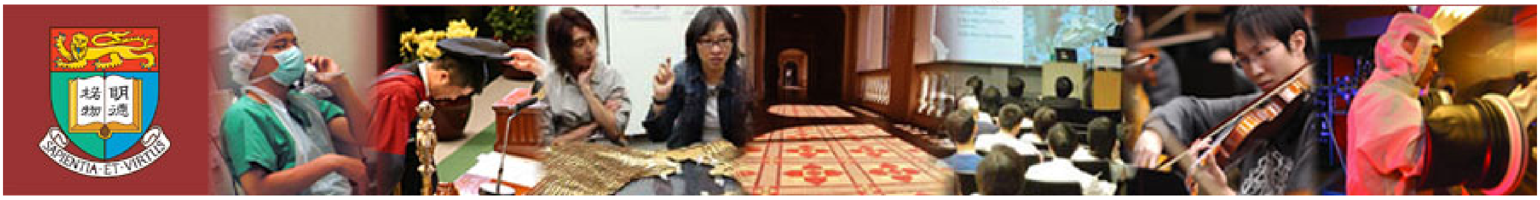

Title	**Long-term average spectral characteristics of different Cantonese opera singing styles**
Author(s)	**Wong, Sai; 黃璽**
Citation	
Issued Date	**2010**
URL	**http://hdl.handle.net/10722/173728**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

Long-Term Average Spectral Characteristics of Different Cantonese Opera Singing Styles

Wong, Sai Chryso

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, 2010.

Abstract

Cantonese Opera is a valuable cultural heritage populates in China. Basic singing styles consist of *zi hou*, *ping hou* and *da hou*. However, objective parameters measuring voice qualities in Cantonese opera singing are lacking. The current study examined the sound quality associated with *zi hou*, *ping hou* and *da hou* singing styles in comparison to conversational voice by means of Long-Term Averaged Spectra (LTAS). Continuous singing and speech samples were obtained from professional Cantonese opera singers and naïve speakers of Cantonese. All singing and speech samples were digitized at 44 kHz and 16 bits/ sample. Parameters including the first spectral peak (FSP), mean spectral energy (MSE), spectral tilt (ST) and high frequency energy (HFE) were derived from the LTAS contours by using Praat. Different singing styles exhibited different LTAS contours and were associated with significantly a higher ST value than conversational voice, implying a difference in resonance. Further investigation on the phonatory mechanism is indicated.

Introduction

As an important part of Cantonese culture originating from the Southern China, Cantonese opera is one of the major categories of Chinese opera with a history of at least 150 years (Liu & Huang, 2006). Among other valuable cultural heritage, Cantonese opera populates in the vicinity of the Guangdong and Guangxi provinces (RTHK & EDB, 2007). Though the origin of Cantonese opera is still in dispute (Li, 2009), it is generally agreed that it was imported from the Northern part of China during the Song (宋) and Yuan (元) dynasties (in late 13th century) and later prevailed in the Southern provinces (RTHK & EDB, 2007). Cantonese opera is characterized by the unique gestures and singing styles. According to Huang (1999), basic Cantonese opera singing styles of Cantonese opera include (1) *ping hou* (平喉), (2) *zi hou* (子喉), and (3) *da hou* (大喉), or *ba qiang* (霸腔). Although these singing styles differ in a number of aspects, different singing styles show significantly different voice qualities. For example, *ping hou* is commonly used by male characters known as xiao sheng (小生), and it often refers to a voice quality produced probably close to using the modal register. Meanwhile, *da hou* is used by male characters, wu sheng (武生), da hua mian (大花面) and er hua mian (二花面) with voices close to a modal register as in *ping hou*. However, it exhibits a pitch that is four to five intervals higher in musical scale (Chen, 2007). *Da hou* exhibits a throaty and rough singing voice representing heroic characters (Patzak, 1998) with aroused emotion in male characters (Chen, 2007). *Zi hou* is used by female characters known as hua dan (花旦) except lao dan (老旦) (RTHK & EDB, 2007), referring to a falsetto voice which possesses a voice fundamental frequency that is about an octave higher than *ping hou* (Huang, 1999). Physiologically, it is produced with constricted vocal folds and an elevated laryngeal tension.

Similar to many other traditional skills, the singing skills of Cantonese opera is acquired through a master-apprentice system; an apprentice learns the skills based on subjective comments given by the pedagogue, shi fu (師傅). The apprentice learns to sing correctly by exploring the best resonance voice via manipulating the respiratory, laryngeal, and articulatory systems, which is notoriously the most difficult part of learning (Liu, 2006). Apprentices have to learn to master the various skills essentially via trial-and-error, rendering an ineffective learning with a flat learning curve. To become proficient singers, years of training are required and progress varies among individuals (Guang dong yue qu ming jia, n. d.). Controversial views in learning to produce the best voice quality by different pedagogues have also been noted. According to Huang (1999), maximizing the pharyngeal space and relaxing laryngeal muscles may contribute to a good voice quality. Liu (2006) suggested using appropriate breathing method to achieve the best resonance voice. In singing at the falsetto register as in *zi hou* style, Chen (2007) suggested a traditional shouting method known as han sang (喊桑) which helps to expand the pitch range. Apprentices have to maximize their pitch with the help of gliding and breathing method (Liu, 2006). Laryngeal tension needs to be elevated during practice, yet, this adversely affects vocal health. During traditional Cantonese opera vocal training, pedagogues teach their apprentices how to sing by listening to the voices, based on which verbal feedback is provided. However, different pedagogues perceive voices differently based on their subjective feelings of how the voice is produced. This way of teaching is apparently unreliable and not scientific in helping the apprentice singers to acquire the most appropriate and correct voice quality for different singing styles.

Voice quality of operatic singing was defined acoustically by Sundberg (1987) as a peak of energy represented by the clustering of vowel formants F3, F4 and F5, known as singer’s formant, and which is a resonance phenomenon contributing to the ability to project the voice in a large theatre. A number of previous studies have examined the acoustic characteristics of singing voice quality. The study by Barnes, Davis, Oates, and Chapman (2004) recorded the voices of six professional operatic sopranos performing vowel and song tasks. The results were analyzed acoustically and good correlations between high spectral energy levels and performance level among operatic sopranos were reported. Mitchell and Kenny (2004) examined the effects of a singing technique known as open throat in female classical voices of six advanced singing students. Three different conditions: “optimal”, “sub-optimal” and “loud sub-optimal” were compared acoustically and finding indicated visual inspection of spectral change were more sensitive than conventional long-term average spectrum (LTAS) analysis to compare vocal timbre. Despite the various studies of classical singing in foreign countries, research on Cantonese opera and parameters to objectively quantify a good voice quality are still lacking. Understanding the vocal quality for different singing styles is useful in order for singing training to be effective.

