Articulatory Strategies in Obstruent Production in Mandarin Esophageal Speech

Fang Hu[1], *Yungang Wu*[2], *Wen Xu*[2*], *Demin Han*[2]

[1]Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
[2]Department of Otolaryngology, Beijing Tongren Hospital, Beijing, China
[1]hufang@cass.org.cn, [2*]xuwenlily@yahoo.com.cn

Abstract

Based on the comparison between 4 esophageal speakers and 4 normal laryngeal speakers, this study investigated the voice onset time (VOT) characteristics and the linguopalatal articulation in the production of Mandarin obstruent consonants. Results show that esophageal speakers distinguish unaspirated vs. aspirated plosives or affricates in a similar way as laryngeal speakers do. However, the aspirated plosives and affricates have a shorter VOT whereas the unaspirated plosives and affricates have a longer VOT in esophageal speech than in laryngeal speech. Interestingly, esophageal speech exhibits a significantly more extensive linguopalatal contact than normal speech does. Results suggest that articulatory strategies have been adjusted to facilitate the linguopalatal articulation as well as the sub-to-supra-laryngeal coordination by using a narrower air way in the production of esophageal speech.

Index Terms: articulatory strategies, obstruent production, Mandarin esophageal speech

1. Introduction

Esophageal speech is a commonly adopted alaryngeal speech after total laryngectomy. A main concern after the operation is how to improve the intelligibility of esophageal speech ([1]; [2]). Using the pharyngoesophageal (PE) segment as the neoglottis, esophageal speech has a significantly different voice quality from normal laryngeal speech. Previous studies have examined various aspects of esophageal speech, such as aerodynamics ([3]; [4]), acoustics including fundamental frequency (F0), intensity, voice quality, and resonance frequencies ([5]; [6]; [7]), and the perception ([8]; [9]; [10]). It has been found that esophageal speech is typically characterized by a perceptually hoarse voice quality, and significantly reduced airflow volume, fundamental frequency (F0), intensity, and speech rate.

However, previous studies mainly focused on the acoustic and perceptual aspects of esophageal speech and less attention has been paid on the articulation of esophageal speech. This study examined the temporary coordination between the sub- and supra-systems of speech organs and the linguopalatal articulation in the obstruent production in Mandarin Chinese. It is assumed that since the use of the PE neoglottis dramatically changes both the phonatory setting and speech airstream mechanisms, speech motor system has to be reprogrammed during the articulation of esophageal speech.

2. Methodology

Four Mandarin esophageal speakers, 2 male and 2 female, participated in this study. They were all born, grew up, and were living in Beijing. They aged 57 to 63 years old, with an average of 60 years old, at the time of recording. They had been trained to use esophageal speech for 2 to 10 years after total laryngectomy. The intelligibility of their esophageal speech is above 95%, according to the test conducted at Beijing Tongren hospital by using the Chinese Intelligibility Wordlist ([11]).

Four normal Mandarin speakers, 2 male and 2 female, were used as control. They were all Beijing natives and aged 30 to 45 years old at the time of recording. They had normal dentition and had no reported history of speech and hearing problems.

Meaningful monosyllables containing the target obstruents were used as test syllables. That is, [t t^h k k^h ts ts^h tʂ tʂ^h tɕ tɕ^h s ʂ ɕ] for the linguopalatal articulation study and [p p^h t t^h k k^h ts ts^h tʂ tʂ^h tɕ tɕ^h] for the VOT study. When possible, test syllables have a high level tone with a balanced vowel environment [i] [u] or [a]. That is, there are totally three test syllables for each target obstruent. The test syllable was embedded in a carrier frame [X, wo tu X tsʅ] '*lit.* X, I read X Chinese character'. Five repetitions were recorded for each test syllable.

The electropalatograph, the WinEPG system ([12]), was used for the recording of linguopalatal articulation. Each speaker wore a customized pseudo-palate during the recording. An example of the pseudo-palate and the schematic arrangement of the 62 electrodes on the pseudo-palate were shown in Figure 1.

