*** Title Page, including Author Details**

Long-term Average Spectral Characteristics of Cantonese Alaryngeal Speech

Manwa L. Ng[1], Hanjun Liu[2], Qin Zhao[3], Paul K. Y. Lam[4]

[1]Speech Science Laboratory
Division of Speech and Hearing Sciences
Faculty of Education
The University of Hong Kong
5/F Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong, SAR China

[2]Department of Communication Sciences and Disorders
Northwestern University
2240 Campus Drive
Evanston, IL. 60208, USA

[3]The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology
Xi'an Jiaotong University, Xi'an, 710049, P. R. China

[4]Hong Kong Voice and ENT Center
Room 601, 8 Wyndham Street
Central, Hong Kong, SAR China

Correspondence:
Manwa L. Ng
Speech Science Laboratory
Division of Speech and Hearing Sciences
Faculty of Education
The University of Hong Kong
5/F Prince Philip Dental Hospital
34 Hospital Road, Sai Ying Pun
Hong Kong, SAR, China

Telephone: +852-28590582
Fax: +852-25590060
Email: manwa@hku.hk

Long-term Average Spectral Characteristics of Cantonese Alaryngeal Speech

Introduction

With the larynx severed, laryngectomized patients need to adopt an alternative phonation method after total laryngectomy. To date, esophageal (SE), tracheoesophageal (TE), and electrolaryngeal (EL) speech, and use of a pneumatic artificial larynx (PA) are the most common alaryngeal phonation methods used in Hong Kong [1,2]. All of these alaryngeal phonation methods differ in how sound is created: both SE and TE speakers make use of the pharyngoesophageal (PE) segment as the new sound source (i.e., neoglottis) [3-6], while EL and PA speakers rely on an external device for sound generation [2].

The use of an alternative voicing method by laryngectomees inevitably changes the voice quality. A large number of studies have investigated the different sound characteristics associated with SE, TE, and EL phonation [cf., 2,6-21]. Yet, few studies focused on the sound characteristics of PA phonation, which was likely a result of its limited use in the North America and Europe [2,22,23]. These studies sought to describe the sound characteristics of different alaryngeal phonation through the examination of acoustic and aerodynamic parameters. Among the most frequently examined are the average fundamental frequency (F_0), F_0 range and perturbation, phonation intensity, vowel duration, and voice onset time. However, studying the voice source of laryngeal and alaryngeal speech has not been easy as it is always "contaminated" by the effect of vocal tract filter. According to the source-filter theory, voices of laryngeal (NL) speakers are products of the laryngeal and supralaryngeal resonance systems, both of which are independent of each other [24]. Accordingly, alaryngeal phonation is the product of the neoglottal sound source and the supra-neoglottal resonance system, determined by the vocal tract configuration. Despite the wide range of acoustic and aerodynamic studies of alaryngeal phonation, few studies critically examined the vibratory behavior of the voicing source of different alaryngeal phonation, despite the fact that such knowledge is of paramount importance in restoring postlaryngectomy verbal communication. To study the sound source of different alaryngeal phonation, the effect of vocal tract filter needs to be eliminated. This can be done by inverse-filtering the aerodynamic signals, or by means of long-term average spectral analysis of acoustic speech signals.

Analysis of long-term average spectra (LTAS) of speech offers a unique and reliable approach for estimating the vibratory characteristics of the alaryngeal sound source. Calculation of the LTAS involves excluding pauses and voiceless portions from a continuous sample of phonatory behavior and acoustically examining the remaining signal as discrete spectra derived from consecutive temporal intervals of phonatory activity [25]. By averaging these individual spectra, the LTAS levels out the short-time variations present in the human voice due to the filtering properties of the vocal tract [26]. Three common features extracted from the LTAS are first spectral peak (FSP), mean spectral energy (MSE), and spectral tilt (ST). The FSP is the frequency value associated with the first amplitude peak across the LTAS display. The FSP is assumed to provide a representation of the average F_0 across a phonatory sample [27]. The MSE is the average amplitude value across the frequency range of 0 – 8000 Hz. Physiologically, MSE is thought to represent the constant properties of the vocal source, as the LTAS averaging process eliminates any dynamic features induced by articulatory movement during vocalization [25]. The ST is a representation of how quickly the amplitudes of the harmonics decline, with a low ST corresponding to a hyperadductional phonatory state [27].