Acoustic analysis provides an objective description of voice quality during singing. For example, proper singing resonance has been found to relate to the extra formant (energy), the singer’s formant, observed in spectrogram of the singer’s voice which indicates high spectral energy near 3 kHz (Sundberg, 1974). Formant frequency values directly indicate the location of tongue during singing (Ryalls, 1996). Based on these parameters, singers will be able to know their voice during singing. To study singing voice quality, one needs to examine the pattern and characteristics of vocal fold vibration. However, as articulators are continuously moving during singing, the effect of vocal tract resonances needs to be eliminated before vocal fold vibration can be examined. Traditionally, the effect of vocal tract resonance may be eliminated by inverse-filtering. Yet, this procedure cannot be done at the time of singing. Real time feedback therefore cannot be obtained. Another approach of removing the effect of vocal tract resonance is by means of long-term average spectrum

(LTAS) analysis. The procedure is a reliable approach for estimating the vibratory characteristics of the vocal folds during singing, particularly for analyzing running speech and singing with an extended recording sample as LTAS typically stabilizes after 30-40 seconds of continuous singing or speech samples (Cleveland, Sundberg, & Stone, 2001). Short-term variations due to the phonetic structures and filtering properties of the vocal tract can be averaged out and provide spectrum characteristics of the voice source (Löfqvist & Mandersson, 1987). Moreover, as an acoustic analysis tool, LTAS is non-invasive and safe. Parameters obtained from LTAS spectra reveal information on voice quality (timbre) characterized by formant frequencies and which is relevant to the perception of voice (White, 2001). Thus, LTAS is a suitable tool for comparing the voice quality among different singing styles and normal conversational voices.

A large number of investigations in comparing the acoustical differences of singing styles and speaking voices have been carried out. Kovačić, Boersma, and Domitrović (2003) compared Klapa and Dozivački styles of folk singing voices and found differences in the spectral slope and energy concentration obtained from LTAS analysis. Similarly, Stone, Cleveland, Sundberg, and Prokop (2002) compared the aerodynamic and acoustical differences of operatic and Broadway singing styles with reference to speech characteristics. Broadway singing style was found to be more similar to speaking than operatic singing style. Cleveland et al. (2001) compared the resonance characteristics of speech and country singing by analyzing LTAS and found similar LTAS characteristics in speaking and singing.

However, despite the many studies of singing voices, none examined the unique singing styles of Cantonese opera singing. The present study investigated the inter-style differences between *ping hou*, *zi hou* and *da hou* in comparison with normal conversational voices. LTAS spectra associated with *ping hou*, *zi hou*, *da hou* were quantified and compared based on four different parameters: (1) first spectral peak (FSP), (2) mean spectral energy

(MSE), (3) spectral tilt (ST) and (4) high frequency energy (HFE). The FSP is the frequency value associated with the first amplitude peak across the LTAS display (Goberman & Robb, 1999). It is assumed to represent the average fundamental frequency (F0) across phonatory sample and corresponds to changes in subglottal pressure and vocal stiffness. The MSE is the average amplitude value relative to the maximum energy peak across the frequency range of 0 - 8000 Hz. Physically, it is thought to represent the constant properties of vocal source and correlate with the tension of the laryngeal musculature (Fuller & Horii, 1988). The ST, which is defined as the ratio of energy (sum of amplitudes) between 0 - 1 kHz and 1 - 5 kHz, represents how quickly the amplitudes of the harmonics decline and low ST corresponds to hyperadductional state of vocal folds. The HFE, which is defined as the sum of amplitudes across frequency range 5 kHz - 8 kHz, is associated with high frequency noise elements during phonation (Goberman & Robb, 1999).

The general research question for the present study was: *Are there any significant differences between different singing styles and in comparison with conversational voices, as judged from the LTAS parameters including FSP, MSE, ST and HFE?*

Method

Participants

Eleven professional singers of Cantonese opera were included in the present study. Five different singers were included in the three groups of different singing style. Singers were selected based on the following criteria: (1) They were judged to be experienced and well-qualified professional singers by the local Cantonese singing communities, and (2) their singing was recommended by experienced Cantonese opera pedagogues to be representative of the singing styles investigated in the study. Singers specialized in more than one singing styles were preferred as to minimize the effect due to individual variation in vocal tract configuration. Table 1 summarizes the demographic characteristics of the participants and the type of singing styles they specialized in.

For comparison, three female and two male naïve speakers of Cantonese were recruited and served as controls. Their age range was from 22 to 40 years old with a mean of 30 years. The participants were considered to be naïve speakers as they had no prior training in singing Cantonese opera. Naïve speakers were preferred as to ensure singing techniques in Cantonese opera singing were not present in the speech samples collected. All participants were reported to have no known history of speech and hearing problem and were in good vocal and physical condition at the time of recording.

Table 1

Summary of demographic characteristics and sing styles of the professional singers

Subject	Sex	Singing style
1	Female	*Zi hou*
2	Female	*Zi hou*
3	Female	*Zi hou*
4	Female	*Zi hou*
5	Female	*Zi hou*, *Ping hou*, *Da hou*
6	Male	*Ping hou*, *Da hou*
7	Male	*Ping hou*, *Da hou*
9	Female	*Ping hou*, *Da hou*
10	Male	*Ping hou*
11	Male	*Da hou*

Singing and speech materials

Based on the assumptions of LTAS, variations due to varying phonetic structure in singing or running speech can be averaged out (Löfqvist & Mandersson, 1987). Variations in speech and singing materials do not account for the variations in LTAS. Thus, standardization of the materials for singing and speech was not needed in the present study.

Design and procedures

The singing samples produced by professional singers were selected from commercial audio discs available. The extracted samples were classified into the three different singing styles. This was based on the musical notation of the pieces selected and assured by an experienced singer in Cantonese opera. Portions of only the vocal part were used for LTAS analysis. One *da hou* singing sample was recorded by a professional singer. The recordings obtained were all digitized at 44.1 kHz with 16 bits/sample quantization rate and stored in a computer for later LTAS analysis.

