A preliminary study estimating the articulatory space of Cantonese-speaking 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[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

Introduction

Ultrasound tongue imaging (UTI)

UTI has been employed in studying tongue movement & contour

- sociophonetics (Lawson et al., 2008; Liu et al., 2023),
- speech & language therapy (Cleland, 2023; Campos & Ristau, 2022)

Also in various languages,

- French (Laporte & Ménard, 2018);
- Persian (Baghban et al., 2020);
- Hungarian (Csapó & Xu, 2020);
- Mandarin (Faytak et al., 2020; Luo, 2020; Ahn et al., 2024)

Introduction - Aims

<u>Ultrasound tongue imaging (UTI)</u>

However, limited study (e.g., Havenhill et al., 2024) has utilised UTI to investigate Cantonese.

The current pilot study aims to

- Estimate the articulatory space of Cantonese-speaking healthy adults;
- If articulatory space is affected by gender and phonological context

Method

Participants

- 5 male and 5 female heathy Hong Kong Cantonese speakers
- Normal hearing with no known speech disorder, oral, facial surgery
- Mean age: 23.5 years old
- SD: 2.4 years

Stimuli - Vowels

- Isolated context, 5 repetitions
- CV/CVC context, 14 repetitions
- All tone 1

Isolated context		CV/CVC context	
Character	IPA	Character	IPA
啊	aː	卡	kʰaː
衣	iː	家	ka
烏	uː	巔	tin
		天	tʰin

Method

Procedure

- *Ultrasound*: SuperSonic Aixplorer; Frame rate: 32 FPS; depth: 9cm
- *Microphone*: UGREEN CM592; sampling rate: 44.1 kHz
- Participants read the speech stimuli while holding the ultrasound probe
- Medical sonographer ensured the images centred between the shadows of mandible and hyoid bones
- Audio and image signals recorded synchronously using an external PC via OBS studio

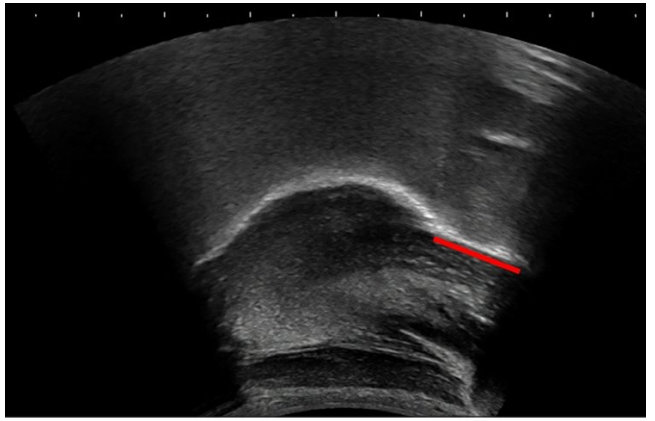
Method

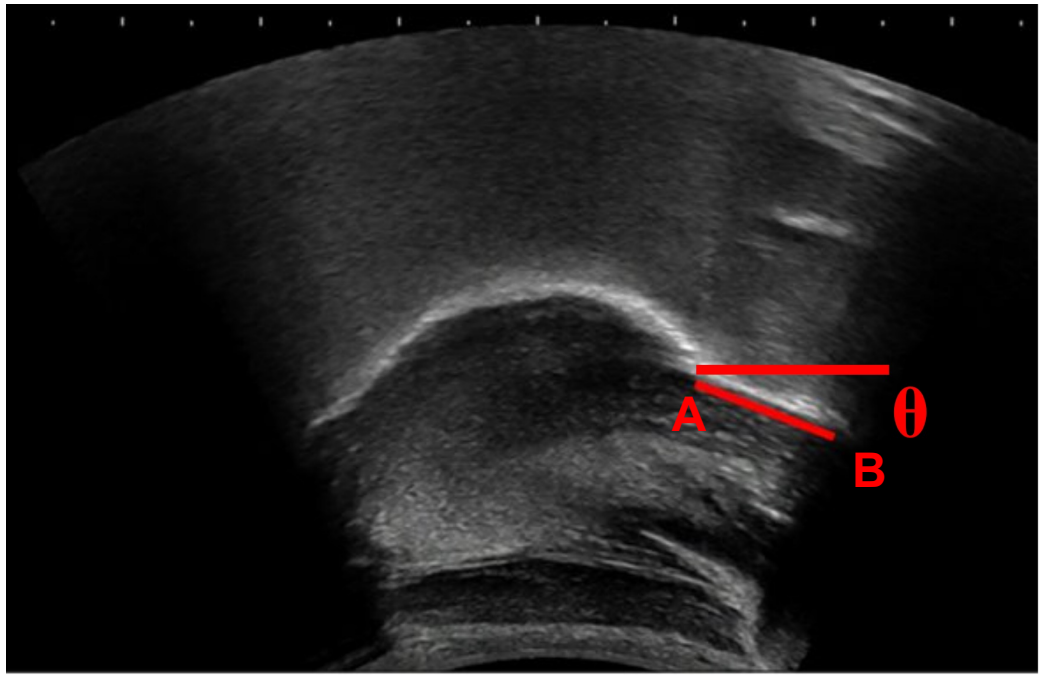
Procedure

At the beginning of the recording, participants were instructed to:

a. bite down on a biteplate and press their tongue against it

b. produce a sustained /a/



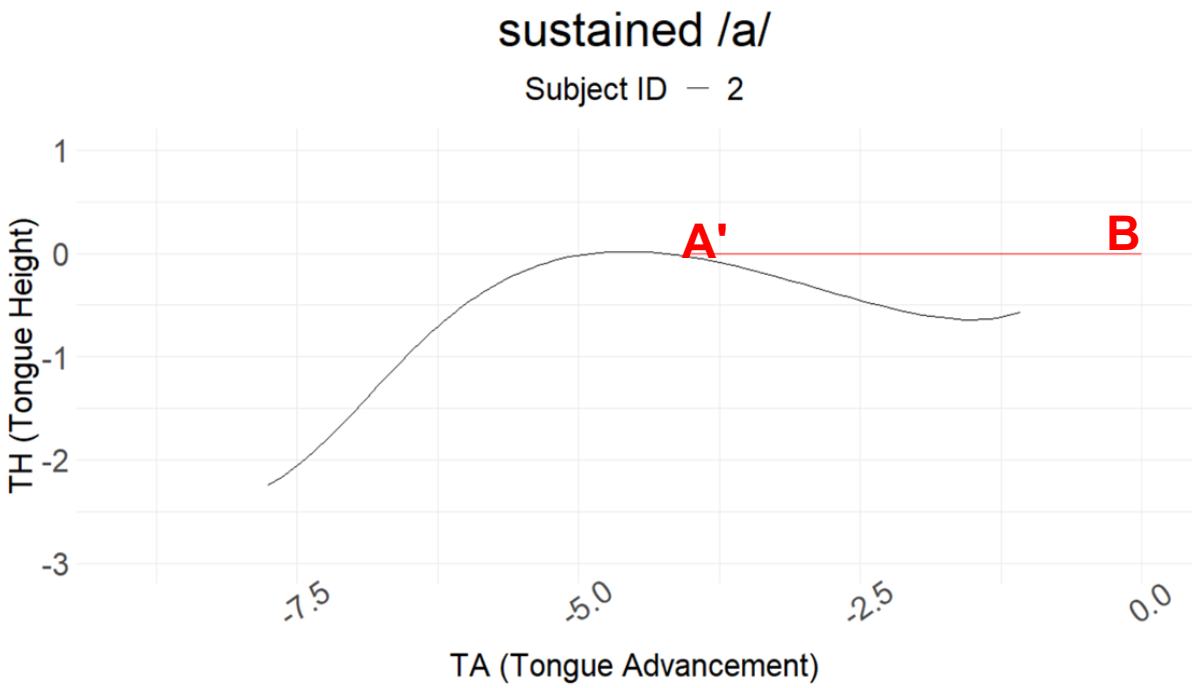


Trace of biteplate:

- solid red line indicates the trace of the biteplate
- serves as the occlusal plane (4 cm)
- right end indicates the upper incisors, the most front location of the vocal tract
- images are rotated based on the theta value to definition of horizontal and vertical orientation in the vocal tract

Method

Procedure



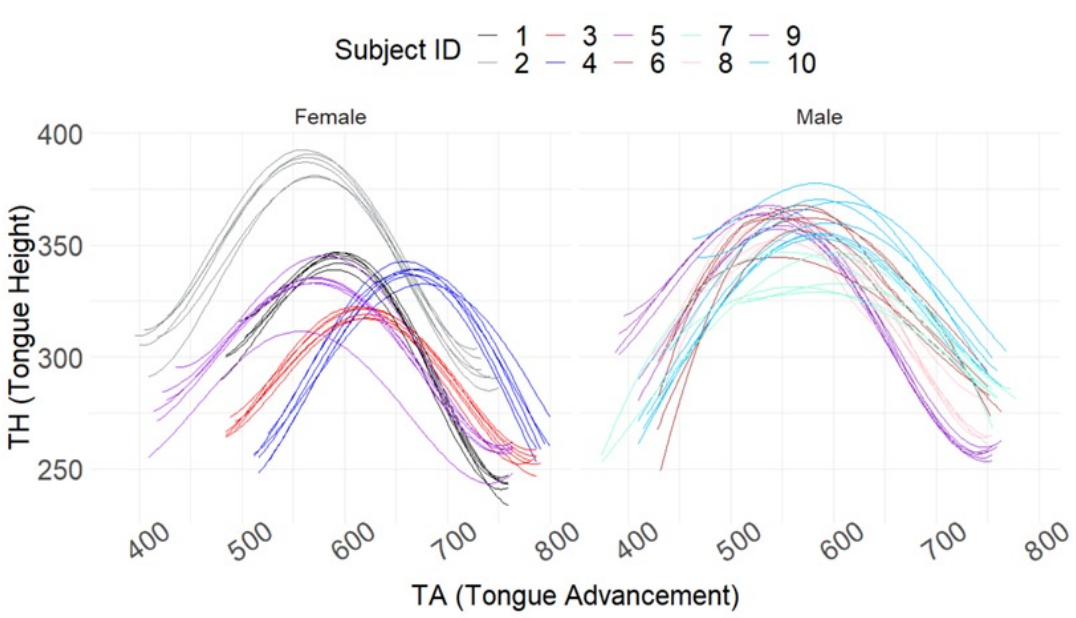
Normalisation:

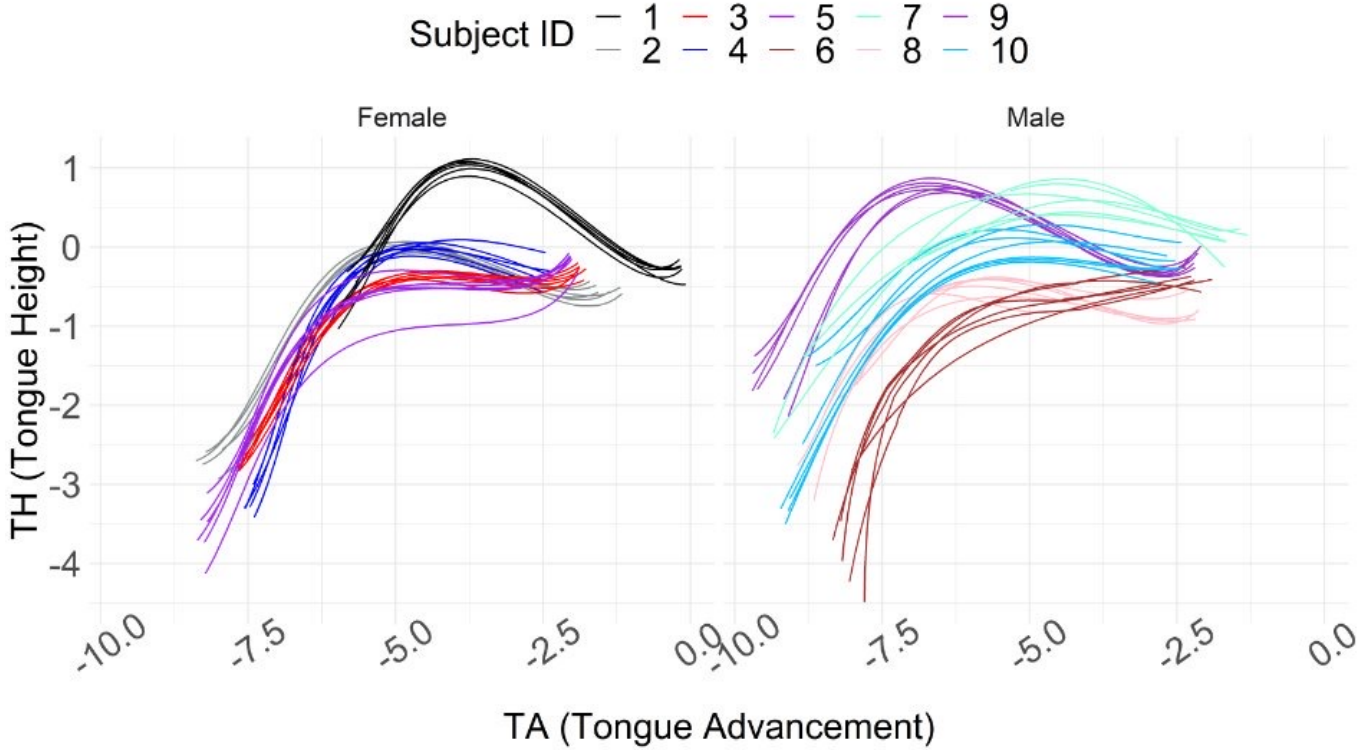
- TA normalisation factor: extend point A to A' where it intercepts tongue surface when producing sustain /a/

