A preliminary study estimating the articulatory space of Cantonese-sepaking healthy adults using ultrasound tongue imaging

Min Ney WONG[1-4], Bruce Xiao WANG[1,5], Yu SUN[1,6] Zhen SONG[1,6], Ho Yee NG[4], Hang Ching LAM[1,4] & Yong Ping ZHENG[1,6]

[1]Research Institute for Smart Ageing / [2]Research Centre for Language, Cognition, and Neuroscience / [3]The HK PolyU-PekingU Research Centre on Chinese Linguistics / [4]Department of Chinese and Bilingual Studies / [5]Department of English and Communication / [6]Department of Biomedical Engineering, Hong Kong Polytechnic University

Ultrasound tongue imaging (UTI) has been extensively employed in studying tongue movement during speech production in various areas, such as sociophonetics (Lawson et al., 2008; Liu et al., 2023) and speech and language therapy (SLT). For example, Cleland (2023) showcases the advantages of using UTI in research and clinical practice for people with cleft lip and palate, and Campos & Ristau (2022) investigated the effectiveness of UTI biofeedback training for speech intervention. However, previous studies have primarily focused on English and other European languages such as French (Laporte & Ménard, 2018), Persian (Baghban et al., 2020) and Hungarian (Csapó & Xu, 2020). The current study is the first of its kind aiming to estimate the articulatory space of Cantonese-speaking healthy adults using UTI.

The current on-going project aims to recruit 5 male and 5 female heathy Hong Kong Cantonese speakers ranging from 18 to 30 years old. Single word stimuli[1] were designed to elicit the articulatory gestures of target segments, namely consonants (C) and vowels (V), covering three monophthongs, four diphthongs and two consonants (/t/ and /k/). Each CV combination has at least 5 repetitions. The designed target words with monophthongs (/i, a, u/) allow us to capture the most extreme configurations of one's articulatory space, and those with diphthongs (e.g., /ai/, /au/) which capture the movement and the transition between the extreme configurations. The alveolar (/t/) and velar (/k/) plosives cover the most anterior and posterior lingual-oral constrictions in Cantonese.

UTI was collected using SuperSonic Aixplorer collecting 60 frames per second during speech production. Acoustic signal (speech recordings) was simultaneously collected using UGREEN CM592 microphone. Target words and vowels were firstly located in speech recordings using a *TextGrid* in Praat (Boersma & Weenink, 2023), the time information was then retrieved to extract corresponding ultrasound images (USIs). Tongue contours were automatically fitted using a U-Net model from (Zhu et al., 2019) with manual correction by a medical sonographer. Figure 1 (upper panel) shows the tongue contour producing three monophthongs from one male speaker, solid contours are smoothed using polynomial fitting and shaded contours represent each production. Horizontal dashed line from -4 to 0 on x-axis indicates the trace of the bite plate used at the beginning of each recording session, and later used for rotation to make tongue contours more comparable (0 point is the location of upper incisors). The fitted tongue contours were processed by extracting two parameters, namely the highest point (i.e., TH; tongue height) and its corresponding x-coordinate (i.e., TA; tongue advancement). Figure 1 (lower panel) shows the TH-TA plot of three monophthongs from the same male speaker, representing the articulatory vowel space. Both the tongue contour and articulatory vowel space show some degree of within-speaker variation. More speakers are being recorded and analysed. Detailed data collection and USIs processing methods as well as results will be presented and discussed in the conference.

[1] Diadochokinesis tasks and free speech were also used for speech production, but only monophthongs are presented here due to limited space.

