HHS Public Access

Author manuscript

Folia Phoniatr Logop. Author manuscript; available in PMC 2022 March 23.

Published in final edited form as:

Folia Phoniatr Logop. 2022 ; 74(2): 103–111. doi:10.1159/000517676.

The Effect of Clear Speech on Cantonese Alaryngeal Speakers' Intelligibility

Tak Fai Hui[a], **Steven Randall Cox**[b], **Ting Huang**[c], **Wei-Rong Chen**[c], **Manwa Lawrence Ng**[a]

[a]Speech Science Laboratory, University of Hong Kong, Hong Kong SAR, China;

[b]Department of Communication Sciences and Disorders, Adelphi University, New York, NY, USA;

[c]Haskins Laboratories, New Haven, CT, USA

Abstract

Background/Aim: The purpose of this study was to provide preliminary data concerning the effect of clear speech (CS) on Cantonese alaryngeal speakers' intelligibility.

Methods: Voice recordings of 11 sentences randomly selected from the Cantonese Sentence Intelligibility Test (CSIT) were obtained from 31 alaryngeal speakers (9 electrolarynx [EL] users, 10 esophageal speakers and 12 tracheoesophageal [TE] speakers) in habitual speech (HS) and CS. Two naïve listeners orthographically transcribed a total of 1,364 sentences.

Results: Significant effects of speaking condition on speaking rate and CSIT scores were observed, but no significant effect of alaryngeal communication methods was noted. CS was significantly slower than HS by 0.78 syllables/s. Esophageal speakers demonstrated the slowest speech rate when using CS, while EL users demonstrated the largest decrease in speaking rate when using CS compared to HS. TE speakers had the highest CSIT scores in HS (listener 1 = 81.4%; listener 2 = 81.3%), and esophageal speakers had the highest CSIT scores in CS (listener 1 = 87.5%; listener 2 = 89.7%). EL users experienced the largest increase in intelligibility while using CS compared to HS (9.1%) followed by esophageal speakers (8.9%) and TE speakers (1.4%).

Conclusion: Preliminary data indicate that CS may significantly affect Cantonese alaryngeal speakers' speaking rate and intelligibility. However, intelligibility appeared to vary considerably across speakers. Further research involving larger, heterogeneous groups of speakers and listeners

Correspondence to: Manwa Lawrence Ng, manwa@hku.hk.

Author Contributions

T.F.H. was responsible for data collection, data analysis, and manuscript preparation. S.R.C. was responsible for data analysis, and manuscript preparation and revision. W.-R.C. was responsible for data analysis and manuscript revisions. T.H. was responsible for data analysis and manuscript revisions. M.L.N. was responsible for data analysis, and manuscript preparation and revisions.

Statement of Ethics

All subjects have given their written informed consent, and the ethics of the study was approved by the Faculty Research Ethics Committee of the Faculty of Education, University of Hong Kong, on November 18, 2019.

Conflict of Interest Statement

The corresponding author of the article is currently an Associate Editor of *Folia Phoniatrica et Logopaedica*. All authors have no conflicts of interest to declare.

alongside longer and more refined CS training protocols should be conducted to confirm that CS can improve Cantonese alaryngeal speakers' intelligibility.

Keywords

Alaryngeal speech; Cantonese; Clear speech; Laryngectomy; Speech intelligibility

Introduction

Total laryngectomy is an invasive surgical procedure that is commonly used to treat patients with advanced laryngeal cancer [1]. Total laryngectomy involves the removal of the entire larynx, which is essential for phonation. Individuals who undergo total laryngectomy (i.e., laryngectomees) experience a host of physical, psychological, and socioemotional changes [2–4]. For example, they lose the ability to generate voicing, and as a result, have to learn an alternative alaryngeal speaking method to regain verbal communication. Three common methods of alaryngeal speech include esophageal speech (ES), tracheoesophageal speech (TE), and electrolaryngeal speech.

Alaryngeal speaking methods differ in the generation of sound energy used for voice and speech production. These methods can be generally categorized into 2 groups: ES and TE utilize an intrinsic sound source for phonation [5], and electrolaryngeal speech requires the use of an extrinsic sound source (i.e., an electrolarynx [EL]) [6, 7]. More specifically, ES and TE speech rely on the vibration of the pharyngoesophageal segment, which is also known as the neoglottis, as a postlaryngectomy voicing source. Speakers using ES generate voice by expelling air that is stored in the upper esophagus, and the outward airflow triggers vibration of the pharyngoesophageal segment to generate sound [8]. In TE speech, a TE puncture is surgically created through the shared wall between the trachea and esophagus [9]. A one-way valve (i.e., TE prosthesis) is placed inside the puncture and allows air to flow from the lungs to the esophagus, where the air vibrates the pharyngoesophageal segment and generates sound that is articulated into speech. Electrolaryngeal speech requires an external, hand-held electronic device, known as the EL, to produce postlaryngectomy voice. The vibratory head of the EL transmits sound energy through neck tissues into the vocal tract, which vibrates the air column inside and is articulated into speech [10].

Regardless of the postlaryngectomy communication method, all alaryngeal speech appears to result in reduced intelligibility compared to typical, laryngeal speech. Differences in acoustics, aerodynamics, and auditory-perceptual characteristics (e.g., intelligibility, speech acceptability, listener comfort, etc.) associated with alaryngeal communication methods have been reported [5, 11–19]. For example, EL users have been consistently shown to have lower voice-related quality of life scores [20], lower intelligibility [21], and less acceptable speech compared to other forms of alaryngeal speech [22]. Further, some aspects of TE speech, such as speaking rate and inflection, have been reported to be as acceptable as laryngeal speech [23]. It appears, then, that the level of noise in an alaryngeal signal may distract listeners and, consequently, impact their ability to comprehend an alaryngeal speaker's message.