Method

Pre-post normalization

- Tongue shapes producing isolated /aː/ by all speakers before (left panel) and after (right panel) rotation and scaling
- Tongue contours are more aligned across speakers after rotation and scaling



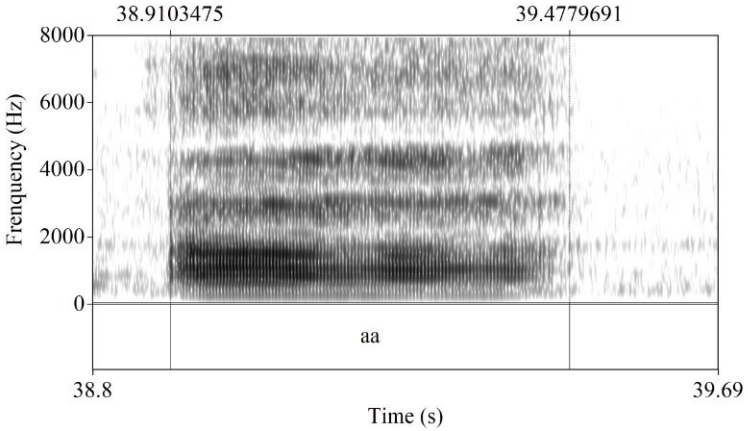


Method

Image processing

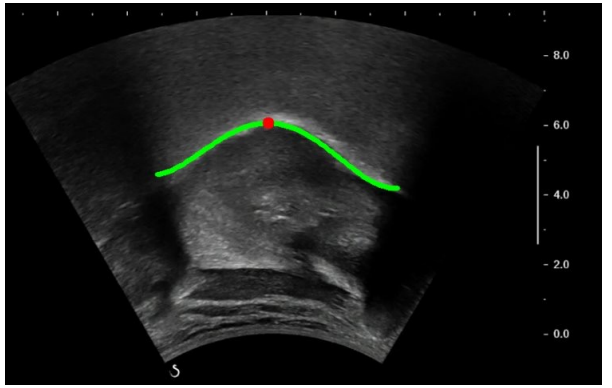
- Target vowels and consonants located in speech recordings using a TextGrid in Praat (Boersman & Weenink, 2024) → Retrieve time information
- Extract the middle frames between the onset and offset of the target vowel and consonants
- Automatic tongue contours recognition using a U-Net model (Zhu et al., 2019) with manual correction