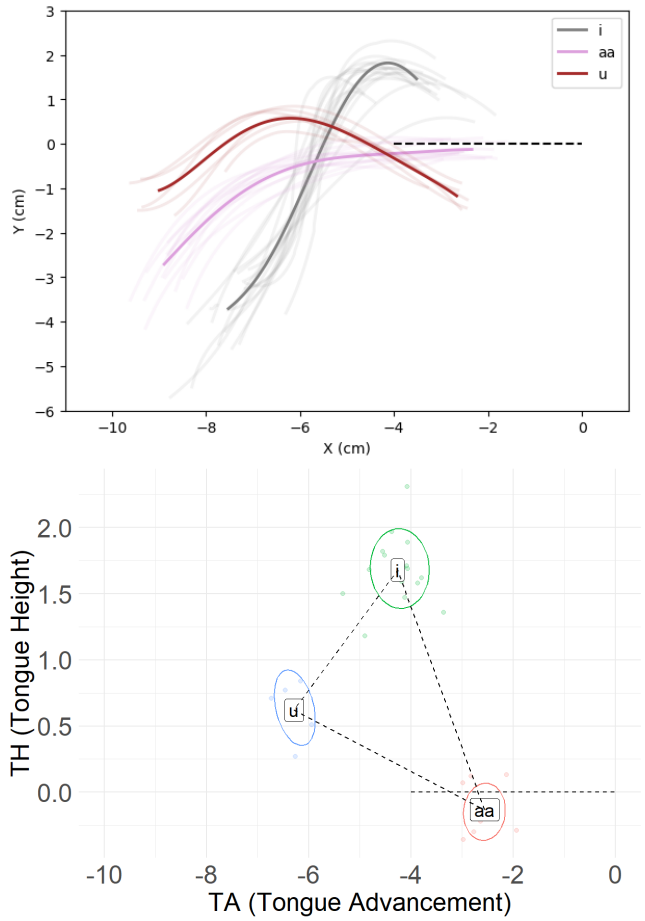

Figure 1. Tongue contour (upper panel) and vowel space (lower panel) from one Cantonese male speaker.

References

Baghban, K., Zarifian, T., Adibi, A., Shati, M., & Derakhshandeh, F. (2020). The quantitative ultrasound study of tongue shape and movement in normal Persian speaking children. *International Journal of Pediatric Otorhinolaryngology*, *134*, 110051. https://doi.org/10.1016/j.ijporl.2020.110051

Boersma, P., & Weenink, D. (2023). *Praat:Praat: Doing phonetics by computer [Computer program]*. (6.3.08) [Computer software]. http://www.praat.org/

Campos, A. R., & Ristau, J. (2022). Effectiveness of an Ultrasound Visual Biofeedback Training for Tongue Shape Assessment During Speech Sound Production. *Language, Speech, and Hearing Services in Schools*, *53*(3), 825–836. https://doi.org/10.1044/2022_LSHSS-21-00102

Cleland, J. (2023). Ultrasound Tongue Imaging in Research and Practice with People with Cleft Palate ± Cleft Lip. *The Cleft Palate Craniofacial Journal*, 10556656231202448. https://doi.org/10.1177/10556656231202448

Csapó, T. G., & Xu, K. (2020). Quantification of Transducer Misalignment in Ultrasound Tongue Imaging. *Interspeech 2020*, 3735–3739. https://doi.org/10.21437/Interspeech.2020-1672

Laporte, C., & Ménard, L. (2018). Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech. *Medical Image Analysis*, *44*, 98–114. https://doi.org/10.1016/j.media.2017.12.003

Lawson, E., Stuart-Smith, J., & Scobbie, J. M. (2008). *Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English*. https://test-eresearch.qmu.ac.uk/handle/20.500.12289/142

Liu, Y., Tong, F., Boer, G. de, & Gick, B. (2023). Lateral tongue bracing as a universal postural basis for speech. *Journal of the International Phonetic Association*, *53*(3), 712–727. https://doi.org/10.1017/S0025100321000335

Zhu, J., Styler, W., & Calloway, I. (2019). *A CNN-based tool for automatic tongue contour tracking in ultrasound images* (arXiv:1907.10210). arXiv. https://doi.org/10.48550/arXiv.1907.10210