Gandour and Weinberg [24] highlighted that ES and TE speakers are more proficient at producing intonational contrasts compared to EL users. EL users were generally found to lack the ability to control F0, leading to difficulties producing required intonation patterns of American English [25]. A follow-up study by Gandour et al. [26] attempted to understand the extent to which ES speakers and EL speakers using a Servox® can produce tonal contrasts in Thai (i.e., a tonal language). Findings suggested that Thai alaryngeal speakers were not able to accurately produce phonemic tones [26]. Gandour et al. [26] stated that their results contrasted with their prior work involving American English alaryngeal speakers, and the "discrepancy may be related to differences in phonological and/or phonetic characteristics between the 2 languages" (p. 28) rather than F0-based acoustic properties. Overall, the postsurgical neoglottis and Servox® EL might be inadequate for producing tonal contrasts in Thai [26]. Tone production associated with alaryngeal speech of other tonal languages has also been reported, including Mandarin [27, 28] and Cantonese [17–18, 29–32].

Besides tone production, prior studies have compared other speech performances of Cantonese alaryngeal speakers. For example, Law et al. [14] compared the intelligibility of 49 Cantonese alaryngeal speakers using the different types of alaryngeal speech. The Cantonese Sentence Intelligibility Test (CSIT) [33] was used to evaluate intelligibility, and higher intelligibility scores were associated with EL speech (77.3%), followed by TE (61.5%) and ES speech (59.7%). Alternately, Ng et al. [18] found no significant differences in intelligibility among superior alaryngeal speakers using a 7-point equal-appearing interval scale ranging from 1 (poor intelligibility) to 7 (excellent intelligibility; i.e., 4.15 for EL, 4.41 for TE, and 3.89 for ES). The discrepant findings might be attributed to the fact that only "superior" alaryngeal speakers were recruited by Ng et al. [18], while the proficiency of the alaryngeal speakers was not rated by Law et al. [14]. These findings highlight the significance of postlaryngectomy speech rehabilitation and training. As a result, Cantonese alaryngeal speakers might be able to speak with a comparable level of intelligibility across alaryngeal communication methods with continued efforts to improve alaryngeal voice and speech rehabilitation.

Clear speech (CS) is a deliberate way of speaking "clearly" to improve one's intelligibility [11, 34–38]. CS features a distinct style of speaking that involves increasing vocal intensity, slowing speech rate and overarticulating during speech production [34, 36, 39]. Such features of CS can be elicited with explicit instructions (e.g., "to speak clearly," "to overarticulate," "to speak slower"). Further, CS is associated with differences in a range of acoustic properties when compared to conversational speech. For example, CS has led to an increase of 5–8 dB in vocal intensity, a slower speaking rate, and a larger vowel space [39–43]. Originally used to enhance speech understanding amongst persons with hearing impairment [36], CS has been more recently used by speakers with dysarthria while communicating with normal-hearing listeners [44–46].

For laryngectomees who experience a significant structural change in their speech apparatus and compromised intelligibility, it is crucial to learn and adapt strategies to improve their communication. Recent research has expanded the use of CS from individuals with dysarthria to laryngectomees. Cox and Doyle [11] evaluated the influence of CS on naïve

listeners' ratings of speech acceptability and listener comfort. They found that CS negatively affected speech acceptability but not listener comfort. However, it should be noted that higher ratings of speech acceptability, which are based on pitch, rate, understandability, and voice quality, do not always result in high levels of speech intelligibility [14, 33]. A follow-up study by Cox et al. [47] examined the effect of CS on vowel production by 10 EL users and found that vowels produced using CS had a longer duration, but there was no effect on vowel identification. Though both studies indicated no statistically significant benefits in terms of acceptability and vowel identification in EL speech [11, 47], the effect of CS on intelligibility at the sentence level was not investigated. Moreover, the 2 studies only focused on EL users whose primary language was American English. Therefore, it is not known whether CS can significantly improve the intelligibility of alaryngeal speakers who speak a tonal language.

While generalizations across languages should be made with caution, the work of Gandour et al. [26] suggested that alaryngeal speakers may have difficulty producing intelligible speech in tonal languages. Cantonese is a tonal language in which the manipulation of lexical tones can affect the message conveyed [14]. However, the effect of CS on intelligibility in Cantonese alaryngeal speakers is unknown. The present research aimed to address the following research questions: (1) Is there an effect of CS on Cantonese alaryngeal speakers' intelligibility? And (2) is there a difference of the effect of CS on different types of alaryngeal communication methods?

Methods

Participants

Speakers.—Thirty-one male alaryngeal speakers were recruited from the New Voice Club of Hong Kong, which is a self-help organization of laryngectomees. The demographic information is summarized in Table 1. All speakers were referred by a practicing speech therapist at the New Voice Club with more than 10 years of experience in alaryngeal voice and speech rehabilitation. The recruitment criteria of alaryngeal speakers included: (1) they were proficient speakers of alaryngeal speech as judged by the experienced speech therapist, (2) they were physically healthy with no other speech, language, and hearing problems, except those associated with laryngectomy, and (3) they were native speakers of Cantonese. The speakers had adopted 1 of 3 alaryngeal speaking methods: EL (n = 9), ES (n = 10) and TE speech (n = 12). Their ages ranged from 35 to 91 years (mean = 66.48 years, SD = 11.31 years) and the duration of using a particular type of primary alaryngeal speech ranged from 3 months to 25 years (mean = 7.15 years, SD = 6.74 years). All EL users used the Servox® Digital neck-type EL (mean = 79.6 Hz; SD = 10.1 Hz), and all TE speakers digitally occluded their stoma during speech production. The possible effect on intelligibility of the brand of the EL and the mode of occlusion used by TE speakers were thus minimized. Speakers with any known form of cognitive impairment were excluded from the study.