Singing samples lasting for at least one minute in duration and conversational speech samples with approximately four minutes in duration were obtained from professional singers and naïve speakers respectively. Participants were informed about the purpose of the study and a signed written consent form was obtained. Prior to the recording, participants were given a brief period of practice to familiar themselves with the recording environment and to prepare for the speech and singing task. Each singer was asked to sing a song he was familiar with using a particular singing style in a way approximating an actual performance. Naïve non-singers were asked to talk on one to two topics familiar to them at a comfortable loudness level by using a normal speaking rate for about four minutes. Conversational task was chosen to obtain a longer speech sample and to ensure fluency of speech.

Recording of conversational speech was carried out in a sound-treated booth. The recording of one *da hou* sample was obtained in a quiet setting. Speech and singing samples were recorded using a dual-channel audio recorder via a dynamic microphone. A mouth-to-microphone distance of approximately 10 cm was maintained throughout the recording. The recorded samples were digitized at 44,100 Hz or above and stored in computer for LTAS analysis.

Signal preparation

The LTAS analysis was done by using Praat (version 5.1.29) which is a signal analysis software. It allows digitization, editing, storage and analysis of the recordings. To proceed with data analysis, an amplitude-by-time waveform of each sample was displayed on the computer screen. Since the analysis was restricted to the voiced portions of the acoustic signals as to reveal vocal fold vibratory characteristics, voiceless portions and silent periods were excluded prior to LTAS analysis. Portions dominated by high frequency components were defined as voiceless (Löfqvist & Mandersson, 1987). Silent periods were defined by visible portions with no amplitude deviation away from the baseline (Goberman & Robb, 1999). The voiceless portions and silent periods were removed manually from the waveform using a pair of vertical cursors superimposed on the waveform. Subsequently, a waveform with only continuous voiced sample remained for analysis.

A normalization procedure was carried out before the LTAS analysis. This is necessary since sound pressure level (SPL) was one of the variables affecting the contour of spectrum particularly in frequency region above 2000 Hz (Mitchell & Kenny, 2004). The normalization was done using Praat and the mean SPL was scaled to 80dB for each sample by multiplying the amplitudes proportionally.

Analysis

The analysis was accomplished using a 1048-point analysis and the bandwidth was set to be 22 Hz. The LTAS display derived from fifteen singing samples and five speech samples was created. The result provided an amplitude-by-frequency representation of an average energy concentration across 0 - 8 kHz. Parameters including FSP, MSE, ST and HFE were obtained and calculated to quantify the contour. The average fundamental frequency (F0) was also obtained from the audio signals for the purpose of comparison.

Reliability

Editing of the samples was considered as a critical procedure in performing LTAS analysis (Goberman & Robb, 1999). Approximately 10% of the samples (at least 1 minute of recordings from each singing style and speech voice) were randomly selected for re-editing. Intra-judge reliability was examined by re-editing the selected voice sample. Duration of the original edited sample was compared with re-edited sample. Based on the comparison, intra-judge agreement was found to be 94%. Inter-judge agreement was found by comparing the original voice sample and re-edited voce sample by an outside judge. The inter-judge agreement was found to be 95%. Both values indicated that the editing done by the primary investigator was consistent and reliable.

Results

The overall average LTAS spectra associated with different singing styles and conversational voices are displayed in Figure 1. Visual inspection shows that the singing groups demonstrated a noticeable energy peak near 3000 Hz, while the control group shows a gradual falling contour. Furthermore, singing styles differ in the location of energy peaks. While *ping hou* voice showed an additional slight peak near 4000 Hz, *da hou* style showed multiple peaks near 4000 Hz, 5000 Hz and 6000 Hz respectively. Moreover, singing voices showed a relatively flat contour after a sharp rise in amplitude at low frequency range.

Values obtained from the four parameters were used as the main outcome measures. To investigate the overall differences between different singing styles and conversational voice, the non-parametric Kruskal-Wallis test was carried out for the parameters obtained. A non-parametric test was employed due to the small sample size in each group, which failed to comply with the normality assumption for any parametric statistical procedures. Prior to the analysis, Levene test of homogeneity of variance was done which yielded no significant difference for FSP, MSE, ST and HFE ($p > 0.05$). However, significant differences were found for the average fundamental frequency ($p > 0.05$). The results of the statistical analysis performed were summarized as follow:

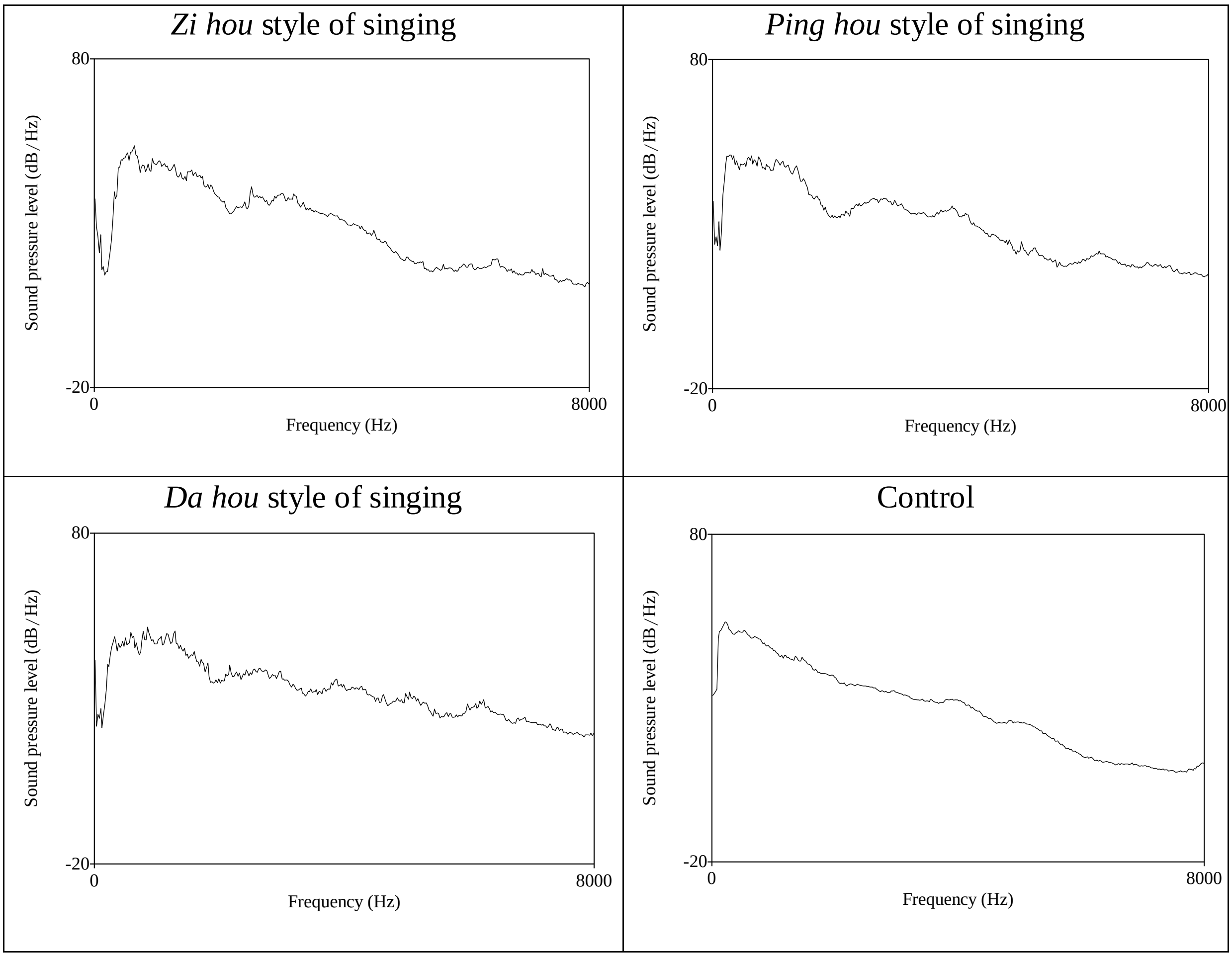

Figure 1. Average long-term average spectra (LTAS) associated with *zi hou*, *ping hou*, *da hou* and conversational speech.

First spectral peak (FSP)

Averaged FSP values associated with different singing styles and conversational speech are listed in Table 2. Significant main effect was found (χ^2 (3, 16) =11.178, $p < 0.05$). Homogeneity was assumed as no significant difference was revealed in the Levene test result. Tukey HSD post hoc tests were conducted to further evaluate the between-group differences. It was revealed that the *zi hou* style of singing showed significantly higher FSP than *ping hou* style of singing ($p < 0.05$). Both *zi hou* ($p < 0.01$) and *da hou* ($p < 0.05$) style of singing showed significantly higher value than the control group respectively. There were no significant differences in FSP values between the remaining groups.

Mean Spectral Energy (MSE)

Averaged MSE values associated with different singing styles and conversational speech are listed in Table 2. Kruskal-Wallis test yielded no significant differences between groups (χ^2 (3, 16) = 0.12, $p > 0.05$). Tukey HSD post hoc tests for multiple comparisons were carried out. No significant differences were revealed between different singing styles and the conversational voice ($p > 0.05$).

Table 2

Long term average spectrum (LTAS) characteristics for zi hou, ping hou, da hou styles of singing and control group. The table contains mean rank and range of the first spectral peak (FSP) and the mean spectral energy (MSE).

Styles of singing	FSP (Hz)[a]			MSE (dB)[b]		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Zi hou	459.8	45.62	385-495	-13.44	1.79	-16.29 – -11.36
Ping hou	297.0	67.81	231-385	-12.69	1.40	-13.70 – -10.49
Da hou	385.0	76.21	319-495	-12.82	1.72	-14.88 – -10.96
Naïve speaker	226.6	121.9	121-429	-13.17	1.31	-14.57 – -10.99

[a] FSP was expressed as the maximum amplitude in the frequency range 0 - 500Hz

[b] MSE was expressed as the average amplitude value (dB) relative to the maximum amplitude value across the entire frequency range (0 - 8000Hz).

Spectral Tilt (ST)

Average ST values associated with different singing styles and conversational speech are listed in Table 3. Kruskal-Wallis revealed main effect for the groups (χ^2 (3, 16) =11.8, $p <$ 0.05). Multiple comparisons between the groups were performed with Turkey HSD tests. Significant results were revealed between different styles of singing and the control group. *Zi hou*, *ping hou* and *da hou* style of singing demonstrated significantly lower ST value than the control group ($p < 0.05$). No significant between-group difference was found in different styles of singing.

High frequency energy (HFE)

Average HFE values exhibited by different singing styles and conversational speech are listed in Table 3. Kruskal-Wallis was performed to examine the overall differences in high frequency energy value across different singing styles and the conversational voice. The main effect was found to be insignificant (χ^2 (3, 16) =7.343, $p > 0.05$). Tukey post hoc test was performed for multiple comparisons between groups. *Da hou* style of singing exhibited significantly higher HFE value compared with conversation voice ($p < 0.05$) and other groups did not show significant differences.

Fundamental Frequency (F0)

The average F0 was obtained for singing groups and control group and the mean, standard deviation and range of F0 values are listed in table 4. Results of non-parametric test Kruskal-Wallis test indicated significant main effect (χ^2 (3, 16) =16.7, $p < 0.01$). Levene test revealed significant difference for the mean F0 value suggesting homogeneity of variance is not assumed. Thus, Tamhane post hoc test was done for multiple comparisons. *Zi hou* style of singing showed significantly higher value than *ping hou* style of singing ($p < 0.01$) and the control group ($p < 0.05$). *Da hou* style of singing showed significantly higher F0 than *ping* *hou* style ($p < 0.05$) and conversational voice ($p < 0.05$). *Ping hou* style of singing did not show significant differences with the conversational voices.

Table 3

Long term average spectrum (LTAS) characteristics for zi hou, ping hou, da hou styles of singing and control group. The spectral characteristics include the spectral tilt (ST) and the high frequency energy (HFE).