A number of previous studies have used LTAS to examine normal and disordered voice characteristics [cf., 27-36]. Despite the many LTAS studies examining voice production in various speaker groups, application of LTAS to examine alaryngeal phonation has been rare. Globek, Stajner-Katusic, Musura, Horga, and Liker [37] reported that the LTAS spectral timbres of SE and TE voices were similar. However quantitative information concerning the specific spectral features of SE and TE phonation were not reported. Weinberg, Horii, and Smith [38] examined LTAS in SE speakers and found a considerably lower average spectral level (7 – 10 dB) compared to NL phonation. Weiss, Yeni-Komshian, and Heinz [39] observed that the LTAS of EL phonation decreased in amplitude at approximately 500 Hz, while the LTAS of NL phonation began to decrease in amplitude at approximately 200 Hz. In addition, EL phonation in the frequency region of 2 – 4 kHz remained 5 – 10 dB higher compared to NL phonation.

Based on the above discussion, several drawbacks arise. (1) Results of past LTAS studies examining alaryngeal phonation are inconclusive and fail to provide a detailed depict of the underlying source characteristics of different alaryngeal speech. Past studies failed to quantify LTAS spectra of alaryngeal speech by using parameters such as FSP, MSE, ST of alaryngeal speech. Information drawn from these studies has been based on isolated and peculiar parameters derived from the source spectrum. (2) Perhaps due to the scarcity of PA speakers, comprehensive LTAS study of all kinds of alaryngeal speech in comparison with NL phonation is lacking. Such data would provide valuable clinical information concerning the similarities and differences across the various alaryngeal phonation methods, as well as NL speech. (3) Previous LTAS research only focused on English speakers. Information on how different alaryngeal speech of a tone language is not available. By examining the performance of alaryngeal speakers of a tone language, additional knowledge of alaryngeal speech characteristics with regard to the control of tonal variation will be obtained. Due to the relative inability in pitch manipulation, alaryngeal speakers of a tone language, especially those using EL and PA speech, may find it more problematic as pitch variation in EL and PA phonation is reportedly lacking.

In response to the drawbacks from previous studies, the present study examined alaryngeal speakers of Cantonese. In a tonal language such as Cantonese, tone (i.e., F_0

regulation) is primarily used to signal word-type [40]. The general research question for the present study was: *Do LTAS measures (FSP, MSE, ST) differ significantly among NL, SE, TE, EL and PA modes of phonation?* The LTAS spectra associated with NL, SE, TE, EL, and PA speech of Cantonese were examined, based on which LTAS parameters including FSP, MSE and ST were derived and compared.

Method

Participants

Sixty-three adult male native speakers of Cantonese consisting of 10 laryngeal and 53 alaryngeal speakers participated in the present study. The alaryngeal participants consisted of 15 SE, 12 TE, 15 EL, and 11 PA speakers. The TE speakers were all using the Provox-type valve, EL speakers using Servox electrolarynx, and PA speakers using a custom-made pneumatic device. The average age of each speaker group was 63 years, with participants ranging in age from 48 to 80 years. None of the participants had a reported history of speech and hearing problems, except for those problems associated with laryngectomy for the alaryngeal speakers. The alaryngeal speakers were selected from the New Voice Club of Hong Kong and rated as “good” or “excellent” speakers by a local speech-language pathologist. All speakers were literate and volunteered to participate in the present study.

Design and Procedure

The speech material included a 136-word passage selected from a third grade Chinese reading book, which was used in previous studies on Cantonese alaryngeal speech [cf., 2]. The recording was carried out in a sound-treated booth. Speech samples were recorded using a dual-channel audio recorder (Nakamichi, MR-2) via a dynamic microphone (Shure, SM58). During the recording, the microphone was placed at a distance of approximately 10 inches from the speaker’s mouth. To minimize the recording of stoma noise, the microphone was covered with a mesh screen.

Prior to the actual recording, the participants were given a brief period of practice to familiarize themselves with the recording environment and speech material. The participants were instructed to read aloud the reading passage using a normal speaking rate and comfortable loudness level. The speakers were provided with a card on which Chinese characters of the reading passage were printed. The recorded speech samples were then digitized at 20 kHz with 16 bits/sample quantization and stored in a computer for later LTAS analyses.