Listeners.—Two healthy adult females (aged 23 and 27 years) were recruited through advertisements at the University of Hong Kong. Both listeners were native speakers of Cantonese and reported no history of speech, language, or hearing problems. Listeners

completed an undergraduate degree and were considered “naïve” to alaryngeal speakers as they had not yet studied, worked, or undertaken research with this patient population.

Acquisition of Speech Stimuli

Speech stimuli used in the study were selected from the CSIT [33]. The CSIT consists of a pool of more than 1,000 Cantonese sentences that vary in length from 5 to 15 words and includes 100 sentences for each sentence length. These sentences contain real words and approximate spontaneous speech rather than nonsense sentences. Eleven sentences (one from each sentence length) containing a total of 220 words were randomly drawn from the sentence pool and randomized to form a set of sentence stimuli for each speaker. The sentences were printed out on sheets of A4 paper using a font size of 48 points.

Recording of Speech Stimuli

Recording of speech stimuli was conducted in a quiet room at the New Voice Club of Hong Kong. The background noise level was monitored using a portable sound level meter (1350A, TES Electrical Electronic Corp., Taiwan). Speech samples were obtained using a professional-grade condenser microphone (Shure SM58) and a preamplifier (M-Audio USBPre) connected to a laptop computer with Praat [48]. During the recording, the microphone was placed 15 cm from the participant’s mouth. All recordings were digitized at a sampling rate of 44 kHz and quantized at 16 bits/sample. Each speaker was first asked to read the sentence stimuli (11 sentences containing 220 words) on the paper provided to them in habitual speech (HS), just as they would read in daily communication. A period of practice time was given to the speakers to familiarize and raise any questions regarding the sentence presented each time to ensure that the speaker knows every word in the stimuli before the actual recording procedure. During the procedure, the experimenter noted errors while speakers were reading (obvious reading error owing to inaccurate identification of words) and provided the correct pronunciation of the misread word(s). Sentences were rerecorded if speakers misread words. After a brief break, prerecorded instructions to produce CS were provided to participants. Specific instructions “to overarticulate” (in Cantonese: 誇張咁讀) and “to slow down their speech” (in Cantonese: 減慢語速) were provided [11, 36, 39]. Each speaker was asked to read a practice sentence using CS. A second demonstration was provided when the participant failed to correctly produce the practice sentence in CS. A maximum of 2 demonstrations was provided. After the practice session, the participants were asked to read the same set of sentence stimuli, and their productions were recorded. Across all 31 speakers, 682 sentences (11 sentences per speaker × 31 speakers × 2 speech conditions) were recorded and presented to each listener in a randomized order for intelligibility assessment.

Evaluation of Intelligibility

Evaluation of intelligibility was completed in a quiet room that was monitored using a portable sound level meter (1350A, TES Electrical Electronic Corp., Taiwan). The listeners were seated in front of a laptop computer and instructed to listen to speech stimuli presented through the speaker of the laptop computer (Mac-Book pro, Apple). Listeners transcribed sentences by typing each word into an Excel file. More specifically, each sentence was inputted on a designated cell in an Excel file with the total number of words labeled so

that listeners were aware of how many words should be transcribed for each sentence. Each sentence was presented a total of 2 times. This procedure was repeated until all sentences were transcribed.

Data Analysis

Acoustic Analysis.—The recorded speech was transcribed, forced aligned by Montreal forced aligner [49], and then the syllable and segment boundaries were manually adjusted. We calculated the speaking rate as the number of syllables per second for each sentence, excluding interruptions such as coughing or long pauses. A linear mixed effect model was fitted by using the "lme4" [50] package in R [51] to assess the differences in speaking rate between alaryngeal communication method and speaking conditions. Post hoc comparisons were carried out by using the "multcomp" package [52] and p values adjusted by false discovery rate [53]. We entered "CONDITION" (2 levels: HS and CS) and "METHOD" (3 levels: EL, ES, and TE) as fixed effects, and by-speaker random intercept and random slope of "CONDITION" as random effects into the model. The interaction of "CONDITION" and "METHOD" did not improve the model, based on the result of a likelihood ratio test, and was thus not included.

Intelligibility Analysis.—The CSIT intelligibility scores associated with HS and CS were calculated by dividing the number of correctly transcribed words by the total number of words in the speech stimuli, yielding an intelligibility percentage score. The scores from the 2 listeners were averaged and used for further statistical analyses. To assess the effect of speaking condition on alaryngeal intelligibility, a two-way repeated measures analysis of variance (ANOVA) was conducted, and a post hoc pairwise comparison was carried out for the CSIT scores. Effect sizes were determined using partial η^2 and were interpreted using guidelines by Cohen (i.e., 0.01 = small effect, 0.06 = medium effect, and 0.14 = large effect) [54]. An α-level of $p < 0.05$ was used for all statistical analyses.

Results

Reliability Analysis

Both intrarater and interrater reliabilities were assessed using 10 sets of speech stimuli (10 × 11 = 110 sentences; approx. 15% of sentences). Reliability stimuli were assessed after all of the primary study stimuli had been evaluated. Pearson product-moment correlations were used to assess intrarater reliability. A high intrarater reliability was obtained for the CSIT intelligibility scores (listener 1: $r = 0.993$, $p < 0.05$; listener 2: $r = 0.999$, $p < 0.01$). For interrater reliability, the intraclass correlation coefficient (2, k) and their 95% CIs were calculated based on the mean of k raters with a two-way mixed-effects model (intraclass correlation coefficient [2, k] model). The intraclass correlation coefficient for the CSIT score was 0.749 with 95% CI of 0.586–0.847, which indicated moderate interrater reliability [55].