Styles of singing	ST [a]			HFE (dB)[b]		
	Mean	Standard Deviation (10^{-3})	Range	Mean	Standard deviation	Range
Zi hou	0.29	13.7	0.27-0.31	45215	10644	27202-52741
Ping hou	0.31	16.5	0.29-0.33	48041	17070	25510-67194
Da hou	0.29	16.5	0.27-0.31	68505	16287	52543-92552
Naïve speaker	0.35	22.8	0.32-0.37	42771	92953	30874-55588

[a] ST was expressed as Σ amplitude value (dB) $_{0 - 1000\ Hz}$/ Σ amplitude value (dB) $_{1000 - 5000\ Hz}$

[b] HFE was expressed as Σ amplitude value (dB) $_{5000 - 8000\ Hz}$

Table 4

Average fundamental frequency value for zi hou, ping hou, da hou styles of singing and control group.

Styles of singing	Fundamental frequency (Hz)		
	Mean (Hz)	Standard Deviation	Range (Hz)
Zi hou	353	6.05	345-361
Ping hou	264	17.8	241-290
Da hou	324	27.1	281-353
Naïve speaker	183	55.4	120-243

Discussion

In the present study, LTAS spectra analysis was used to describe the average voice quality and the contour reflects the voice source characteristics associated with different singing styles of Cantonese opera. Since only a relatively small sample size was used for comparison between groups, generalization based on the results should be made with caution. In general, the LTAS spectra showed a typical downward sloping contour, with amplitude diminishes as the frequency increases (see Figure 1). This gradual decrease in magnitude is typical in voicing using modal register (Klatt, 1980). Upon visual inspection of the LTAS contour, singing voices show relatively smaller amplitude attenuation than conversational voices and *da hou* style showed the least attenuation across the entire frequency range among the three Cantonese opera singing styles. This observation is consistent with the calculated LTAS parameters including ST and HFE, and the relevant physiological implications which are discussed below.

A further examination of the LTAS contours reveals different location of energy peaks compared among the groups. This may imply that the difference in resonance characteristics of vocal tract between groups. During the production of sounds, glottal waves originated from the vocal fold vibration get propagated along the vocal tract. The physiological configuration of the vocal tract suppresses the transfer of sound energy at certain frequency and maximizes energy at other frequencies (Losco, 2007). Previous studies investigating the LTAS curves compared different types of voices found presence of formants (or energy peaks) at various frequencies. In the investigation by Cleveland et al. (2001), classically trained singers produced a strong spectral peak at about 2.8 kHz followed by the steep fall-off in the upper part of spectrum known as singer's formant, while speaking voices showed only a slight peak between 3 kHz and 4 kHz known as speaker's formant. Similar speaker's formant was found in Croatian folk singing voices (Kovačić et al., 2003). The

LTAS contour of singing voices in Cantonese opera more likely resembles the speaker's formant as a relatively weak amplitude level was observed near 3 kHz rather than a prominent strong peak. Nevertheless, despite the similarity presence of speaker's formant, singing voices in Cantonese opera may exhibit a unique resonance pattern. As can be seen in Figure 1, a relatively concentrated energy which resulted in a flat contour was noted at the low frequency range. This was found to be distinguishable from the country and classical singers (Cleveland et al., 2001).

On the other hand, the LTAS contour was found to vary among the three different singing styles of Cantonese opera. This represents differences in the voice quality associated with different resonance effects. From the subjective view of a singing pedagogue, Liu (2006) suggested that different singing styles used different parts of the body as the resonator. For example, while *ping hou* singers should feel major vibration in the chest, *zi hou* singers should feel vibration more above the soft palate. It may be possible to distinguish different singing styles based on the location of peaks present in the LTAS contour. Yet, as mentioned earlier, the LTAS pattern obtained from small number of subjects may not be representative of the general pattern in Cantonese opera singing. From the data summarized in Tables 1 - 4, a relatively large range of values was indicated in some LTAS parameters. Thus, variations in individual spectra may be present and the curve produced from averaged data may not adequately represent the unique quality associated with different singing styles. Comparisons should be made with cautions. Future studies with a larger pool of samples will allow individuality of spectra be further investigated. To determine if the shape of LTAS contour can objectively justify the voice quality, correlation between the LTAS contour and the subjective rating of singing voice quality can be further investigated to establish the relationship between contour pattern and the perceptual singing standard.

First Spectra Peak (FSP)

The present data indicated that both *zi hou* and *da hou* singing voices were associated with a significantly higher FSP value conversational voice and only *zi hou* singing voice revealed a significantly higher FSP value than *ping hou* singing voice. As suggested by Goberman and Robb (2005), the FSP value is assumed to relate to the rate of vibration of vocal folds during singing and speaking, and it is supposed to represent the fundamental frequency (F0) across the sound samples. Thus, *zi hou* singing voice is assumed to phonate at the highest F0. This is consistent with the definition from Huang (1999) which stated that *zi hou* singing voices exhibited F0 generally with, perceptually, an octave higher than *ping hou* singing voices. Yet, FSP value obtained is inconsistent with the definition of *da hou* singing voices which should exhibit pitch four to five intervals higher in musical scale than *ping hou* singing voices (Huang, 1999). Direct comparison of FSP and F0 values from each group revealed relatively greater variations in FSP values obtained (see Figure 2). Different results were obtained from the statistical analysis in comparison among the groups using FSP and F0. Though similar results were obtained in comparison of *zi hou* singing voice among the groups, *da hou* singing voice differed significantly in comparison to the *ping hou* singing group only as regard to the F0 value. Thus, F0 obtained directly from the audio wave spectra correlates better with the definition of singing voices by Huang (1999). Alternatively, *ping hou* singing voices demonstrated insignificant FSP and F0 values when comparing across the groups. This is probable as *ping hou* singing generally consists of musical notes ranges from G2 (98Hz) to F4 (349Hz) in the low-mid frequency range of singing voice (Huang, 1999) and conversational voice consists of F0 ranges from 120 Hz - 243 Hz (see Table 4). The two groups yielded comparable average F0 values.

Based on the above discussion, it may be more reliable to use the average F0 value across the entire voice sample rather than the FSP value obtained from LTAS to compare the pitch of different styles of singing voices. However, obtaining average F0 value or the FSP value depends on the music material. Perhaps, future acoustical study of Cantonese opera singing voices should control the musical material. Analysis should be based on the same singing materials performed with different singing styles.