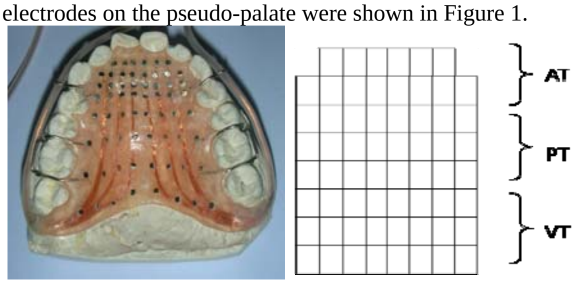


Figure 1: *The customized pseudo-palate (left) and the schematic arrangement of the 62 electrodes (right).*

Each square represents for a silver electrode, and signals will be recorded when the tongue contacts with any sensors on the pseudo-palate. The two anterior rows were defined as alveolar region; the three following rows were defined as post-alveolar region; and the three posterior rows were defined as velar region. In the analysis, the linguopalatal contact totals over the alveolar region (AT), the postalveolar region (PT), the velar region (VT), and the whole palate (WT) were measured. Besides, center of gravity (CoG) was calculated to examine the anteriority of the contact; the mean lateral measure (MLM) was defined to examine whether there is more contact close to the midline or towards the sides of the palate; the mean asymmetry measure (MAM) was designed to examine whether there is more contact towards one side or the other ([13]).

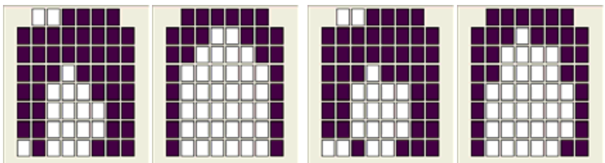
Separate audio recordings were conducted for the VOT study. VOT was measured directly from the audio waveform with reference to the corresponding wideband spectrogram.

3. Results

3.1. Linguopalatal articulation

3.1.1. *stops*

As shown in Figure 2, the articulation of [t t^h] (/d t/ in Pinyin) mainly involves linguopalatal contact in the alveolar region for laryngeal speakers. But for esophageal speakers, both alveolar and postalveolar regions are contacted. Moreover, esophageal speakers have more contacts in velar region than laryngeal speakers. In other words, esophageal speakers exhibit a generally heavier linguopalatal contact pattern than laryngeal speakers. This is also clearly demonstrated by the results from quantitative analyses in Figure 3. Esophageal speakers have a significantly greater mean PT, VT, and WT than laryngeal speakers. This results in a significantly greater CoG, i.e. more anteriority in control than esophageal speakers. Laterally, the contact pattern is closer to the midline in esophageal speakers than in laryngeal speakers. But this doesn't suggest a different control in laterality between the two groups. Instead, this can simply be attributed to the heavier contact pattern in esophageal speakers, as both groups share a similar mean lateral symmetry pattern.



d-eso. d-ctr. t-eso. t-ctr.

Figure 2: *Linguopalatal contacts for /d/ ([t̥]) and /t/ ([t^h]): esophageal (eso.) vs. control (ctr.) speakers.*

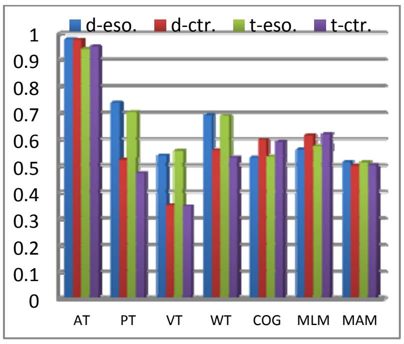
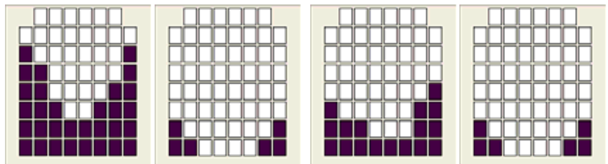


Figure 3: *Mean AT, PT, WT, CoG, MLM and MAM for the alveolar stops /d/ ([t̥]) and /t/ ([t^h]) in esophageal (eso.) and laryngeal (ctr.) speakers; n=120. Paired t-tests yielded significant difference between eso. and ctr. speakers at 95% confidence level for all indices except AT and MAM.*

As shown in Figure 4, the articulation of velar stops [k k^h] (/g k/) mainly involves linguopalatal contact in the velar region for laryngeal speakers. But for esophageal speakers, a significantly heavier contact pattern is detected, and the contact extends from velar region to postalveolar region or sometimes even alveolar region. And this is also demonstrated by AT, PT, VT and WT in Figure 5. As a result, the velar stops in esophageal speakers are anterior to those in laryngeal speakers, as manifested by a significant greater CoG value in esophageal speakers than in laryngeal speakers.

Although velar stops are fronted and alveolar stops are backed in terms of CoG, velar and alveolar stops definitely have different linguopalatal contact patterns in esophageal speakers. In other words, esophageal speakers can well distinguish velar stops from alveolar stops, although with heavier linguopalatal contact patterns than laryngeal speakers.