Time info retrieving



Extract the middle frame
& automatic fitting

Tongue contour tracking



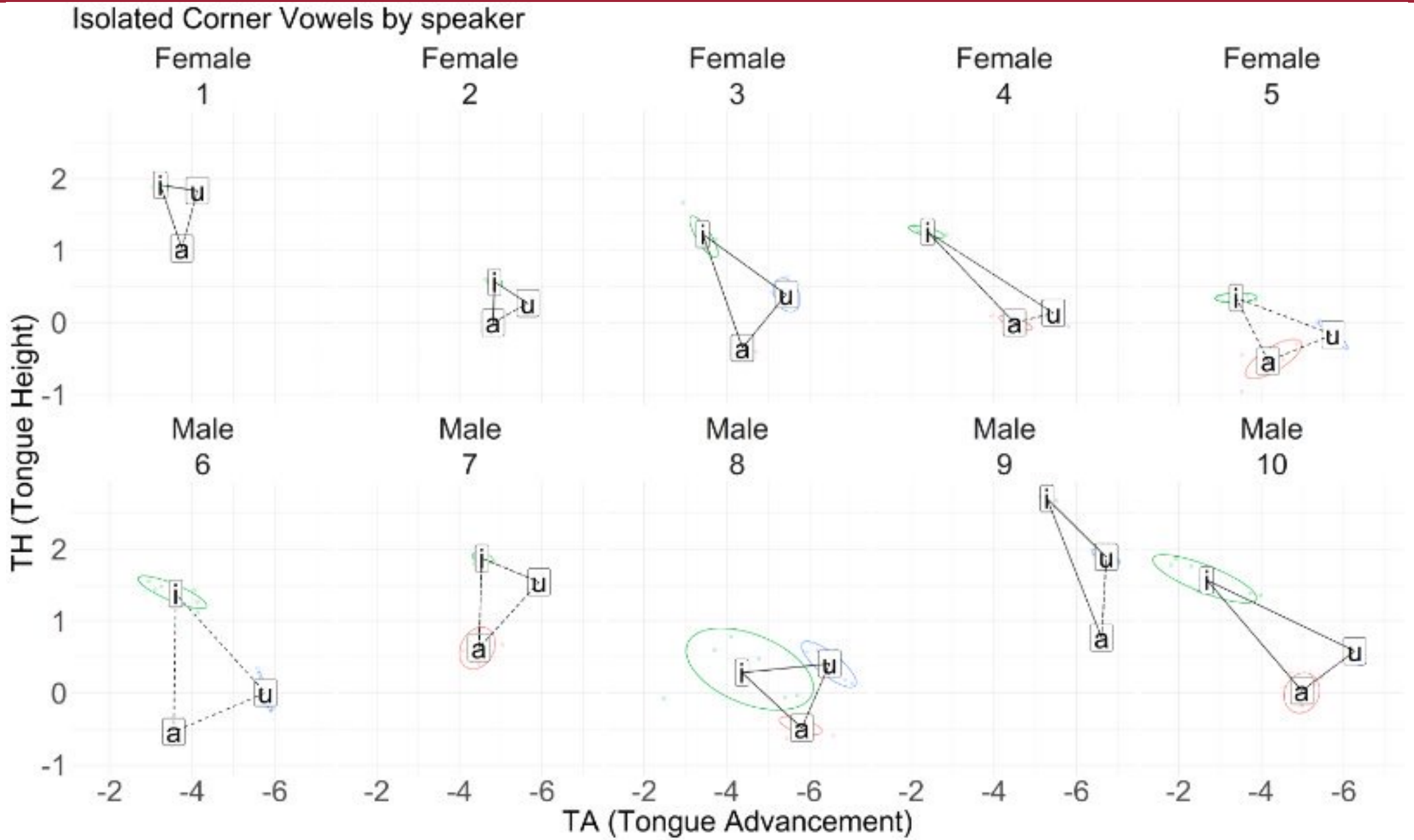
Method

Statistical analyses

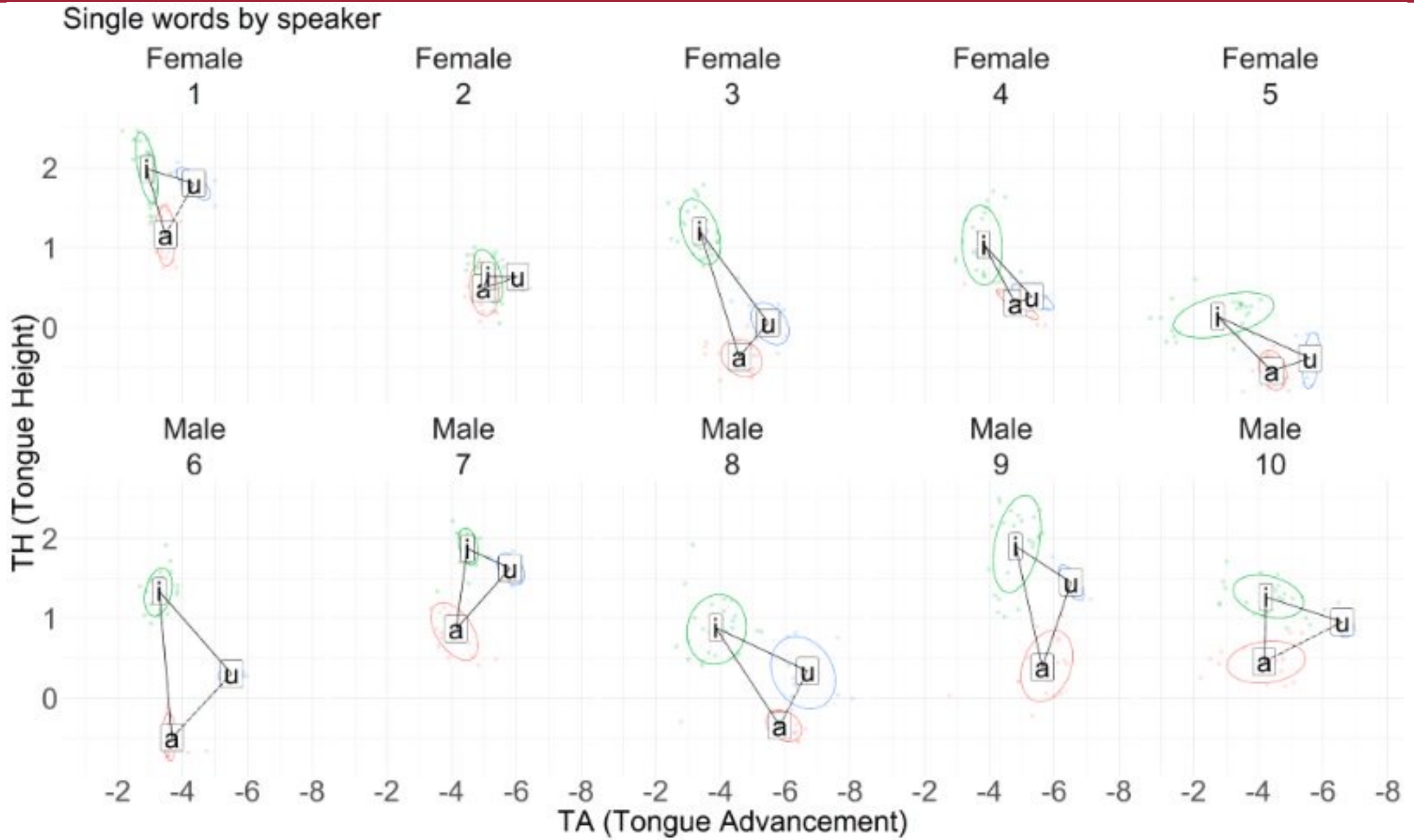
Articulatory space:

- Tongue advancement (TA) and tongue height (TH) changes were analysed using linear mixed-effect model (LMM)
  - Fixed effects:
    - ✓ gender (male, female)
    - ✓ vowel type (/aː, iː, uː/)
    - ✓ context (isolated, CV/CVC)
    - ✓ interaction between gender and vowel, and context and vowel
  - Random effect:
    - ✓ participants

Results: Isolated vowels (UTI)