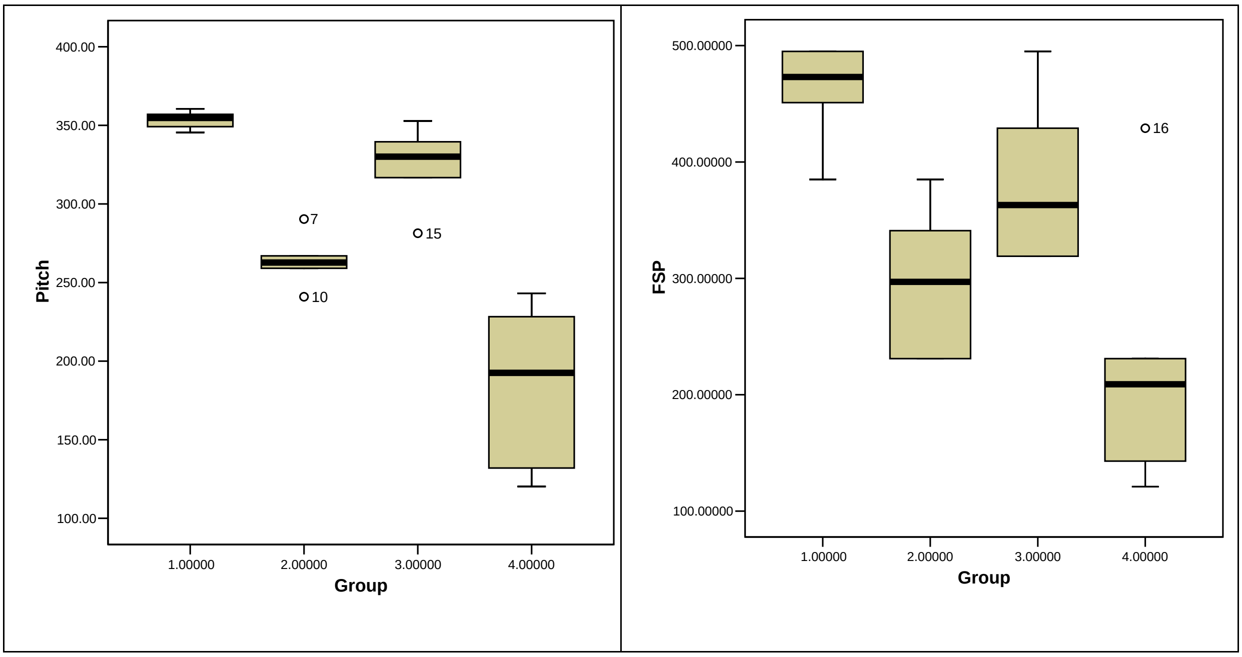

Figure 2. Box plot of fundamental frequency (left) and First Spectra Peak (right) for subjects in groups of (1) *Zi hou* (2) *Ping hou* (3) *Da hou* (4) Control.

Mean Spectra Energy (MSE)

The average amplitude (in dB) across the frequency range 0 - 8 kHz relatively to the maximum amplitude peak across the frequency range was obtained from the LTAS. In the study of Fuller and Horii (1988) examining infant cries, tense cry quality was found to associate with reduced spectral energy variation. Greater (less negative) MSE values indicate less variation in spectral energy corresponds to increased tension in voice quality. In the present study, however, insignificant difference was obtained between the groups indicating comparable magnitude of tension present in the vocal source. This is contrary to the normal physiology in pitch variation. Colton, Casper, and Leonard (2006) stated that vocal fold tension is positively correlated with the vibrating frequency. From the analysis, *zi hou* and *da hou* style of singing exhibited a significantly higher F0 than *ping hou* style of singing. Deducted from this relationship, greater tension should be present in *zi hou* and *da hou* style of voicing. Apart from that, *da hou* singing voices were considered to associate with throaty voice and may associate with high subglottal pressure and glottal hypertension (Laukkanen, Sundberg, & Björkner, 2004). This is not consistent with the physiological implications of the MSE value obtained from the current study. Additional explanation may account for the comparable glottal source tension in singing and conversational voices.

Resonance phenomenon which is supposed to be essential in Cantonese opera singing may account for the reduced vocal fold tension in singing. Huang (1999) emphasized easy phonation when singing with high F0 in both *zi hou* and *da hou* styles. Liu (n. d.) claimed appropriate breathing and phonation methods are effective in avoiding vocal abuse. Desirable phonation method includes “mian zhao chang fa (面罩唱法)” which suggests sensory perception of voice around the facial region. This aims at generating a soft and relaxed voice quality. Comments made by Cantonese opera pedagogue Huang (1999) and Liu (n. d.) are indeed consistent with the resonant voice proposed in literature. Titze (2001) defined resonant voice as the voice production that is easy to phonate with vibratory sensation in the face. Physiologically, the resonant voice employed a barely abducted vocal fold gesture. This glottal configuration is demonstrated to maximize the ratio of vocal output intensity to the mechanical stress of vocal folds during impact (Berry et al., 2001). In this way, production of voice can be achieved with the least effort. The tension in the vocal fold may be minimized.

Comparable MSE value, which relates to tension of vocal source, obtained in the current study may be attributed to the resonant singing technique. However, learning to acquire the best resonance voice is formidable because of individuality of structures and tissues (Liu, 2006). Further studies examining the location of resonators in achieving the best resonant voice quality may give insight to the singing training in Cantonese opera. Titze (2001) interpreted the sensation around the face as an indication of effective conversion of aerodynamic energy to acoustic energy instead of utilizing the facial part as a resonator. Resonance is indeed considered as reinforcement between vocal cord vibration and supraglottal acoustic pressure. Future studies may be focused on examining the supra-glottal laryngeal tract configuration and the laryngeal muscles involved using techniques such as Magnetic Resonance Imaging (MRI). Clinically, MSE value may be a good indicator of potential vocal trauma. Professional singers are prone to vocal abuse with inappropriate phonation particularly in achieving high pitch. This may relate to the increased stress of impact as impact stress on vocal folds is an important leading cause for vocal nodules (Colton et al., 2006). It may be possible for singers to make an intra-subject comparison of the MSE value. Significantly greater MSE value may indicate potential vocal abuse. Perhaps, this can be further investigated by intra-group comparison of singing and speaking voices in trained and non-trained Cantonese opera singers.