Data analysis

The LTAS analysis was performed according to the procedures established by Goberman and Robb [27], and Löfqvist and Mandersson [26]. The waveform of each participant’s speech sample was displayed on a computer using a signal analysis software (Praat, version 5.0.05). Since only voiced segments were used for the quantification of LTAS, silent periods were identified and removed from the speech samples prior to LTAS analysis. Silent periods were demarcated using a pair of vertical cursors superimposed on the waveform and subsequently removed, leaving a waveform of continuous phonation. The LTAS of the edited waveform was obtained by using a custom-made software program written in Matlab (45 ms Hamming window, LPC autocorrelation using 24 coefficients). To quantify the LTAS contours, FSP, MSE, and ST were extracted.

Reliability

Both intra-judge and inter-judge reliability measures were made for the derivation of LTAS. The editing of the original speech samples was viewed as the critical measurement procedure necessary to perform the LTAS calculations. Therefore, approximately 30% of the entire data corpus (20 speech samples) (at least three from each phonation type) were randomly selected from the recorded signals and re-edited. Based on comparison of the original edited speech sample durations to the re-edited speech sample durations resulted in an intra-judge agreement of 98%. Inter-judge agreement was performed by comparing the original measurements to those of a second investigator who was also experienced in acoustic analysis. Inter-judge agreement was found to be 96%.

Results

The overall average LTAS spectra associated with the NL, SE, TE, EL, and PA speaker groups are shown collectively in Figure 1. The FSP, MSE, and ST values were calculated from the LTAS spectra which are represented in Table 1.

First Spectral Peak

To evaluate whether FSP differed significantly between the speaker groups, a one-way analysis of variance (ANOVA) was performed. Significant main effect was found [$F(4, 58) = 11.209$, $p = 0.000$]. Tukey post-hoc t-tests revealed that PA group demonstrated significantly lower FSP than SE, TE, NL, and EL groups ($p < 0.001$). The remaining four groups did not differ significantly in regard to FSP values.

Mean Spectral Energy

Results of a one-way ANOVA indicated significant main effect for groups [$F(4, 58) = 311.370$, $p = 0.000$]. Tukey post-hoc t-tests revealed that PA speakers exhibited the greatest average MSE value (-47.79 dB) and SE speakers showed the smallest average MSE value (-18.64 dB) among the speaker groups ($p < 0.01$). There were no significant differences in MSE values between EL and NL groups.

Spectral Tilt

One-way ANOVA revealed significant main effect for groups [$F(4, 58) = 27.056$, $p = 0.000$]. Tukey post-hoc t-tests revealed that NL speakers showed significantly greater ST values than each of the four alaryngeal speaker groups ($p < 0.01$). The ST values of the four alaryngeal groups did not differ significantly.

(INSERT FIGURE 1 HERE)

(INSERT TABLE 1 HERE)

Discussion

In the study, LTAS spectra were used to average out the short-term dynamic features caused by articulatory movements, and a depict of sound source characteristics is obtained [25-26]. As can be seen in Figure 1, NL, SE, TE, EL, and PA speech demonstrated similar patterns of LTAS contours; all LTAS contours are downward sloping, with higher amplitude (energy) at lower frequency, and diminishing amplitudes as frequency increases. This amplitude attenuation with frequency was found in the LTAS spectra of all types of phonation methods. The trend that amplitude diminishes as frequency increases is, in fact, not uncommon in many mechanical vibratory systems.

Despite the similarity in amplitude attenuation, NL speakers exhibited the greatest attenuation than the other speaker groups (see Figure 1). According to Pickett [41], the glottal spectrum obtained from laryngeal excitation is found to have a roughly constant attenuation of -12 dB/octave. As frequency doubles, the amplitude of harmonics decreases by 12 dB. With the use of PE segment as the new sound source, SE and TE speakers of Cantonese exhibited more gentle downward sloping LTAS contours. This is likely to be due to the different voicing mechanism used in SE and TE speech from NL speech. Furthermore, a careful examination of the LTAS of SE and TE speech reveals a small amplitude peak at around 6 – 7 kHz, indicating an increase in energy around this frequency. The finding that the LTAS of SE and TE speech dropped more slowly with frequency than NL speech is consistent with that reported by Weinberg et al. [38], and Qi and Weinberg [42]. Such discrepancy in LTAS is believed to contribute to the unique sound quality of SE and TE phonation.