Speaking Rate

Speaking rate data for each speaker group are shown in Figure 1. The fitted linear mixed effect model revealed that Cantonese alaryngeal speakers used a rate that was significantly ($p < 0.0001$) slower in CS (mean = 2.11 syllables/s, SD = 0.67) compared to HS (mean

= 2.89 syllables/s, SD = 0.69). EL users demonstrated the biggest decrease in speaking rate when using CS (mean = 2.39 syllables/s) compared to HS (mean = 3.22 syllables/s), followed by TE speakers (mean = 2.91 syllables/s in HS; mean = 2.12 syllables/s in CS) and ES speakers (mean = 2.57 syllables/s in HS; mean = 1.86 syllables/s in CS). ES speakers used the slowest speaking rate, as compared to EL and TE. However, post hoc comparisons showed no significant differences in speaking rate between any pair of the 3 alaryngeal communication methods (EL-ES = 0.45, $p = 0.24$; TE-ES = 0.21, $p = 0.38$; EL-TE = 0.24, $p = 0.38$).

CSIT Intelligibility Scores

CSIT intelligibility scores for listeners and speakers are shown in Table 2 and Figure 2, respectively. The mean CSIT scores for speakers ranged from 72.4% (EL speech) to 81.3% (TE speech) in HS, while the CSIT scores ranged from 81.5% (EL speech) to 88.6% (ES speech) in CS. Overall, EL users had the lowest CSIT intelligibility scores in HS (listener 1 = 71.2%; listener 2 = 73.6%), followed by ES speakers (listener 1 = 80.3%; listener 2 = 79.3%) and TE speakers (listener 1 = 81.4%; listener 2 = 81.3%). ES speakers had the highest mean intelligibility score in CS (listener 1 = 87.5%; listener 2 = 89.7%), followed by TE speakers (listener 1 = 83.3%; listener 2 = 82.2%) and EL users (listener 1 = 86.5%; listener 2 = 76.6%). EL users had the greatest increase in their CSIT scores while using CS (+9.1%), while ES and TE speakers' CSIT scores increased by 8.9 and 1.4%, respectively.

A repeated-measures ANOVA indicated a significant effect of speaking condition on CSIT scores, $F(1, 29) = 4.317$, $p = 0.047$, partial $\eta^2 = 0.13$. The magnitude of the effect suggested that the speaking condition had a medium effect on CSIT scores. However, the repeated-measures ANOVA did not indicate a significant effect of speaker group on CSIT scores, $F(1, 29) = 0.566$, $p = 0.574$.

Discussion

The purpose of this study was to provide preliminary data concerning the effect of CS on Cantonese alaryngeal speakers' speaking rate and intelligibility. Two naïve listeners evaluated the intelligibility of 32 alaryngeal speakers (9 EL, 10 ES, 12 TE) by providing orthographic transcriptions for a total of 1,364 sentences. Alongside prior research that reported significant improvements in intelligibility in CS condition [44–46], CS had a significant effect on speaking rate and intelligibility in the current study.

Prior research has reported that CS can improve the intelligibility of individuals with speech impairments resulting from different etiologies (e.g., dysarthria following stroke, traumatic brain injury, and Parkinson's disease) [44–46]. Yet, there is a dearth of literature that has investigated the effect of CS on alaryngeal speakers. Research involving alaryngeal speech has only focused on EL users whose primary language was American English, and the primary outcome measures have been speech acceptability, listener comfort, and vowel identification [11, 47]. Thus, to explore the potential of CS for non-English alaryngeal speakers, the current study investigated the effect of CS on Cantonese alaryngeal speakers' intelligibility at the sentence level.

General trends in intelligibility were observed. For example, TE speakers had the highest intelligibility in HS, ES speakers had the highest intelligibility in CS, and EL users appeared to have the largest increase in CSIT scores when moving from HS to CS. A systematic review reported by van Sluis et al. [21] summarized 8 studies investigating the intelligibility of different alaryngeal communication methods using either a rating scale or accuracy of word transcription. TE was found to be the most intelligible when compared with ES and EL speech in most of the studies reviewed [21]. However, contradictory findings have been observed in 2 studies involving Cantonese alaryngeal speakers [14, 18]. For example, Ng et al. [18] found that TE was the most intelligible, but the intelligibility ratings of EL were higher than that of ES speakers. Meanwhile, Law et al. [14] reported that EL speech was the most intelligible alaryngeal speaking method when compared to TE and ES. The discrepancy in the prior literature may be explained by the fact that both studies had specific selection criteria; speakers were either deemed as having achieved the “maximum ability in acquiring the new speaking methods” [14, p. 705] or were “expert” alaryngeal speakers [18]. However, the proficiency of the alaryngeal speakers was not experimentally controlled in the present study. Instead, all of the alaryngeal speakers were referred by an experienced speech therapist from the New Voice Club of Hong Kong.

