

Geniohyoid Thickness Variation Across Cantonese Consonants: Evidence from Ultrasound Imaging

Shuming Huang¹, Bruce Xiao Wang^{1,4}, Alisher Myrgyyassov², Yu Sun², Min Ney Wong^{3,4}, Yongping Zheng^{2,4}

¹English and Communication Department, Hong Kong Polytechnic University, Hong Kong, China

²Biomedical Engineering Department, Hong Kong Polytechnic University, Hong Kong, China

³Department of Language Science and Technology, Hong Kong Polytechnic University, Hong Kong, China

⁴Research Institute for Smart Ageing, Hong Kong Polytechnic University, Hong Kong, China

The tongue's role as a central articulator has been widely explored using ultrasonography, with most studies focusing on tongue surface movement [1–3]. For instance, [2] identified a cross-linguistic lateral bracing posture in six languages, while [3] showed that coronal consonants in the Arrernte language demonstrated distinctive tongue root positions in different manners of articulation. However, as a complex muscular organ, the tongue's shape and motion result from the coordinated contraction of multiple muscle groups. Activities of these muscle groups, such as changes in muscle thickness, cannot be fully captured by imaging the tongue surface alone [4]. Early electromyographic research [5] revealed different activation patterns in tongue muscles, including the geniohyoid (GH), during Japanese consonant production. Yet, tongue muscle activity remains underexplored in ultrasound studies. This study addresses this gap by examining GH thickness during Cantonese consonant production using ultrasound imaging.

Ten native Cantonese speakers (5 females, 5 males, mean age: 24) participated in the study. Each produced three Cantonese onset fricatives (/f/, /s/, /h/) and three Cantonese onset nasals (/m/, /n/, /ŋ/) under a same vowel-consonant-vowel (aCa) context, all with tone 1. Each participant had four repetitions per consonant. Ultrasound images were collected using the *SuperSonic Imagine Aixplorer*® system and a curved probe. Audio was simultaneously recorded with a Shure SM58 microphone positioned approximately 30 cm from the participant. Each fricative and nasal token was segmented using Praat [6] to determine its duration and midpoint.

We used the UltraUNet [7] deep learning model for GH segmentation and thickness measurement. The model was trained on 1650 manually labelled GH image frames, using a combined loss function incorporating focal loss [8] and Dice loss (80% training; 20% validation; 30 training epochs; Intersection over Union = 0.904 ± 0.009 STD). The trained model was then applied to the central two frames of each fricative and nasal production. Each frame of interest was eyeball examined. Frames were excluded if the shadows of the mandible or hyoid bone were not clearly visible, if any portion of the GH muscle was absent, or if the segmentation of the GH was inaccurate. Each remaining frame was first processed by the trained segmentation network to produce a binary mask of the GH (Figure 1, top left). The largest connected component was then extracted, followed by edge smoothing (Figure 1, top right). Next, the mask was skeletonised [9] to facilitate further analysis (Figure 1, bottom left). Finally, GH thickness was calculated as twice the mean distance from the skeletal reference line to the boundary of the segmented mask (Figure 1, bottom right).

Descriptive statistics for geniohyoid (GH) thickness across Cantonese consonants and genders are presented in Table 1. Among female speakers, mean GH thickness ranged from 6.15 mm (SD = 0.60) for /f/ to 6.51 mm (SD = 0.41) for /m/. For male speakers, means ranged from 6.18 mm (SD = 0.57) for /f/ to 6.59 mm (SD = 0.97) for /m/. Standard errors were generally low, indicating consistent measurements within groups. Overall, GH thickness values were similar across fricatives and nasals, with slightly higher means observed for /m/ in both genders. With ongoing data collection, further statistical analyses will be discussed and presented in the conference.