Again, as can be seen from Figure 5, the linguopalatal contact pattern is closer to the midline in the production of velar stops for esophageal speakers than for laryngeal speakers (MLM); and the lateral symmetry pattern is similar in both groups (MAM).



g-eso. g-ctr. k-eso. k-ctr.

Figure 4: *Linguopalatal contacts for /g/ ([k̥]) and /k/ ([k^h]): esophageal (eso.) vs. control (ctr.) speakers.*

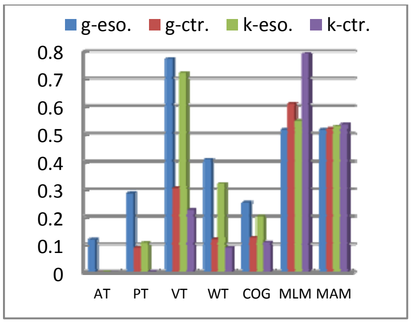
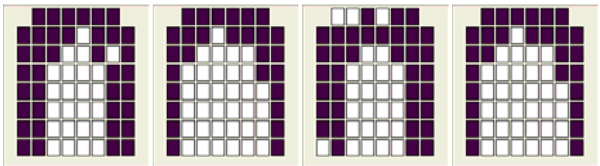


Figure 5: *Mean AT, PT, WT, CoG, MLM and MAM for the velar stops /g/ ([k̥]) and /k/ ([k^h]) in esophageal (eso.) and laryngeal (ctr.) speakers; n=120. Paired t-tests yielded significant difference between eso. and ctr. speakers at 95% confidence level for all indices except AT and MAM for both stops and MLM for/g/.*

3.1.2. *affricates*

As shown in Figure 6 and 7, alveolar affricates demonstrate a similar pattern as alveolar stops do. First, esophageal speech exhibits a heavier linguopalatal contact pattern than laryngeal speech does. Second, according to CoG, alveolar affricates in esophageal speech are less anterior than those in laryngeal speech. Third, the lateral symmetry pattern is similar in both groups (MAM), while the contact pattern is laterally closer to the midline in esophageal speakers than in laryngeal speakers (MLM).



z-eso. z-ctr. c-eso. c-ctr.

Figure 6: *Linguopalatal contacts for /z/ ([ts]) and /c/ ([tsʰ]): esophageal (eso.) vs. control (ctr.) speakers.*

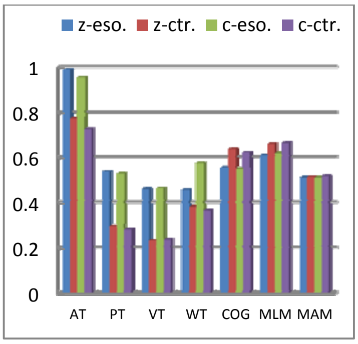
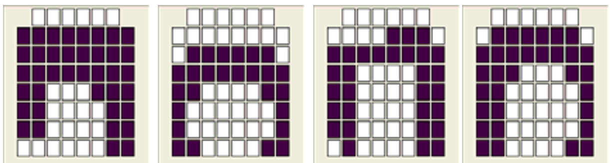


Figure 7: *Mean AT, PT, WT, CoG, MLM and MAM for the alveolar affricates /z/ ([ts]) and /c/ ([tsʰ]) in esophageal (eso.) and laryngeal (ctr.) speakers; n=120. Paired t-tests yielded significant difference between eso. and ctr. speakers at 95% confidence level for all indices except MAM.*



zh-eso. zh-ctr. ch-eso. ch-ctr.

Figure 8: *Linguopalatal contacts for /zh/ ([tʂ]) and /ch/ ([tʂʰ]): esophageal (eso.) vs. control (ctr.) speakers.*

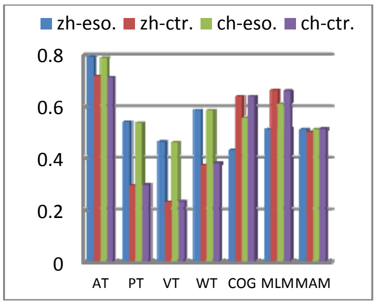
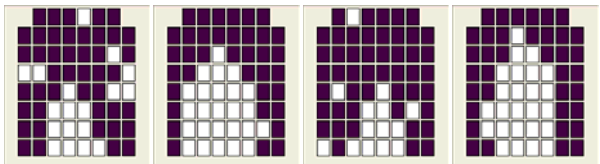