Group results indicated significant effects of speaking condition on speaking rate and CSIT scores, but no significant effect of speaking condition was noted between alaryngeal communication methods. Improvements in CSIT scores were observed in all the speaker groups when alaryngeal speakers were using CS. For example, a 9.1% improvement was observed in the CS condition for EL users. Such an increase was higher than the improvement in word intelligibility as reported in Cox [50], who reported a 1.3% improvement in English word identification involving orthographic transcription. The differences between the 2 languages may be a major factor leading to the different results. One of the hypothesized underlying principles of CS in improving the EL’s intelligibility is the reduction of speech rate and overarticulation, and these could be achieved by the possible increase in the number and duration of pauses during speech production [41]. Unlike English, however, Cantonese is a tonal language in which lexical tone plays an important role in distinguishing and understanding Cantonese speech. It has been documented that, for a tonal language, when pitch information is not available, other perceptual cues such as vowel duration and intensity might be used by listeners to more accurately detect lexical tones (e.g., 5759). As tone variation is not common with the use of a Servox® Digital, slowing down and carefully articulating each syllable in CS might help listeners better understand and, therefore, yield better intelligibility. Another possible factor specific to EL speech is that better coordination in voicing and manipulation of the device may be achieved in CS, thanks to the slower speech rate [11]. These potential benefits on intelligibility may thus not be reflected in the intelligibility measure at the word or phoneme level, of which the duration of production was much shorter and the phonotactic structure is less complex compared to the production of sentences. Moreover, the more contextualized linguistic context at the sentence level may also lead to a higher intelligibility score when compared to that of word or phoneme level [47, 56].

An increase of 8.9% was observed in the CSIT intelligibility score in ES speakers while using CS, which might be partly attributed to ES speakers having the slowest speaking

rate in CS. The implementation of CS may positively affect how ES speakers control and coordinate their already limited volume of air for speech, potentially leading to the increases in intelligibility. Consider that ES speakers are reported to have particular difficulty with articulation of aspirated consonants due to the necessity to build up air pressure during their production [60, 61]. In fact, Ng et al. [61] found that ES speakers exhibit a significantly greater articulatory contact pressure during speech production as compared to other laryngeal speakers in an attempt to compensate for the reduced intelligibility by overarticulating. It was hypothesized that the implementation of CS would lead to a further increase in the extent of overarticulation and, subsequently, might lead to a higher intelligibility. Given the results in the present study, future research should assess acoustics, aerodynamics, and muscle activity (e.g., articulatory contact pressure) when Cantonese ES speakers are using CS to verify these possibilities.

TE speakers showed the smallest increase in CSIT intelligibility scores (1.4%) compared to EL users and ES speakers. This might be the result of TE speakers having an overall higher baseline intelligibility in the HS condition; the TE speakers' mean intelligibility score of 81.3% was higher than those who have achieved "maximal proficiency" as reported by Law et al. [14]. In fact, 8 out of 12 TE speakers had a CSIT intelligibility score over 85% in the present study. This "ceiling" of intelligibility in HS may suggest that the majority of the TE speakers in this study may have reached a level of "maximal proficiency." These findings are consistent with the systematic review by van Sluis et al. [21], in which TE speech was suggested to be the most intelligible among the 3 types of alaryngeal speech. Further, certain features of CS (i.e., overarticulation, slower rate) do not appear to benefit this speaker group when compared to EL users and ES speakers. This could be the result of TE speech production being supported by pulmonary air, which is similar to voice production for laryngeal speakers [18]. TE speakers do not need to coordinate their speech in a similar manner to ES speakers and EL users, and as a result, did not derive a similar CS benefit.

Several notable limitations of the present study should be acknowledged. To warrant a sufficient sample size, the proficiency of alaryngeal speakers was not experimentally controlled. Instead, recruitment was based on referrals from an experienced speech therapist at the New Voice Club of Hong Kong. Future research might consider requiring all alaryngeal speakers to use their form of alaryngeal speech for at least 2 years (e.g., Cox and Doyle [11]) alongside auditory-perceptual ratings (e.g., speech acceptability) from a panel of speech-language pathologists prior to the intelligibility assessment. Another possible limitation is the lack of longer and more refined training sessions to elicit CS. In the present study, CS was elicited through brief, prerecorded instructions and modeling by the primary investigator. Though similar verbal instructions used to elicit CS in healthy or speakers with speech impairments were used in previous studies [11, 36, 41, 42, 46, 47], it is possible that laryngectomees may require more time and training to produce CS. A CS training program that incorporates more detailed instructions, feedback, and intensive practice, such as that used by Krause and Braida [34] or Park et al. [45], should be attempted in the future to ensure successful acquisition of CS. Also, there is a possibility that some of the alaryngeal speakers were already using some form of CS in the HS condition. Future studies should assess additional acoustic features, such as vowel formants, to examine the difference of speech production in HS and CS conditions. Lastly, the current study only included male

alaryngeal speakers and 2 female listeners. Male alaryngeal speakers form the majority of members (>90%) at the New Voice Club of Hong Kong, and only 2 females responded to advertisements at the University of Hong Kong. Future research should include a larger heterogeneous group of speakers and listeners to confirm the potential benefits of CS in Cantonese alaryngeal speakers.

Conclusion

The present study suggests that CS had a significant effect on speaking rate and intelligibility for Cantonese alaryngeal speakers. While a statistically significant effect was not observed between alaryngeal communication methods, mean intelligibility scores increased for all forms of alaryngeal speech, suggesting a potential benefit of CS for Cantonese alaryngeal speakers. EL speech had the greatest increase in the CSIT intelligibility scores followed by ES and TE speakers. Overall, the preliminary data presented in this study serve as an initial step in understanding the effect of CS on Cantonese alaryngeal speakers' intelligibility. Further research involving larger, heterogeneous groups of speakers and listeners alongside longer and more refined CS training protocols must be conducted to confirm that CS can improve Cantonese alaryngeal speakers' intelligibility.

Funding Sources

This work was supported in part by National Institutes of Health Grant R01 DC-002717 to Haskins Laboratories.

Data Availability

All data generated and analyzed during this study are included in the article. Further enquiries can be directed to the corresponding author.