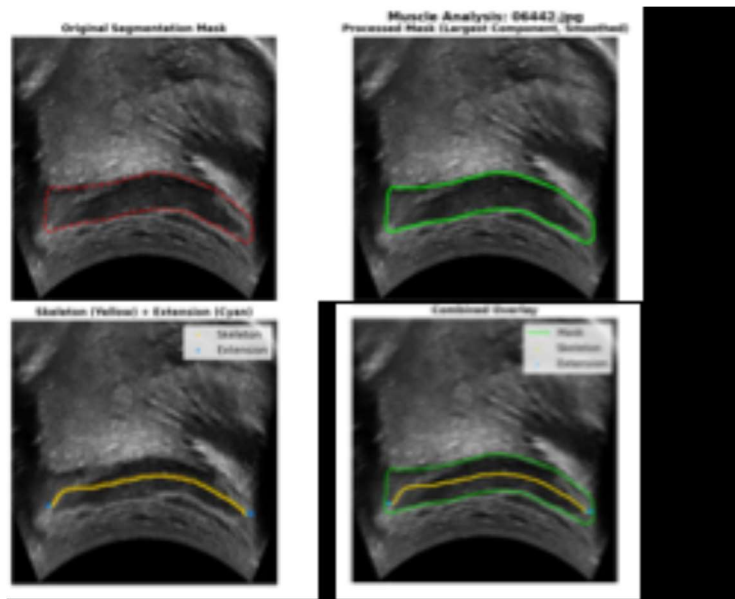


Fig. 1. An example of GH thickness measurement producing /h/ under aCa context. Coloured polygon indicates the area of GH.

Table 1. Descriptive statistics of GH thickness by consonant and gender.

Consonant	Females				Males			
	N	Mean (mm)	SD	SE	N	Mean (mm)	SD	SE
/f/	20	6.15	0.6	0.13	10	6.18	0.57	0.18
/h/	19	6.32	0.54	0.12	15	6.47	0.86	0.22
/m/	20	6.51	0.41	0.09	15	6.59	0.97	0.25
/n/	22	6.29	0.56	0.12	15	6.56	0.81	0.21
/s/	17	6.37	0.46	0.11	14	6.47	0.86	0.23
/ŋ/	22	6.41	0.51	0.11	14	6.3	0.6	0.16

References

- [1] Chiu, C., Wei, P.-C., Noguchi, M., & Yamane, N. (2020). Sibilant Fricative Merging in Taiwan Mandarin: An Investigation of Tongue Postures using Ultrasound Imaging. *Language and Speech*, 63(4), 877–897. <https://doi.org/10.1177/0023830919896386>
- [2] Liu, Y., Tong, F., De Boer, G., & Gick, B. (2023). Lateral tongue bracing as a universal postural basis for speech. *Journal of the International Phonetic Association*, 53(3), 712–727.
- [3] Tabain, M., Kochetov, A., & Beare, R. (2020). An ultrasound and formant study of manner contrasts at four coronal places of articulation. *The Journal of the Acoustical Society of America*, 148(5), 3195–3217.
- [4] De Campos, D., Jotz, G. P., Heck, L., & Xavier, L. L. (2014). Sexual Dimorphism in the Histologic Organization of the Muscle Fibers in Human Tongue. *Journal of Voice*, 28(4), 424–429. <https://doi.org/10.1016/j.jvoice.2013.12.013>
- [5] Miyawaki, K. (1975). A preliminary report on the electromyographic study of the activity of the lingual muscles. *Annu. Bull. RILP*, 9, 91–106.
- [6] Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer [Computer program]*. (Version 6.3.08) [Computer software]. <http://www.praat.org/>
- [7] Myrghyassov, A., Song, Z., Sun, Y., Wang, B. X., Wong, M. N., & Zheng, Y. (2025). UltraUNet: Real-Time Ultrasound Tongue Segmentation for Diverse Linguistic and Imaging Conditions. *arXiv preprint arXiv:2509.23225*.
- [8] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988).
- [9] T. Y. Zhang and C. Y. Suen, “A fast parallel algorithm for thinning digital patterns,” *Commun ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984, doi: 10.1145/357994.358023