However, direct comparison of LTAS spectra of EL and PA speech with NL speech should be cautioned because of the difference in the way sound source is coupled with the vocal tract transfer function. In NL, SE and TE speech, the sound source is located at the deep end of the vocal tract. During speech production, sound is propagated from this end outward, and almost the entire vocal tract anterior to the sound source is used as the resonator. The posterior resonating tube is significantly shorter and contributes less to the resonance. However, in EL and PA speech, the coupling between sound source and vocal tract is different. In EL speech, sound generated by the mechanical vibration of the EL device was transcervically transmitted to the oral cavity, during which process the sound was unavoidably filtered and attenuated. In PA speech, sound generated from the pneumatic device was delivered into the oral cavity via a straw-shaped tube. The location at which the sound source in EL and PA speakers was coupled with the vocal tract filer was apparently different from that of NL, SE and TE speakers. The difference in coupling between sound source and the vocal tract in EL and PA speech may cause the vocal tract to resonate the sound in a way that was different from NL, SE and TE speech.

First Spectral Peak

Our data indicate that NL speakers exhibited an average FSP value of 232.00 Hz. As suggested by Goberman and Robb [27,30], FSP measured from the LTAS is assumed to correspond to the rate of vibration of the sound source. It follows that, in NL speakers, FSP should resemble how fast the vocal folds are vibrating during speech production. However, Cantonese is a tone language consisting of six lexical tones; the same syllable carries different meanings if produced at different lexical tones. When reading the passage, the participants uttered each syllable with a preset lexical tone. Therefore, the average F_0 value, or the FSP value, therefore largely depends on the lexical tones of the characters forming the passage.

NL, SE, and TE speakers exhibited comparable FSP values. This indicates that, despite the use of PE segment, the SE and TE speakers of Cantonese were able to produce an average F_0 value similar to NL speaker using vocal folds. This is in line with the findings obtained from Cantonese SE speakers reported previously [18]. Considering the fact that our SE and TE speakers were superior speakers selected by speech-language pathologists, it is believed that the SE and TE speakers were able to manipulate pitch by controlling the rate of PE segment vibration to achieve linguistic purposes.

However, FSP associated with PA speech appears to be markedly lower than that associated with NL, SE, and TE speech (see Table 1). This may reflect the rate of vibration at which the rubber reed located inside the pneumatic device used by PA speakers was vibrating. Although control of the vibratory rate of the pneumatic device seems possible, it is not certain if and how the rate of vibration can be controlled due to the insufficient information on PA speech. In a perceptual study of SE, TE, EL, and PA speech of Cantonese, Ng et al. [2] reported that PA sound was associated with less amount of hoarseness and noise when compared with SE and TE speech. But it is generally agreed that PA sound is perceived as strange, loud, and lack of pitch variation. This is likely the reason for the decreasing prevalence of PA speech among laryngectomees nowadays. Future studies perhaps should focus on the vibratory behavior of a pneumatic device, and the control over the rate of vibration by PA speakers.

FSP of NL, SE, and TE speech was found to be highly correlated with the average F_0 value in continuous speech production. However, this may not be true for EL. The average FSP value of EL speakers of Cantonese was 217.93 Hz, while all EL speakers used the

Servox electrolarynx with a preset pitch level of approximately 80 Hz [18]. Despite the pitch variability of the EL device, none of the EL speakers was able to signal different lexical tones by manipulating the pitch control during the experiment. The discrepancy between the pitch of electrolarynx and the average FSP may be due to the energy distribution of its particular sound. EL sound quality is notoriously associated with the radiated noise [2,43] which could interfere with the energy distribution of EL device, resulting in low-frequency energy deficit in speech spectrum. As shown in Figure 1, EL speech is depleted of energy in the low frequency region (0 – 500 Hz). This is consistent with findings from previous studies [39,42,43]. For frequency below 500 Hz, some EL speakers may demonstrate an energy that is more than 30 dB lower than NL speakers [cf., 43]. The lack of low frequency energy may be contributed by the directed radiated noise of EL speech, which has been discussed by other researchers previously [cf., 44]. Since EL sound energy was being interfered by the radiated noise, the first spectral peak was found in the higher frequency rather than the average F_0. As FSP only reflects the maximum amplitude across the entire frequency range, the interference of the radiated noise in EL speech implies that the average FSP value of EL speech does not represent the pitch of the electrolarynx being used. The effect of radiated noise in EL speech renders the unique EL sound quality. Different strategies are under development to reduce or eliminate this noise component [43,44].