References

1. Silverman DA, Puram SV, Rocco JW, Old MO, Kang SY. Salvage laryngectomy following organ-preservation therapy – an evidence-based review. Oral Oncol. 2019 Jan;88:137–44. [PubMed: 30616784]
2. Chan JY. Practice of laryngectomy rehabilitation interventions: a perspective from Hong Kong. Curr Opin Otolaryngol Head Neck Surg. 2013 Jun;21(3):205–11. [PubMed: 23572016]
3. Eadie TL. The ICF: a proposed framework for comprehensive rehabilitation of individuals who use alaryngeal speech. Am J Speech Lang Pathol. 2003 May;12(2):189–97. [PubMed: 12828532]
4. Eadie TL. Application of the ICF in communication after total laryngectomy. Semin Speech Lang. 2007 Nov;28(4):291–300. [PubMed: 17935014]
5. Ng ML. Aerodynamic characteristics associated with oesophageal and tracheoesophageal speech of Cantonese. Int J Speech Lang Pathol. 2011 Apr;13(2):137–44. [PubMed: 21480810]
6. Cox SR. Review of the electrolarynx: the past and present. Perspect ASHA Spec Interest Groups. 2019 Feb;4(1):118–29.
7. Ng ML. The use of the Lombard Effect in Improving Alaryngeal Speech. J Voice. 2021 Jan;35(1):18–28. [PubMed: 31350113]
8. Doyle PC, Finchem EA. Teaching esophageal speech: a process of collaborative instruction. In: Doyle PC, editor. Clinical care and rehabilitation in head and neck cancer. Cham: Springer; 2019. p. 145–61.
9. Singer MI, Blom ED. An endoscopic technique for restoration of voice after laryngectomy. Ann Otol Rhinol Laryngol. 1980 Nov-Dec;89(6 Pt 1):529–33. [PubMed: 7458140]

10. Nagle KF. Elements of clinical training with the electrolarynx. In: Doyle PC, editor. Clinical care and rehabilitation in head and neck cancer. Cham: Springer; 2019. p. 129–43.
11. Cox SR, Doyle PC. The influence of clear speech on auditory-perceptual judgments of electrolaryngeal speech. J Commun Disord. 2018 Sep-Oct;75:25–36. [PubMed: 30005317]
12. Doyle PC, Eadie TL. The perceptual nature of alaryngeal voice and speech. In: Doyle PC, editor. Contemporary considerations in the treatment and rehabilitation of head and neck cancer. Texas: Pro-Ed; 2005. p. 113–40.
13. Globlek D, Stajner-Katusic S, Musura M, Horga D, Liker M. Comparison of alaryngeal voice and speech. Logoped Phoniatr Vocol. 2004;29(2):87–91. [PubMed: 15260185]
14. Law IK, Ma EP, Yiu EM. Speech intelligibility, acceptability, and communication-related quality of life in Chinese alaryngeal speakers. Arch Otolaryngol Head Neck Surg. 2009 Jul;135(7):704–11. [PubMed: 19620593]
15. Most T, Tobin Y, Mimran RC. Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. J Commun Disord. 2000 Mar-Apr;33(2): 165–80. [PubMed: 10834832]
16. Ng ML, Chu R. An acoustical and perceptual study of vowels produced by alaryngeal speakers of Cantonese. Folia Phoniatr Logop. 2009;61(2):97–104. [PubMed: 19299898]
17. Ng ML, Gilbert HR, Lerman JW. Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech. Folia Phoniatr Logop. 2001 Jan-Feb;53(1):36–47. [PubMed: 11125259]
18. Ng ML, Kwok CL, Chow SF. Speech performance of adult cantonese-speaking laryngectomees using different types of alaryngeal phonation. J Voice. 1997 Sep;11(3):338–44. [PubMed: 9297679]
19. Robbins J, Fisher HB, Blom EC, Singer MI. A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. J Speech Hear Disord. 1984 May;49(2):202–10. [PubMed: 6716991]
20. Moukarbel RV, Doyle PC, Yoo JH, Franklin JH, Day AM, Fung K. Voice-related quality of life (V-RQOL) outcomes in laryngectomees. Head Neck. 2011 Jan;33(1):31–6. [PubMed: 20848430]
21. van Sluis KE, van der Molen L, van Son RJ, Hilgers FJ, Bhairosing PA, van den Brekel MW. Objective and subjective voice outcomes after total laryngectomy: a systematic review. Eur Arch Otorhinolaryngol. 2018 Jan;275(1):11–26. [PubMed: 29086803]
22. Bennett S, Weinberg B. Acceptability ratings of normal, esophageal, and artificial larynx speech. J Speech Hear Res. 1973 Dec;16(4):608–15. [PubMed: 4783798]
23. Pindzola RH, Cain BH. Acceptability ratings of tracheoesophageal speech. Laryngoscope. 1988 Apr;98(4):394–7. [PubMed: 3352438]
24. Gandour J, Weinberg B. Perception of intonational contrasts in alaryngeal speech. J Speech Hear Res. 1983 Mar;26(1):142–8. [PubMed: 6865370]
25. Gandour J, Weinberg B. Production of intonation and contrastive stress in electrolaryngeal speech. J Speech Hear Res. 1984 Dec;27(4):605–12. [PubMed: 6521468]
26. Gandour J, Weinberg B, Petty SH, Dardarananda R. Tone in Thai alaryngeal speech. J Speech Hear Disord. 1988 Feb;53(1):23–9. [PubMed: 3339865]
27. Liu H, Ng ML. Electrolarynx in voice rehabilitation. Auris Nasus Larynx. 2007 Sep;34(3):327–32. [PubMed: 17239553]
28. Liu H, Ng ML, Wan M, Wang S, Zhang Y. The effect of tonal changes on voice onset time in Mandarin esophageal speech. J Voice. 2008 Mar;22(2):210–8. [PubMed: 17055221]
29. Ng ML, Chan MW. Analyzing neoglottal vibration of Cantonese tracheoesophageal speech: preliminary aerodynamic study using inverse filtering. Folia Phoniatr Logop. 2012;64(6):283–9. [PubMed: 23429237]
30. Ng ML, Lerman JW, Gilbert HR. Perceptions of tonal changes in normal laryngeal, esophageal, and artificial laryngeal male Cantonese speakers. Folia Phoniatr Logop. 1998;50(2):64–70. [PubMed: 9624857]
31. Ng ML, Xiong MY. Chinese alaryngeal speech rehabilitation and their acoustical characteristics: A comprehensive review. Rehabil Med. 2015;25(2):44–9.