Figure 9: *Mean AT, PT, WT, CoG, MLM and MAM for the postalveolar affricates /zh/ ([tʂ]) and /ch/ ([tʂʰ]) in esophageal (eso.) and laryngeal (ctr.) speakers; n=120. Paired t-tests yielded significant difference between eso. and ctr. speakers at 95% confidence level for all indices except MAM.*

It can be seen from Figure 8 that the production of postalveolar affricates clearly involves a linguopalatal contact pattern posterior to that of alveolar affricate. This is true for both esophageal and laryngeal speakers. The difference is that esophageal speech demonstrates a heavier contact pattern than laryngeal speech. Again, as shown in Figure 9, postalveolar affricates in esophageal speech (1) are less anterior than those in laryngeal speech, (2) are laterally closer to the midline than those in laryngeal speech, and (3) have a similar lateral symmetry as those in laryngeal speech.

As can be seen from Figure 10, the articulation of palato-alveolar affricates in both control and laryngeal speakers exhibits a linguopalatal contact pattern in both alveolar and postalveolar region. This suggests that the constrict location of palato-alveolar affricates is the whole area from alveolar ridge to the palate, rather than somewhere in-between ([14]). The difference is that esophageal speech demonstrates a heavier contact pattern than laryngeal speakers. Again, as compared to laryngeal speech, esophageal speech (1) has more linguopalatal contacts, (2) is less anterior (CoG), and (3) is laterally closer to the midline (MLM). And both groups share a similar lateral symmetry pattern.



j-eso. j-ctr. q-eso. q-ctr.

Figure 10: *Linguopalatal contacts for /j/ ([tɕ]) and /q/ ([tɕʰ]): esophageal (eso.) vs. control (ctr.) speakers.*

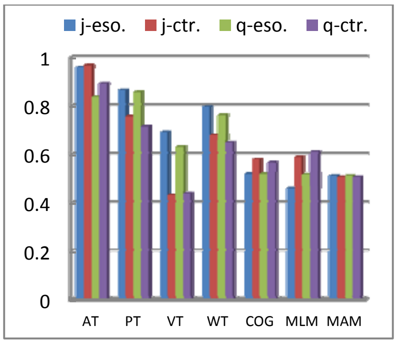


Figure 11: *Mean AT, PT, WT, CoG, MLM and MAM for the palato-alveolar affricates /j/ ([tɕ]) and /q/ ([tɕʰ]) in esophageal (eso.) and laryngeal (ctr.) speakers; n=120. Paired t-tests yielded significant difference between eso. and ctr. speakers at 95% confidence level for all indices except MAM for both affricates and AT for /j/.*

3.1.3. *fricatives*

The distinction between the alveolar and postalveolar fricative is comparable as that between their affricate counterparts. In short, the alveolar fricative shows a linguopalatal configuration anterior to that of the postalveolar fricative. In general, there are more linguopalatal contacts in esophageal speech than in control speech. And consequently, the linguopalatal configuration in esophageal speech is less anterior than that in control speech for the alveolar fricative, but vice versa for the postalveolar fricative (CoG).

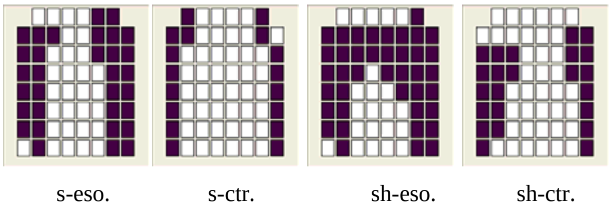


Figure 12: *Linguopalatal contacts for /s/ ([s]) and /sh/ ([ʂ]): esophageal (eso.) vs. control (ctr.) speakers.*

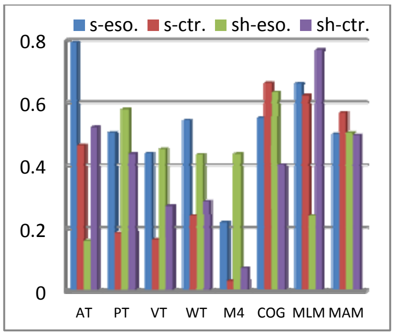


Figure 13: *Mean AT, PT, WT, CoG, MLM and MAM for the fricatives /s/ ([s]) and /sh/ ([ʂ]) in esophageal (eso.) and laryngeal (ctr.) speakers; n=120. Paired t-tests yielded significant difference between eso. and ctr. speakers at 95% confidence level for all indices except MLM for /s/ and MAM for /sh/.*