Mean Spectral Energy

The average amplitude across 0 – 8 kHz was measured as MSE, which is a measurement relative to the maximum amplitude. The greater is the MSE value, the lesser is the difference between spectral energy across the frequency range and the maximal amplitude. In a study of pain-induced infant cries, Fuller and Horii [45] noted that variability of spectral energy was related to the tension of vocal quality; a tense cry quality was associated with a diminished variability of spectral energy, which resulted in a close-to-zero

MSE value. As shown in Table 1, SE speakers exhibited an average MSE value slightly lower than TE speakers. This implies that, although both SE and TE speakers were using the PE segment as the new sound source, SE speech may sound more tense than TE speech. Such difference may be related to the difference the way air is stored in SE and TE speech. While SE speakers use the upper part of the esophagus for air storage and as the new vibratory device, TE speakers made use of the lungs as the air reservoir. The different air reservoir in SE and TE speakers may have contributed to the difference in sound quality between SE and TE speech. This is supported by the aerodynamic findings of TE speakers of English reported by Moon and Weinberg [5]. English TE speakers were found to exhibit decreased airway resistance during speech production when compared with SE speakers. Apparently, more specific information on the tension in the PE segment in SE and TE speakers is needed to confirm this conjecture.

The average (absolute) MSE values of SE and TE speech appear to be closer to zero than NL speech. This may indicate that, in SE and TE speakers, the PE segment was generating a sound that would be more tense than a laryngeal sound made by NL speakers. However, the MSE values of EL and PA speech cannot be directly compared with that of NL speech due to the different sounding mechanisms and coupling between sound source and filter.

Spectral Tilt

The spectral tilt measurement can be understood as a comparison between low-frequency energy (between 0 – 1 kHz) and high frequency energy (between 1 – 5 kHz). NL speakers exhibited the greatest ST value when compared with SE, TE, EL, and PA speakers. This is consistent with the steeper attenuation in the LTAS contours of NL speech discussed earlier. A steeper LTAS contour in NL speech implies a greater difference between low frequency energy and high frequency energy, and thus a “greater spectral tilt”.

Table 1 shows that SE and TE speakers exhibited comparable ST values. According to Löfqvist and Mandersson [26], Mendoza, Muñoz and Valencia-Naranjo [46], and Mendoza, Valencia, Muñoz, and Trujillo [47], a high ST indicates that the LTAS spectrum is dominated by low frequency energy (the fundamental and lower harmonics) and corresponds to a hypofunctional voice. On the contrary, a low ST implies a hyperfunctional voice. However, simple comparison of ST values among different speaker groups may be inaccurate due to the different sounding mechanisms.

Conclusion

A comparison among the average LTAS contours associated with NL, SE, TE, EL, and PA speech of Cantonese reveals that the average LTAS contour of NL speech showed a steeper attenuation rate in amplitude (energy) with frequency when compared with the other speaker groups. While average FSP values of NL, SE and TE speech were comparable, the FSP value of PA was only about one half of that of NL speech. The low FSP value in PA speech is likely to be due to the different rate of vibration of the rubber reed inside the pneumatic device. An average FSP value of 217.93 Hz was found for EL speech which was much higher than the pitch of EL sound. This value failed to reflect how fast the EL device is vibrating, due to the interference of the radiated noise associated with EL speech. Despite using the similar phonation method, SE speakers exhibited a slightly lower (absolute) MSE value than TE speakers. This indicates a slightly higher tension in the PE segment in SE speakers than TE speakers, which may be caused by the different air reservoir system in SE and TE speech. Closely correlated with rate of amplitude attenuation of LTAS contour, ST values of NL speech were found to be significantly greater than the other speaker groups, consistent with the steeper attenuation found in the LTAS contours in NL speech.

References

[1] Law KYI. Communication activity and participation after laryngectomy. MS thesis. Hong Kong: University of Hong Kong; 2005.

[2] Ng M, Kwok CL, Chow SF. Speech performance of adult Cantonese-speaking laryngectomees using different types of alaryngeal phonation. J Voice. 1997;11:338-344.

[3] Damsté PH. Some obstacles in learning esophageal speech. In: Keith RL, Darley FL, editors. Laryngectomy rehabilitation, 3rd ed. Austin, TX: Pro-Ed Inc.; 1994. p. 235-242.