32. Yan N, Lam PK, Ng ML. Pitch control in esophageal and tracheoesophageal speech of Cantonese. Folia Phoniatr Logop. 2012;64(5):241–7. [PubMed: 23051971]
33. Lo A. Intelligibility and acceptability measures of Cantonese dysarthric speech [dissertation]. Hong Kong, University of Hong Kong, 2015.
34. Krause JC, Braida LD. Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility. J Acoust Soc Am. 2002 Nov;112(5 Pt 1):2165–72. [PubMed: 12430828]
35. Payton KL, Uchanski RM, Braida LD. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. J Acoust Soc Am. 1994 Mar;95(3):1581–92. [PubMed: 8176061]
36. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech. J Speech Hear Res. 1985 Mar;28(1):96–103. [PubMed: 3982003]
37. Smiljanić R, Bradlow AR. Speaking and hearing clearly: talker and listener factors in speaking style changes. Lang Linguist Compass. 2009 Jan;3(1):236–64. [PubMed: 20046964]
38. Whitfield JA, Goberman AM. Articulatory-acoustic vowel space: associations between acoustic and perceptual measures of clear speech. Int J Speech Lang Pathol. 2017 Apr;19(2):184–94. [PubMed: 27328115]
39. Uchanski RM. Clear speech. In: Pisoni DB, Remez RE, editors. Handbook of speech perception. Malden: Blackwell Publishers; 2005. p. 207–35.
40. Krause JC, Braida LD. Acoustic properties of naturally produced clear speech at normal speaking rates. J Acoust Soc Am. 2004 Jan;115(1):362–78. [PubMed: 14759028]
41. Lam J, Tjaden K. Intelligibility of clear speech: effect of instruction. J Speech Lang Hear Res. 2013 Oct;56(5):1429–40. [PubMed: 23798509]
42. Lam J, Tjaden K. Clear speech variants: an acoustic study in Parkinson's disease. J Speech Lang Hear Res. 2016 Aug;59(4):631–46. [PubMed: 27355431]
43. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech. J Speech Hear Res. 1986 Dec;29(4):434–46. [PubMed: 3795886]
44. Beukelman DR, Fager S, Ullman C, Hanson E, Logemann J. The impact of speech supplementation and clear speech on the intelligibility and speaking rate of people with traumatic brain injury. J Med Speech-Lang Pathol. 2002 Dec;10(4):237–42.
45. Park S, Theodoros D, Finch E, Cardell E. Be clear: A new intensive speech treatment for adults with nonprogressive dysarthria. Am J Speech Lang Pathol. 2016 Feb;25(1):97–110. [PubMed: 26882004]
46. Tjaden K, Sussman JE, Wilding GE. Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. J Speech Lang Hear Res. 2014 Jun;57(3):779–92. [PubMed: 24687042]
47. Cox SR, Raphael LJ, Doyle PC. Production of vowels by electrolaryngeal speakers using clear speech. Folia Phoniatr Logop. 2020;72(4):250–6. [PubMed: 31121594]
48. Boersma P, Weenink D. Praat: Doing phonetics by computer [computer program on the internet]. Version 6.1.07. Amsterdam [cited 2021 Jan 26]. Available from: www.praat.org.
49. McAuliffe M, Socolof M, Mihuc S, Wagner M, Sonderegger M. Montreal Forced Aligner: trainable text-seech alignment using kaldi. Interspeech. 2017 Aug;2017:498–502.
50. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Statistic Soft. 2015 Oct;67(1). DOI: 10.18637/jss.v067.i01..
51. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [cited 2021 May 22]. Available from: http://www.R-project.org.
52. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J. 2008 Jun;50(3):346–63. [PubMed: 18481363]
53. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics. 2014 May;30(9):1198–204. [PubMed: 24443148]

54. Cohen JW. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates; 1988.

55. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016 Jun;15(2):155–63. [PubMed: 27330520]

56. Cox SR. The application of clear speech in electrolaryngeal speakers [dissertation]. Ontario, Western University, 2016.

57. Chen Y, Ng M. Cantonese lexical tone production by Cantonese speakers using inspiratory phonation. Poster presented at: Annual Convention of the American Speech-Language-Hearing Association; 2014 Nov 20–22; Orlando, FL.

58. Ng M, Lam K, Chen Y. Perception of whispered Cantonese tones. In: Proceedings of the 12th Phonetic Conference of China (PCC2016) (p. 746–50), Inner Mongolia, China, 2016.

59. Ng M, Xiong M, Yan N. Production of Cantonese tones using an ingressive airflow. Poster presented at: International Conference on Asian Language Processing; 2018 Nov 15–18; Bandung, Indonesia.