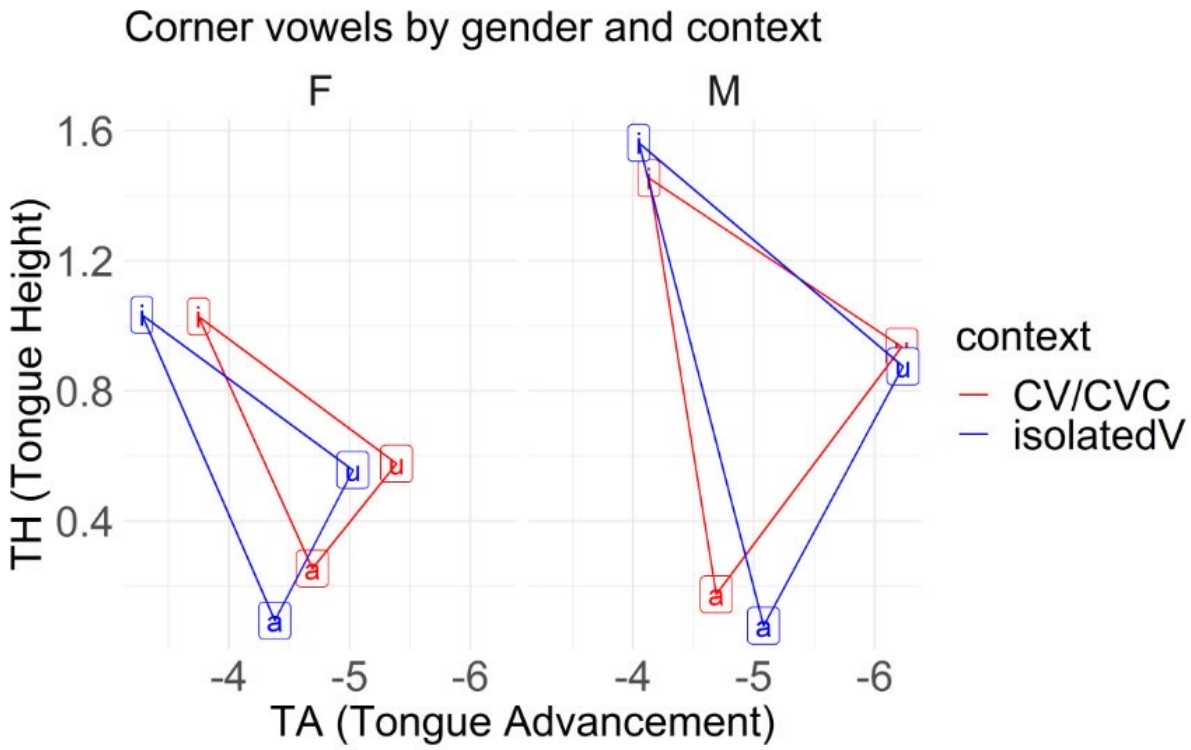


Results: Single words (CV/CVC contexts, UTI)



Results

Articulatory space (UTI)



For both contexts:

- male speakers have a larger articulatory space than female speakers
- close-front vowel /iː/ & back-close vowel /uː/ produced by female speakers is more front and lower than those produced by male speakers
- male and female speakers have similar TA and TH for open-front vowel /aː/

Results

Articulatory space (UTI)

Within male and female speakers

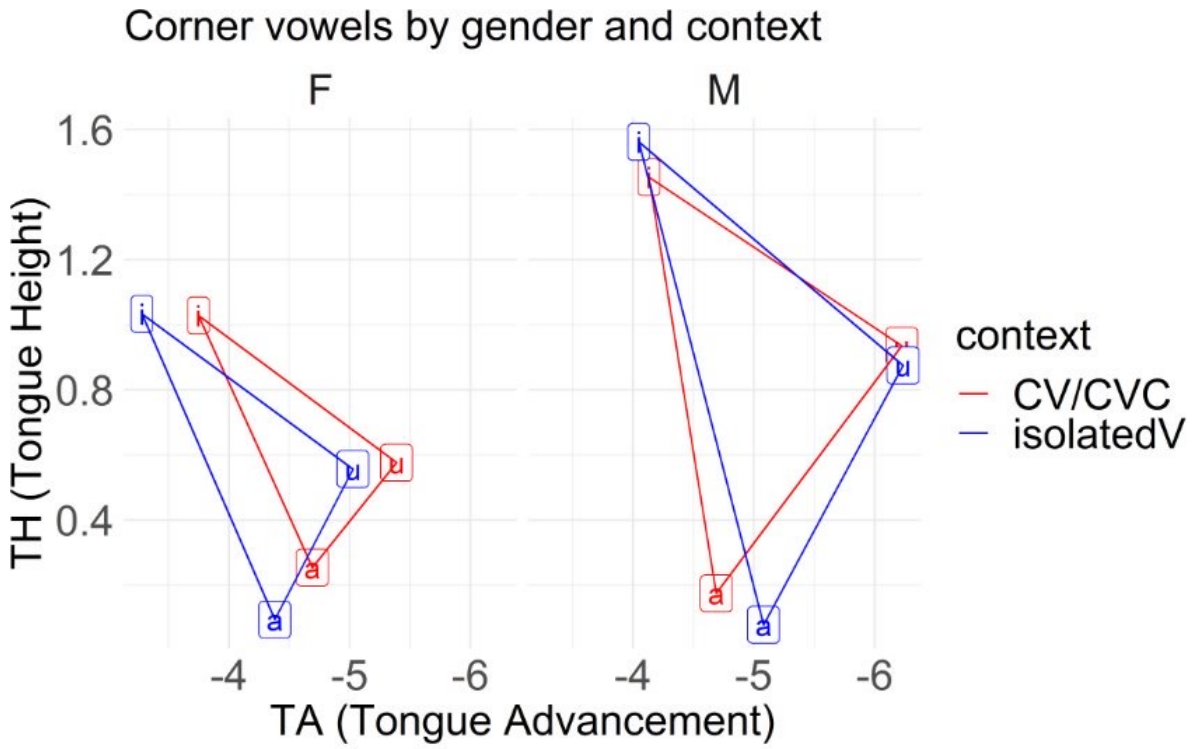
- Vowels in isolated context seem to have similar articulatory space to those under CV/CVC context

Female

- Vowels under isolated context more front than CV/CVC context
- isolated /aː/ more open than CV/CVC /aː/

Male

- Isolated /aː/ is more open and back than CV/CVC /aː/
- /iː/ and /uː/ have similar TH and TA values under different contexts



Results

Articulatory space (UTI)