[4] Hillman RE, Walsh MJ, Heaton JT. Laryngectomy speech rehabilitation – a review of outcomes. In: Doyle PC, Keith RL, editors. Contemporary considerations in the treatment of rehabilitation of head and neck cancer – Voice, speech, and swallowing. Austin, TX: Pro-Ed, Inc.; 2005. p.77-90.

[5] Moon JB, Weinberg B. Aerodynamic and myoelastic contributions to tracheoesophageal voice production. J Speech Hear Res. 1987;30:387-395.

[6] Ng M, Wong J. Voice onset time characteristics of esophageal, tracheoesophageal, and laryngeal speech of Cantonese. J Speech Lang Hear Res. In press.

[7] Gandour J, Weinberg B. Perception of intonational contrasts in alaryngeal speech. J Speech Hear Res. 1983;26:142-148.

[8] Gandour J, Weinberg B, Garzione B. Perception in lexical stress in alaryngeal speech. J Speech Lang Hear Res. 1983;26:418-424.

[9] Gandour J, Weinberg B, Petty SH, Dardarananda R. Vowel length in Thai alaryngeal speech. Folia Phoniatr Logop. 1987a;39:117-121.

[10] Gandour J, Weinberg B, Petty SH, Dardarananda R. Voice onset time in Thai alaryngeal speech. J Speech Hear Disord. 1987b;52:288-294.

[11] Gandour J, Weinberg B, Petty SH, Dardarananda R. Tone in Thai alaryngeal speech. J Speech Lang Hear Res. 1988;53:23-29.

[12] Holly SC, Lerman JW, Randolph K. A comparison of the intelligibility of esophageal, electrolarynx, and normal speech in quiet and in noise. J Commun Disord. 1983;16:143-155.

[13] Liu H, Wan M, Wang S, Niu H. Aerodynamic characteristics of laryngectomees breathing quietly and speaking with electrolarynx. J Voice. 2004;18:567-577.

[14] Liu H, Wan M, Wang S, Wang X, Lu C. Acoustic characteristics of Mandarin esophageal speech. J Acoust Soc Am. 2005;118:1016-1025.

[15] Liu, H., Ng, M., Wan, M., Wang, S. (in print). The effect of tonal changes on voice onset time in Mandarin esophageal speech. Journal of Voice.

[16] Liu H, Ng M, Wan M, Wang S, Zhang Y. Effects of place of articulation and aspiration on voice onset time in Mandarin esophageal speech. Folia Phoniatr Logop. 2007;59:147-154.

[17] Ng M, Chu R. An acoustical and perceptual study of vowels produced by alaryngeal speakers of Cantonese. Folia Phoniatr Logop. In press.

[18] Ng M, Gilbert HR, Lerman JW. Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech. Folia Phoniatr Logop. 2001;53:36-47.

[19] Ng M, Lerman JW, Gilbert HR. Perceptions of tonal changes in normal laryngeal, esophageal, and artificial laryngeal male Cantonese speakers. Folia Phoniatr Logop. 1998;50:64-70.

[20] Robbins J, Fisher HB, Blom EC, Singer MI. A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. J Speech Hear Disord. 1984;49:202-210.

[21] Weinberg B, Bennett S. Selected acoustic characteristics of esophageal speech produced by female laryngectomees. J Speech Hear Res. 1972;15:211–216.

[22] Chen HC, Mardini S, Yang CW. Voice reconstruction using the free ileocolon flap versus the pneumatic artificial larynx: a comparison of patients' preference and experience following laryngectomy. J Plast Reconstr Aesthet Surg. 2006;59:1269-1275.

[23] Ching TY, Williams R, van Hassalt A. Communication of lexical tones in Cantonese alaryngeal speech. J Speech Hear Res. 1994;37:557-563.

[24] Borden GJ, Harris KS, Raphael LJ. Speech science primer: physiology, acoustics, and perception of speech, 4th edition. Baltimore, MD: Lippincott Williams & Wilkins; 2002.

[25] Löfqvist A. Long-time studies of vocal behavior. NICDC Monograph; 1990. p. 96-102.

[26] Löfqvist A, Mandersson B. Long-time average spectrum of speech and voice signals. Folia Phoniatr Logop. 1987;21:221-219.