3.2. VOT

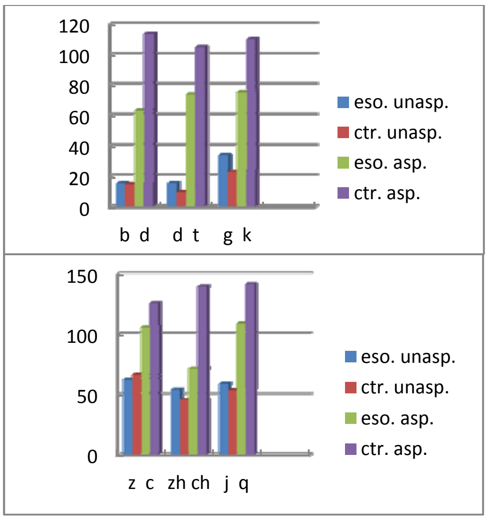


Figure 14: *Mean voice onset time (VOT) in millisecond (n=120) for stops (upper) and affricates (lower).*

Figure 14 summarized the mean voice onset time (VOT) from esophageal vs. control speech. It is clear from the figure that the distinction between aspirated and unaspirated stops or affricates is well manifested by differences in VOT in both esophageal and control speech. That is, aspirated consonants have much greater VOTs than their unaspirated counterparts in both groups. However, it should also be noted that aspirated consonants in esophageal speech have comparatively smaller VOTs than those in control speech, while unaspirated consonants, in general, have relatively greater VOTs than those in control speech. This suggests that the coordination between sub- and supra-systems of articulation is not so well controlled in esophageal speech as in laryngeal speech, since esophageal speakers have to reprogram the sub-to-supra-laryngeal coordination due to the significantly different vibratory behavior of the pharyngoesophageal (PE) segment vis-à-vis vocal folds.

4. Conclusion

Phonological contrasts are well maintained and phonetically implemented in the obstruent production in Mandarin esophageal speech. However, esophageal speakers intend to employ a significantly heavier linguopalatal configuration in obstruent production to compensate for their extremely limited airstream mechanism. And the results also suggest that the coordination between sub- and supra-systems of articulation is reprogrammed in esophageal speech due to the significantly different vibratory behavior of the PE segment vis-à-vis vocal folds.

5. References

[1] Bennett S. and Winberg B., "Acceptability ratings of normal, esophageal, and artificial laryngeal speech", J. Speech Hear. Res., 16:608-615, 1973.

[2] Holly, S. C., Lerman, J. and Randolph, K. A., "A comparison of the intelligibility of esophageal, electrolarynx, and normal speech in quiet and in noise", J. Commun. Disord., 16:143-155, 1983.

[3] Motta, S., Galli, I. and Di Rienzo, L., "Aerodynamic findings in esophageal voice", Arch Otolaryngol Head Neck Surg, 127:700-704, 2001.

[4] Snidecor, J. C. and Isshiki, N., "Air Volume and air flow relationships of six esophageal speakers", J. Speech Hear. Disord., 30:205-216, 1965.

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

[6] Liu, H. and Ng, M., "Formant characteristics of vowels produced by Mandarin esophageal speakers", J. Voice, 23:255-260, 2009.

[7] Robbins, J., Fisher, H. B., Blom, E. C. and Mark, I. S., "A comparative acoustic study of normal, esophgael, and tracheoesophageal speech production", J. Speech Hear. Disord., 49:202-210, 1984.

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

[9] Ng, M., Lerman, J. W. and Gilbert, H. R., "Perceptions of tonal changes in normal laryngeal, esophageal, and artificial laryngeal male Cantonese speakers", Folia Phoniatr., 50:64-70, 1998.

[10] Ng, M., Gilbert, H. R. and Lerman, J. W., "Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech", Folia Phoniatr., 53:36-47, 2001.

[11] Wang, G., Zhu, C., Yuan, W. and Jiang, L., "Chinese wordlist for intelligibility test and its clinical application [in Chinese]", Shanghai J. Stomatology", 4(3):125-127, 1995.

[12] http://www.articulateinstruments.com/

[13] Articulate Instruments Ltd. "Articulate Assistant User Guide: Version 1.18", Edinburgh, UK: Articulate Instruments Ltd, 2010.

[14] Hu, F. "The three sibilants in Standard Chinese", Proc. 8th ISSP, Strasbourg, France, 2008.

(In Proceedings of Interspeech 2012, Mon. 01c. 01, Portland, Oregon, Spetember 9-13, 2012.)