Gender difference by context

	Isolated		CV/CVC	
	TA	TH	TA	TH
~ 1				
~ Gender	*			
~ Gender + Vowel type	***	***	***	***
~ Gender * Vowel type		***	**	***

$^{*}p < .05;\ ^{**}p < 0.01;\ ^{***}p < 0.001$

Post-hoc

Isolated context

- TA value of /u:/ produced by female speakers was significantly larger than that produced by male speakers
- Female speakers have more fronted back vowel than male speakers (estimate = .97, p = .037)

CV/CVC context

- No significant difference in TA and TH between male and female within each vowel category

Results

Articulatory space (UTI)

Context differences within gender

	Male		Female	
	TA	TH	TA	TH
~ 1				
~ Context				
~ Context + Vowel type	***	***	***	***
~ Context * vowel type	*			

$* p < .05$; $** p < 0.01$; $*** p < 0.001$

Post-hoc

Male

- male speakers produced /aː/ with significantly more advanced tongue position in CV/CVC context compared to in isolated context (p = .006).

Female

- female speakers produced /aː/ under isolated context significantly more open than those under CV/CVC context (p = .01).

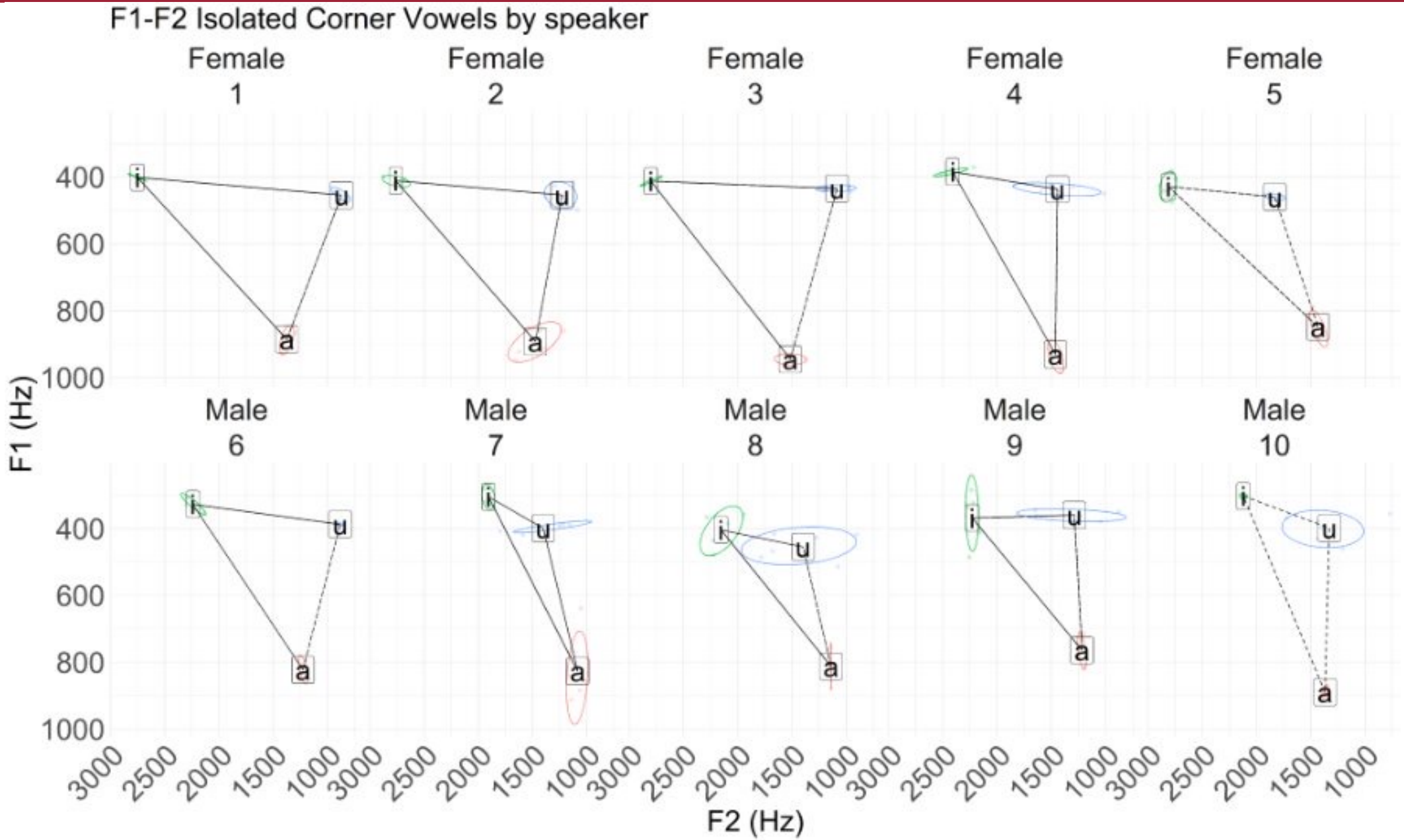
Results – Summary (UTI)

- Vowel type had a significant effect on both TA and TH (aligns with expected patterns).
- Neither gender nor context appeared to have a significant overall effect on the articulatory space.
- EXCEPT: the ONLY significant gender difference: TA for isolated /uː/, where female produced a more fronted /uː/.