Spectral Tilt (ST)

The spectral tilt is measured as the ratio of 0 - 1 kHz to 1 kHz - 5 kHz and can be taken as the comparison between the low frequency energy and high frequency energy (Goberman & Robb, 2005). Large ST value indicates rapid fall of spectrum and energy of fundamental frequency and lower harmonics dominates. Alternatively, small ST value indicates slower drop of the spectrum. The control group exhibited the greatest ST value while the other groups showed comparable ST values. This is consistent with the slopes of contours depicted in Figure 1. The LTAS spectra of singing voices exhibited a more gentle sloping with frequency than conversational voices. At the physiological level, Löfqvist and Mandersson (1987) suggested that differences in ST value were distinguished by the degree of glottal adduction and low ST value implied a hyperadducted state of vocal folds.

Colton et al. (2006) claims that the vocal fold length decreases with increase in F0 when phonating in falsetto register as in *zi hou* singing voice. Tension builds up in the vocal ligaments can probably explain this phenomenon. Incomplete glottal closure is a typical characteristic of falsetto register found in untrained singers. However, complete glottal closure was present in falsetto voices of professional singers disregard the variation in loudness (Sundberg & Högset, 2001). In order to maintain elevated pitch level but avoid breathy phonation which is undesirable in singing, singers may exhibit a more hyperfunctional behavior in comparison to the conversational voice. This may partially account for the low ST value in *zi hou* voice. Alternatively, modal register dominates in *ping hou* and *da hou* singing. Tension exhibits in the vocalis muscles when phonating in modal register with elevated pitch. Vocal folds are stretched and elongated by the actions of intrinsic muscles of larynx (Colton et al., 2006).

Nonetheless, from the result of LTAS, though *ping hou* style of singing demonstrated significantly lower ST value in comparison to the conversational speech, the F0 value was comparable to the conversational speech. This may imply comparable vocal tract configuration. Thus, the lower ST value generally appeared in Cantonese opera singing cannot solely attribute to the hyperfunctional vocal behavior. This hypothesis also does not correspond to the subjective perspective of experienced Cantonese opera pedagogue Liu (2006), whom suggested that relaxing vocal folds is essential in achieving good voice quality in singing. Previous studies (e.g., Niimi, 1996) examined laryngeal gestures of traditional Japanese singing supported this argument. The laryngeal gestures were observed through fiberscopic examination during singing performance. Though traditional Japanese singing was perceived to be strained and pressed, physiological signs suggesting hyperfunction of larynx including hyperadduction of vocal folds and tight glottal closure were not observed. Inferring from the study, hyperfunctional vocal behavior may not be present during the singing in Cantonese opera. However, direct comparison across singing styles should be made with caution. Further research may examine the laryngeal gestures physiologically in Cantonese opera singing during performance to test this hypothesis.

The difference in ST value was in agreement with previous studies. Watts, Barnes-Burroughs, Estis, and Blanton (2006) reported that talented singers demonstrated lower ST compared with the non-talented singers. Singing power ratio (SPR), which is calculated by measuring peak intensities between 2 - 4 kHz and 0 - 2 kHz, represents influence of resonance in supra-glottal vocal tract. From Watts et al. (2006), agreement was found between ST value and SPR value. Thus, low ST value is reliable to represent the effect of resonance discussed earlier. From the comparison of ST value among different singing styles, no significant differences were observed. This may imply comparable importance of resonance in different styles of Cantonese opera singing.

High Frequency Energy (HFE)

High frequency energy measures the sum of energy between 5000 Hz and 8000 Hz. The parameter is likely to correlate with the slope of the spectrum (Löfqvist & Mandersson, 1987). When the slope of source spectrum drops less rapidly, there is greater energy present in the high frequency range. Thus, high frequency energy may partially correlate to the use of hyperfunctional voice as in the physiological representation of spectral tilt value. In the present study, *da hou* singing style exhibited a shallow ST value in combination with higher HFE value when compared with the conversational voice. This may represent a relatively elevated hyperfunctional vocal behavior, probably because *da hou* singing was used to express an aroused emotion (Chen, 2007), and it is generally associated with a louder phonation. According to Colton et al. (2006), one needs to increase the resistance of laryngeal valve for increasing loudness level. The vocal folds have to be adducted strongly which leads to an increased medial compression. This forceful medial compression of the vocal cords is likely to end up in a hyperfunctional state.

Still, higher HFE value may represent the presence of noise component (Löfqvist & Mandersson, 1987) during singing. *Da hou* style was perceived to render a rough and throaty quality (Patzak, 1998). Roughness in voice may indeed acoustically correlates with the perturbation and noisiness of the spectrum. On the other hand, phonatory hyperfunction is perceived to be throaty and pressed (Laukkanen, Sundberg, & Björkner, 2004). The LTAS measurements may not necessarily differentiate the periodic or aperiodic components in the voice source.

Limitation of the Study

One of the limitations with the data derived from LTAS is that they were obtained from commercial audio recordings. The recording environment and instrument were not standardized for objective comparison. Extraneous variables may exist causing potential source of error. Furthermore, loudness affects the amplitude across spectrum (Mitchell & Kenny, 2004). In the present study, the process of normalization used to eliminate the effect of loudness variation is rarely adopted in previous studies. For instance, Cleveland et al. (2001) normalized LTAS contour with respect to the highest spectrum level instead. Therefore, validity in making comparisons across studies may be questionable. Selection of identical singing materials for recording across different singer groups may help minimize the problem in loudness variation in the future.

Conclusion

An array of LTAS parameters including FSP, MSE, ST and HFE values were derived from the LTAS contours for comparison between different singing styles *zi hou*, *ping hou* and *da hou* in Cantonese opera and the conversational voice. Upon visual inspection, singing voices showed slower attenuation across the entire frequency range. This could be reflected by the ST value which was likely to be caused by different resonant singing techniques. While the *zi hou* singing group showed higher FSP value than *ping hou* singing group, FSP values of *ping hou*, *da hou* singing groups were comparable. However, obtaining the fundamental frequency (F0) directly from the voice signals may be more reliable in differentiating the singing styles owing to the relatively smaller variation and finding was more consistent with the definition by pedagogue of Cantonese opera. The MSE values were comparable among the groups, indicating comparable glottal tension. It is hypothesized that glottal tension is minimized with the use of resonant voice. *Da hou* style of singing showed significantly higher HFE value than conversational voice, and this parameter is related to spectral slope. Higher value revealed in *da hou* singing may indicate hyperfunctional vocal behavior as well as the noise components in the voice source.