[27] Goberman AM, Robb MP. Acoustic examination of preterm and full-term infant cries: The long-time average spectrum. J Speech Lang Hear Res. 1999;42:850-861.

[28] Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox R, et al. An international comparison of long-term average speech spectra. J Acoust Soc Am. 1994;96:2108-2120.

[29] Formby C, Monsen RB. Long-term average speech spectra for normal and hearing-impaired adolescents. J Acoust Soc Am. 1982;71:196-202.

[30] Goberman AM, Robb MP. Acoustic characteristics of crying in infantile laryngomalacia. Logoped Phoniatr Vocol. 2005;30:79-84.

[31] Kitzing P. LTAS criteria pertinent to the measurement of voice quality. J Phonet. 1986;14:477-482.

[32] Kovačić G, Boersma P, Domitrović H. Long-term average spectra in professional folk singing voices: A comparison of the Klapa and Dozivački styles. Proceedings of Institute of Phonetic Sciences Volume 25. Stockholm; University of Amsterdam; 2003. p. 53-64.

[33] Laukkanen A, Sundberg J, Björkner E. Acoustic study of throaty voice quality. TMH-QPSR, KTH. 2004;46:13-23.

[34] Wendler J, Doherty E, Hollien H. Voice classification by means of long-term spectra. Folia Phoniatr Logop. 1980;32:51-60.

[35] White P. A study of the effects of vocal intensity variation on children's voices using long-term average spectrum (LTAS) analysis. Logoped Phoniatr Vocol. 1998;23:111-120.

[36] White P. Long-term average spectrum (LTAS) analysis of sex- and gender-related differences in children's voices. Logoped Phoniatr Vocol. 2001;26:97-101.

[37] Globlek D, Stajner-Katusic S, Musura M, Horga D, Liker M. Comparison of alaryngeal voice and speech. Logoped Phoniatr Vocol. 2004;29:87-91.

[38] Weinberg B, Horii Y, Smith BE. Long-time spectral and intensity characteristics of esophageal speech. J Acoust Soc Am. 1980;67:1781–1784.

[39] Weiss MS, Yeni-Komshian GH, Heinz JM. Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx. J Acoust Soc Am. 1979;67:1298–1308.

[40] Wong SL. A Chinese Syllabary Pronounced according to the Dialect of Canton, 2nd edition.. Hong Kong, China: Chung Hwa Book; 1986.

[41] Pickett JM. The acoustics of speech communication – Fundamentals, speech perception theory, and technology. Needham Heights, MA: Allyn & Bacon; 1999.

[42] Qi Y, Weinberg B. Spectral slope of vowels produced by tracheoesophageal speakers. J Speech Hear Res. 1991;34:243-247.

[43] Meltzner GS, Kobler JB, Hillman RE. Measuring the neck frequency response function of laryngectomy patients: Implications for the design of electrolarynx devices. J Acoust Soc Am. 2003;114:1035-1047.

[44] Liu H, Ng M. Electrolarynx in voice rehabilitation. Auris Nasus Larynx. 2007;34:327-332.

[45] Fuller BF, Horii Y. Spectral energy distributions in four types of infant vocalizations. J Commun Disord. 1988;21:251-261.

[46] Mendoza E, Muñoz J, Valencia-Naranjo N. The long-term average spectrum as a measure of voice stability. Folia Phoniatr Logop. 1996;48:57-64.

[47] Mendoza E, Valencia N, Muñoz J, Trujillo H. Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). J Voice. 1996;10:59-66.

