral cricoarytenoid muscles would decrease in fatigue state. At the same time the muscular activites of thyroarytenoid and cricothyroid muscles would increase to compensate for the decrease in activity in lateral cricoarytenoid muscles. Increase in thyroarytenoid and cricothyroid muscles served to tense and elongate the vocal folds in order to stabilize the adduction force. This fatigued point was termed as ‘critical fatigue’ (Boucher & Ayad, 2010).

Such observation was in line with the findings in the present study that the vocal folds were stretched after a vocal demanding task.

The glottal length to width ratio index was another evidence to support the stretching of the vocal fold in fatigued voice. The glottal length to width ratio index represented the shape of the glottis. The statistically significant decrease in the index after the singing task revealed that the shape of the glottis become longer and narrower during the open phase of the vocal fold vibration. In other words, the vocal folds was stretched and elongated, making the glottis thinner in shape. Such physiological observation agreed with the perceptual rating of strained voice reported by the teachers who complained with vocal fatigue (Gotaas & Starr, 1993). It was further believed that the hypothesized compensation behavior might attribute to the finding of an increase in phonation threshold pressure along two hours loud reading task in the study by Solomon and DiMattia (2000). Solomon and DiMattia suggested an increase in the vocal fold viscosity and stiffness after the vocal demanding task. Besides, the participants might over-stretch their fatigued vocal folds, leading to an increase in medial compression. As a result, the subglottal pressure for voicing increased.

Another reason for an increase in stretching and of the vocal fold might have been due to adaptation effect by the participants in the loud singing mode. Gelfer et al. (1996) found a significant increase in amplitude of vocal vibration after one hour of loud reading in untrained singers. In the study by Linville (1995), glottal closure was found to increase after 15 minutes loud reading. Both Gelfer et al. and Linville suggested that participants generalized the loud
reading phonatory mode to the post-task, resulting in the endoscopic observation of increase in vocal fold contact and greater amplitude of vocal fold excursion. In this study, the participants might have adopted a hypertensive mode so that the medial compression increased after the singing task.

It should be noted that there were no significant difference in the actual frequency and intensity level between the pre and post task condition (see Table 1) as these two factors were controlled in this study. Further studies such as investigating the change in pitch and loudness level, and comparing chronic vocal fatigue with the induced fatigue were recommended. They might be able to determine whether such change in glottal configuration was based on a compensatory behavior or simply an adaptation of the loud voicing mode.

Symmetry of vocal fold vibration

Comparison was made between left and right glottal area and glottal width ratio in attempt to check for asymmetry vibration. No statistically significant changes were revealed for both of the indices after the vocal demanding task (see Table 6 and 7). It suggested that asymmetry might not be a significant sign of vocal fatigue. Eustace et al. (1996) found a mildly asymmetrical closure for the participants who complained with chronic laryngeal fatigue. Studies investigating induced fatigue rarely found asymmetry of vocal fold after the vocal demanding task (Gelfer, et al., 1996; Solomon & DiMattia, 2000; Stemple et al., 1995). One might conclude that vocal fatigue induced in a short time did not have significant impact on the symmetry of vocal fold vibration, while long term fatigue might contribute to asymmetrical vibration, which could be considered as an early sign of dysphonia.

Temporal measurements of vocal fold vibration

The fundamental frequency, open quotient and speed quotient were three temporal measurements that determined changes in vibratory waveform after the vocal demanding task.
Results indicated that there was no statistically significant difference in the fundamental frequency calculated by the HSVP program (see Table 8). This was expected as the pitch level of the pre and post trial pairs for each participant were matched beforehand. In addition, there was no significant difference between the fundamental frequency calculated from the HSVP program and from the acoustic speech signal (see Table 9). It shown that the HSVP program was able to check the fundamental frequency accurately based on the vibratory pattern in the video data.

There was no significant difference on the open quotient and the speed quotient between pre and post singing task. It suggested that vocal fatigue did not exert a great impact on the vibratory pattern in terms of the glottal phase. However, this result was different from that reported by Lauri, Alku, Vilkman, Sala and Sihvo (1997). They found an increase in speed quotient and decrease in closing quotient in female participants after a vocal loading task derived from electroglottography (EGG). They hypothesized that these changes were due to an increase in adductory force, reflecting a hyperfunctional vocal adjustment. It should be noted that the temporal parameters like open quotient generated from high speed imaging did not necessarily correlate with that obtained using EGG (Echternach et al., 2010). Further studies were required to determine the agreement of the result of the open quotient and speed quotient generated from different methods on vocal fatigued voice.