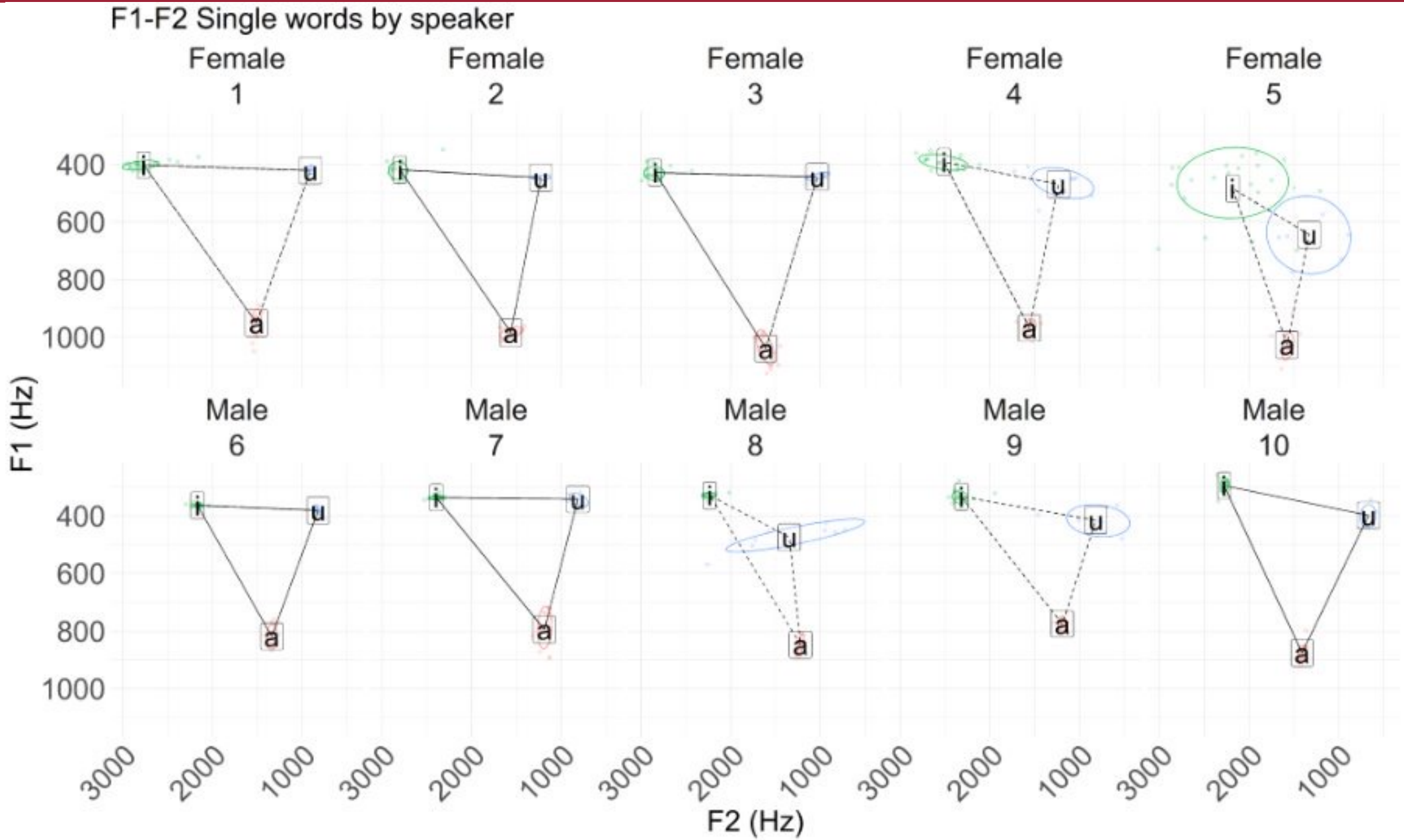
Regarding the effect of context,

- Male speakers produced /aː/ with a significantly more fronted tongue gesture in the CV/CVC context.
- Female speakers produced /aː/ with a significantly more raised tongue gesture in the CV/CVC context.

Results: Isolated vowels (acoustic)

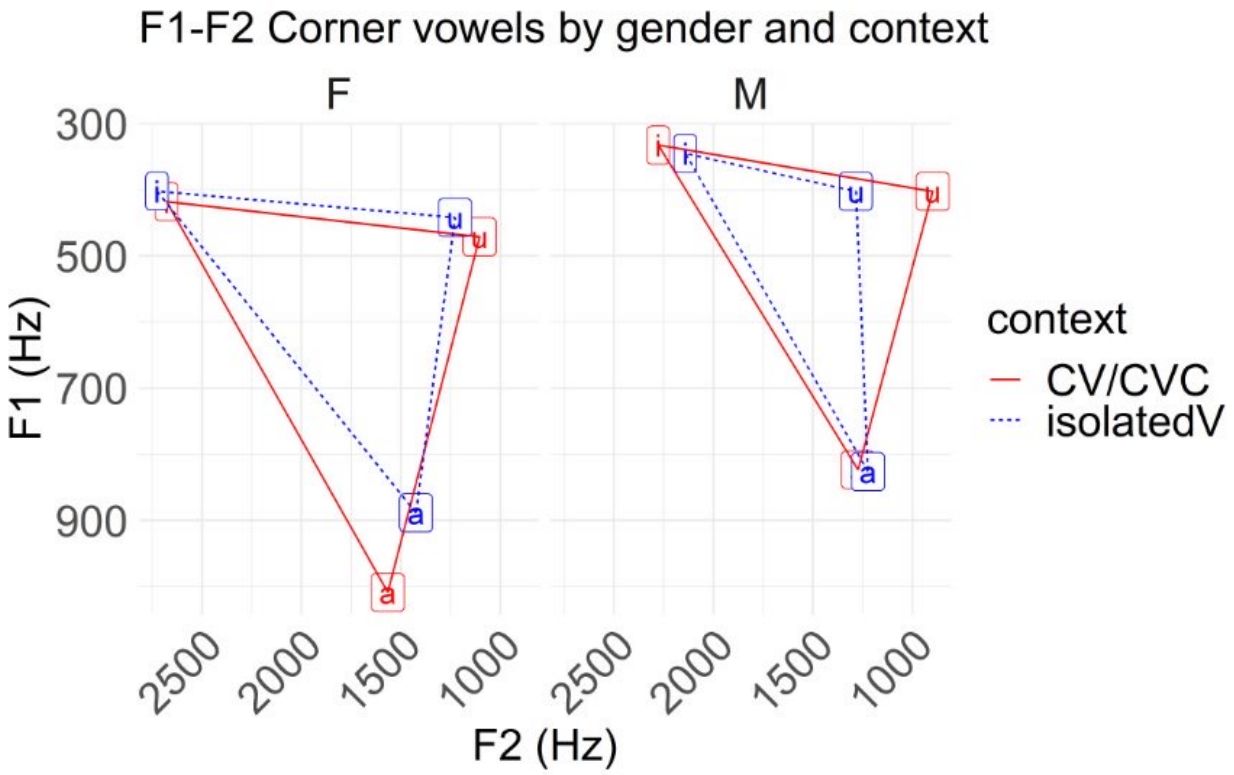


Results: Single words (CV/CVC contexts, acoustic)