Acknowledge

I would like to express my sincere gratitude to Dr. Lawrence Ng for his kind guidance and assistance throughout the course of this dissertation. I would like to deliver my special thanks to Dean and Professor C. F. Chan for his support of the study and provision of information about Cantonese opera despite a very tight schedule, Principal H. K. Lee for his generous assistance in the recording of singing sample. Last but not least, I would like to thank all participants for the willingness to join the study and helpfulness in the reliability rating.

References

Barnes, J. J., Davis, P., Oates, J. & Chapman, J. (2004). The relationship between professional operatic soprano voice and high range spectral energy. *The Journal of the Acoustical Society of America*, *116*, 530-538.

Berry, D. A., Verdolini, K., Montequin, D. W., Hess, M. M., Chan, R. W. & Titze, I. R. (2001). A quantitative output-cost ratio in voice production. *Journal of Speech, Language, and Hearing Research*, 44, 29-37.

Chen, S. R. (2007). *Yue qu de xue he chang*. Xiang Gang: yue qu yin jiu ji wa. 粵曲的學和唱。香港: 粵劇研究計劃。

Cleveland, T. F., Sundberb, J. & Stone, R. E. (2001). Long-term-average spectrum characteristics of country singers during speaking and singing. *Journal of Voice*, *15*, 54-60.

Colton, R. H., Casper, J. K. & Leonard, R. (2006). Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Guang Dong yue qu ming jia. (n.d.). 廣東粵曲名家。Retrieved October 16, 2009, from http://www.chinaculture.org/gb/cn_zgwh/2004-06/28/content_50742.htm

Goberman A. M. & Robb, M. P. (1999) Acoustic examination of preterm and full-term infant cries: the long-time average spectrum. *Journal of Speech, Language and Hearing Research*, *42*, 850-861.

Goberman A. M. & Robb, M. P. (2005) Acoustic characteristics of crying in infantile laryngomalacia. *Logopedics Phoniatrics Vocology*, 30, 79-84.

Huang, S. X. (1999). *Yue qu ji ben zhi shi*. Xiang Gang: Chou pi jiang chu ban you xian gong si. 粵曲基本知識。香港: 臭皮匠出版有限公司。

Fuller, B., & Horii, Y. (1988). Spectral energy distributions in four types of infant vocalizations. *Journal of Communication disorders*, *21*,251-261.

Klatt, D. (1980). Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.

Kovačić, G., Boersma, P., Domitrović, H. (2003). Long-term average spectra in professional folk singing voices: a comparison of the Klapa and Dozivački styles. *In: Proceedings of Institute of Phonetic Sciences*, *25*, 53-64.

Li, R. X. (2008). Yue ju ju zhong yao su, shi bie biao zhi yu yue ju shi de zhen bie duang zhi. *Nan guo hong dou, 5*, 17-22. 粵劇劇種要素、識別標志與粵劇史的甄別斷制。南國紅豆。

Liu, Y.H. (2006). *Liu Yan Hua yue qu jiao cai xuan ji*. 劉豔華粵曲教材選集。 Retrieved September 24, 2009, from http://cofans.org/forum/viewthread.php?tid=401&extra=page%3D1

Liu, Y. H. (n.d.). *Yue qi yan chang ji qiao.* 粵曲演唱技巧。Retrieved Mar 22, 2010, from http://cofans.org/forum/viewthread.php?tid=36&extra=page%3D1

Liu, W.M. & Huang, H. (2006). Wen hua zai ren jian: yue qu. *Nan fang du shi bao.* 文化在人間: 粵曲。南方都市報。 Retrieved September 24, 2009, from http://www.gzlib.gov.cn/shequ_info/ndgz/NDGZDetail.do?id=18237

Löfqvist, A., & Mandersson, B. (1987). Long-time average spectrum of speech and voice signals. *Folia Phoniatrica et Logopaedica*, *21*, 221-219.

Loscos, A. (2007). Spectral processing of the singing voice. Retrieved Mar 22, 2010, from The Pompeu Fabra University, Department of Information and Communication Technologies Website: http://mtg.upf.edu/node/517

Laukkanen, A.M., Sundberg, J. & Björkner, E. (2004). Acoustic study of throaty voice quality. *TMH-QPSR, KTH*, 46, 13-23.

Mitchell, H. F. & Kenny, D. T. (2004). The effects of open throat technique on long term average spectra (LTAS) of female classical voices. *Logoped Phoniatr Vocal, 29*, 99-118.

Niimi, S. (1996). Laryngeal Gestures of traditional Japanese singing. *The Journal of Acoustical Society of America*, 100(4), 2743-2743. Abstract retrieved Mar 22, 2010, from Acoustical Society of America Digital Library database.

Patzak (2003). *Yue qu: da hou ling xiu xiong fei ying.* 粵曲:「大喉領袖」熊飛影。 Retrieved September 24, 2009, from http://www.bh2000.net/special/patzak/detail.php?id=1957

Radio Television Hong Kong (RTHK) & Education Bureau (EDB) (2007). *Yue ju shi chuang*。粵劇視窗。Retrieved October 16, 2009, from http://chineseopera.rthk.org.hk/cantonese/index.htm

Ryalls, J. (1996). A basic introduction to speech perception. San Diego, California: Singular Publishing Group.

Stone, R. E., Cleveland, T. F., Sundberg, J., & Prokop, J. (2002). Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in professional female singer. *TMH-QPSR. KTH, 43*, 17-29.

Sundberg, J. (1974). Articulatory interpretation of the ‘singing formant.’ *The Journal of the Acoustical Society of America, 55*, 838–844.

Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, Illinois: Northern Illinois University Press.

Sundberg, J. & Högset, C. (2001). Voice source differences between falsetto and model registers in counter tenors, tenors and baritones. *Logopedics Phoniatrics Vocology,* 26, 26-36.

Titze, I. R. (2001). Acoustic interpretation of resonant voice. *Journal of Voice*, 15(4), 519-528.

Watts, C., Barnes-Burroughs, K., Estis, J. & Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and non-talented singers. *Journal of Voice*, 20(1), 82-88.

White, P. (2001). Long-term average spectrum (LTAS) analysis of sex- and gender-related differences in children’s voices. *Logopedics Phoniatrics Vocology*, *26*, 97-101.