Limitation

The results in this study should be interpreted with some limitations in mind. The proposed 11 measurements were just some of the many parameters measuring the glottal configuration. The pitch and loudness level were controlled in this study, and the voice onset and offset were excluded in the examination. Therefore, the effect of vocal fatigue on one’s pitch and loudness level, and the pattern of voice onset and offset could not be determined. This study might not give a comprehensive picture on the fatigued vocal vibration pattern.
is recommended to further examine the other parameters such as visual subjective rating on glottal configuration, voice onset and offset using the high speed imaging technique. In addition, high speed imaging could accompany with other instrumental measurements such as phonation threshold pressure and EGG, which should yield a more comprehensive findings in the topic of vocal fatigue.

High speed imaging technique is a powerful tool in investigating the vocal vibration pattern. However, the rigid endoscopy limits the phonation to vowels. In carrying out the endoscopy, the view of the vocal folds could be deflected due to tongue traction or tilting of the endoscope. Any minor deflection could alter the data as the analyzed unit by the HSVP program was in pixels. It is hoped that with innovation on the technology, higher resolution and the use of flexible endoscope for high speed recording will be available in future.

There is room for the HSVP program to improve. Firstly, color image interpretation should be added as the color video might give more details on glottal changes. Secondly, it was found that mucus and digital noises would affect the outlining of the shape of the glottis during measurement extraction. The binarization of the pixels should be modified for better capturing of the glottis. Any unnecessary digit noise should be excluded in glottis delineation. This could be done by modifying the image detection algorithm used in the HSVP program such as one proposed by Zhang, Bieging, Tsui and Jiang (2010). Their proposed method demonstrated better effectiveness in glottis delineation with less digital noise when compared with the use of other image detection algorithm.

**Clinical implication**

Vocal fatigue could be treated as one of the symptoms of voice disorders (Colton, Casper & Leonard, 2006). Early identification on vocal fatigue, especially for those experiencing chronic fatigue would help prevent from developing to various voice disorders. Therefore, a norm for the 11 parameters in this study is necessary. In this study, the use of
high speed imaging technique with quantitative analysis was found to have the potential in identifying signs of fatigued voice. They might not be easily detected by means of subjective measurement. However, the sample size in this study was too small. A large scale data collection in establishing a norm database might make the identification of abnormal vibration pattern more effective and efficient. High speed imaging could be a useful supplementary diagnostic instrument in the future.

**Conclusion**

The present study demonstrated that quantitative analysis on high speed imaging data by using the HSVP program was able to detect the difference on vocal vibration pattern between normal and induced vocal fatigue state. It was believed that the participants might develop a compensatory behavior in the vocal fatigue state. This preliminary conclusion was based on the statistical significant differences between pre and post singing tasks in two ratio indices that measured glottal length and glottal shape. Further studies using multiparameter measurements were required for a better picture on physiological changes in vocal fold under vocal fatigue condition.

**Acknowledgement**

I would like to express my sincere gratitude to my supervisor, Professor Edwin Yiu for his support on this paper. I would like to thank Dr. Karen Chan and Miki Shek for their help in conducting endoscopy, and to Dr. Gaowu Wang of the Peking University for his help in upgrading the HSVP program. Thanks to my colleagues, especially to Rosanna Lee, Ceci Lee and Eric Lau for providing the karaoke instruments, and Morris Poon as being the inter-rater in the reliability analysis.


Appendix A1: Participant’s self rating questionnaire (Chinese version)

姓名:

日期:

時間:

時段: 唱歌前 / 唱歌後

請根據閣下現時的狀況，圈出最適當答案。

1. 你/妳喉嚨感覺不適嗎? (如感到疼痛, 痘癢, 或疲軟等)

   正常 0 1 2 3 4 5 6 7 8 9 10 非常不適

2. 你/妳喉嚨感覺乾涸嗎?

   正常 0 1 2 3 4 5 6 7 8 9 10 非常乾涸

3. 你/妳發聲時比平日更吃力嗎? (即需要出力才能發出聲音)

   正常 0 1 2 3 4 5 6 7 8 9 10 非常吃力

4. 其他感覺, 如有:
Appendix A2 (English version)

Name:

Date:

Time:

Period: before singing / after singing

According to your current situation, circle the most appropriate answer.

1. Do you feel discomfort in your throat? (Such as feeling painful, itching, tired, etc.)

   Normal 0   1   2   3   4   5   6   7   8   9   10 Extreme discomfort

2. Do you feel dryness in your throat?

   Normal 0   1   2   3   4   5   6   7   8   9   10 Extreme dryness

3. Do you require more effort in voicing than in a normal day? (ie. it becomes harder to make sounds)

   Normal 0   1   2   3   4   5   6   7   8   9   10 Very effortful

4. Other feelings, if any:
Appendix B: Montages of the pre and post-singing video data pairs for all participants

Subject 1: Pre Post

Subject 2: Pre Post
Subject 3: Pre Post

Subject 4: Pre Post

Subject 5: Pre Post