Results

Acoustic space (acoustic)

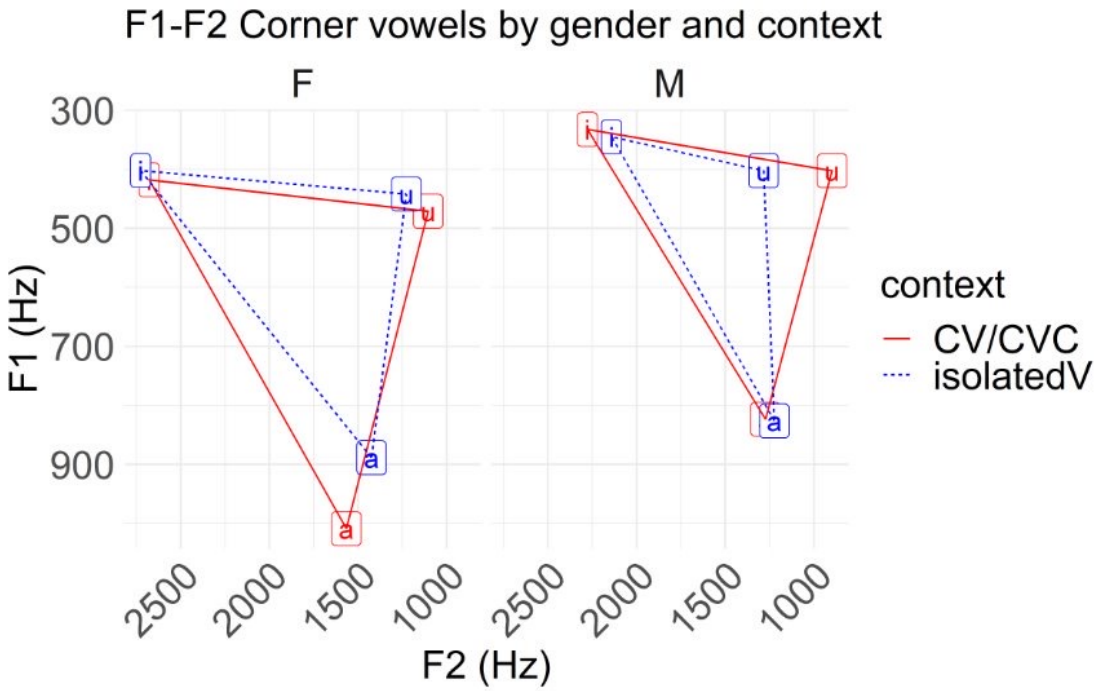


For both contexts

- Male speakers have a smaller acoustic space than female speakers.
- This is unsurprising, larger vocal tract (male) leads to lower formant frequency.
- Close-front vowel /iː/ and back-close vowel /uː/ produced by female speakers is more front (higher F2) and lower (higher F1) than those produced by male speakers.
- Male have lower F1 and F2 for /aː/ than female under both contexts.

Results

Acoustic space (acoustic)



Within male and female speakers

- Vowels in isolated context have smaller acoustic space to those under CV/CVC context.

Female

- Similar acoustic space for /iː/ and /uː/ for both contexts.
- Isolated /aː/ has lower F1 than CV/CVC /aː/. This maps back to articulatory space, as female have more mouth opening for isolated /aː/.

Male

- Similar F1 and F2 for /iː/ and /aː/ under both contexts.
- Isolated /uː/ has higher F2 than CVC /uː/, although there is no much fronting for isolated /uː/ in articulatory space.

Results

Acoustic space (Acoustic)

	Isolated		CV/CVC	
	F1	F2	F1	F2
~ 1				
~ Gender	***	**	***	***
~ Gender + Vowel type	***	***	***	***
~ Gender * vowel type		***	***	

$^{*}p < .05;\ ^{**}p < 0.01;\ ^{***}p < 0.001$

Post-hoc

Isolated context

- Female has significantly higher F1 than male speakers (/aː/ ***; /iː/ ***; /uː/ **)
- Female has significantly higher F2 than male speakers (/aː/ **; /iː/ ***)

CV/CVC context

- Female has significantly higher F1 than male speakers (/aː/ ***; /iː/ ***; /uː/ **)
- Female has significantly higher F2 than male speakers (/aː/ ***; /iː/ ***; /uː/ **)

Results

Acoustic space (Acoustic)

	Male		Female	
	F1	F2	F1	F2
~ 1				
~ Context			***	
~ Context + Vowel type	***	***	***	***
~ Context * vowel type		***	***	**

*$p < .05$; **$p < 0.01$; ***$p < 0.001$

Post-hoc

Male

- Produced CV/CVC /iː/ with significantly higher F2 (**)
- Produced CV/CVC /uː/ with significantly lower F2 (***)

Female

- Produced CV/CVC /aː/ with significantly higher F1 (***) and F2 (*)
- Produced CV/CVC /uː/ with significantly higher F1 (*) and lower F2(*)

References

Ahn, S., Kwon, H., & Faytak, M. (2024). Tongue position in Mandarin Chinese voiceless stops. JASA Express Letters, 4(2).

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

Bailey, "Open Broadcaster Software." [Online]. Available: https://obsproject.com/

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

Faytak, M., et al. 2020 Nasal coda neutralization in Shanghai Mandarin: Articulatory and perceptual evidence. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11(1): 23, pp. 1–29.

Havenhill, J., Oakley, M., Liu, M. (2024). Articulatory-acoustic dynamics in naïve listener imitation of Cantonese vowels. Presentation at LabPhon19, Seoul, South Korea. June 27 – 29, 2024.

J. Zhu, W. Styler, and I. Calloway, "A CNN-based tool for automatic tongue contour tracking in ultrasound images." arXiv, Jul. 23, 2019. doi: 10.48550/arXiv.1907.10210.

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

Luo, S. (2020). Articulatory tongue shape analysis of Mandarin alveolar–retroflex contrast. The Journal of the Acoustical Society of America, 148(4), 1961-1977.

P. Boersma and D. Weenink, "Praat:Praat: doing phonetics by computer [Computer program]." 2023. [Online]. Available: http://www.praat.org/

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Thank you

min.wong@polyu.edu.hk