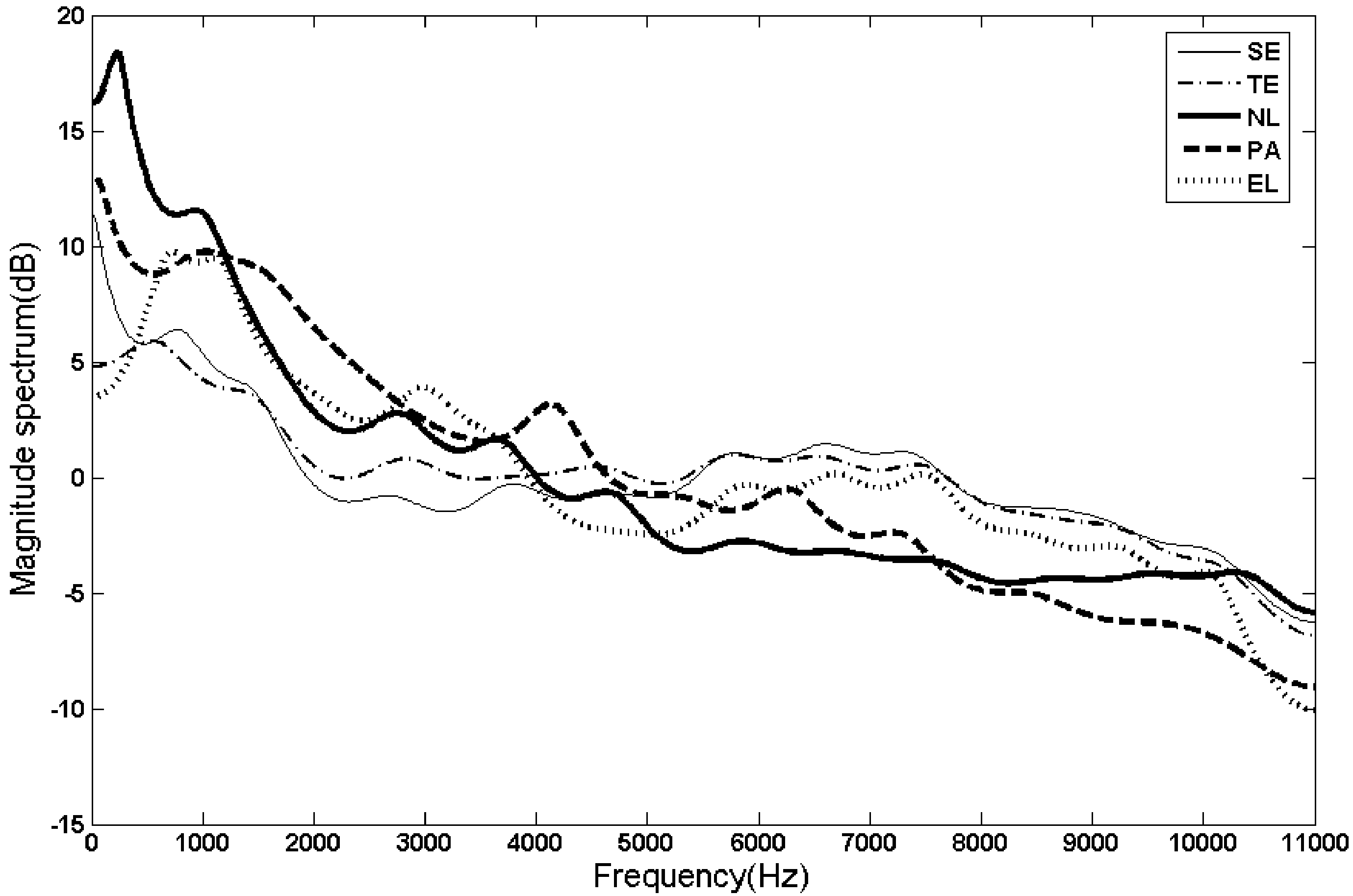
Table 1. Average first spectral peak (FSP), mean spectral energy (MSE), and spectral tilt (ST) associated with normal laryngeal (NL) speakers, standard esophageal (SE) speakers, tracheoesophageal (TE) speakers, electrolaryngeal (EL) speakers, and users of a pneumatic artificial larynx (PA). The corresponding standard deviation is shown in parentheses.

Group	FSP (Hz)	MSE (dB)	ST
NL	232.00 (38.84)	-30.33 (1.91)	1.32 (0.619)
SE	227.75 (66.33)	-18.64 (2.67)	0.36 (0.024)
TE	227.50 (50.43)	-22.61 (1.97)	0.33 (0.050)
EL	217.93 (51.21)	-31.51 (1.79)	0.49 (0.119)
PA	114.91 (24.64)	-47.79 (2.52)	0.56 (0.095)

Figure Caption

Figure 1. Average long term average spectra (LTAS) associated with normal laryngeal (NL) speakers, standard esophageal (SE) speakers, tracheoesophageal (TE) speakers, electrolaryngeal (EL) speakers, and users of a pneumatic artificial larynx (PA).

Figure 1


SE
TE
NL
PA
EL
Magnitude spectrum(dB)
Frequency(Hz)
20
15
10
5
0
-5
-10
-15
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000