60. Liu H, Ng ML, 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(3):147–54. [PubMed: 17556858]

61. Ng ML, Tong ET, Yu KM. Articulatory contact pressure during bilabial plosive production in esophageal and tracheoesophageal speech. Folia Phoniatr Logop. 2019;71(1):1–6. [PubMed: 30466101]

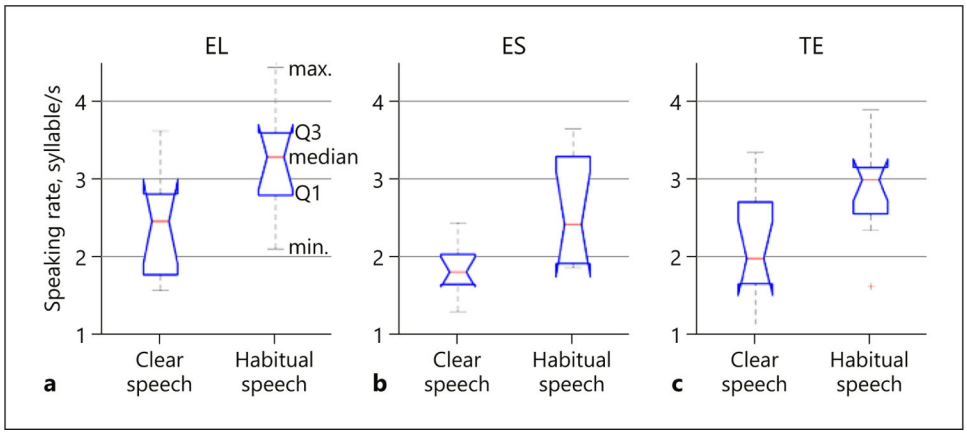

Fig. 1.

Boxplots of speaking rate in syllables per second for each speaker group. **a** Electrolaryngeal (EL) speech. **b** Esophageal (ES) speech. **c** Tracheoesophageal (TE) speech.

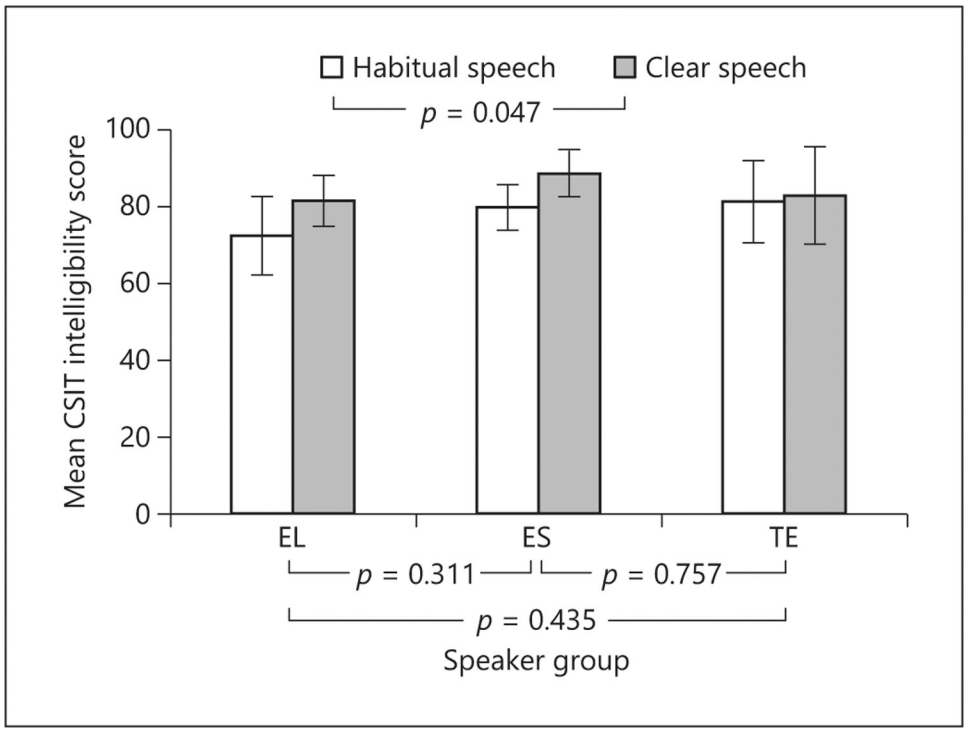

Fig. 2.

Mean CSIT intelligibility score (%) of electrolaryngeal (EL), esophageal (ES) and tracheoesophageal (TE) speakers under different speaking conditions. Error bars represent ±1.96 SE as estimate of 95% CI for the mean.

Table 1.

Demographic information of the alaryngeal speakers

Speaker group	Number of participants	Age, years	Duration of use of alaryngeal speech, years
EL	9	68.33 (7.81)	8.05 (6.89)
ES	10	61.30 (11.25)	5.88 (5.88)
TE	12	69.42 (12.82)	6.33 (7.72)

Age and duration data are reported in means and SDs. EL, electrolarynx; ES, esophageal; TE, tracheoesophageal.

Table 2.

Mean (SD) of CSIT intelligibility score under different speaking conditions

Speaker group	Intelligibility score, %		Change, %
	HS condition	CS condition	
EL	72.4 (20.5)	81.5 (13.3)	+9.1
ES	79.8 (11.9)	88.6 (12.0)	+8.8
TE	81.3 (21.4)	82.8 (25.1)	+1.5

EL vs. ES: $p = 0.311$, ES vs. TE: $p = 0.757$, EL vs. TE: $p = 0.435$; HS vs. CS: $p = 0.47$, $p < 0.05$ considered as statistically significant. HS, habitual speech; CS, clear speech; EL, electrolarynx; ES, esophageal; TE, tracheoesophageal.