Lingual articulation of the Sūzhōu Chinese labial fricative vowels

Matthew Faytak, **Jennifer Kuo**, and **Shunjie Wang** 王顺杰
faytak@ucla.edu

University of California, Los Angeles

ICPhS 19, Melbourne

Scan to download slides

Acknowledgements

- Profs. **Pat Keating** and **Megha Sundara** for comments
- **Yiqiao Liu** 刘伊乔 for contributing to contour extraction
- Prof. **Chen Zhongmin** 陈忠敏 (Fudan Univ.) and **Wang Feifan** 王非凡 (Hong Kong Univ.) for logistics

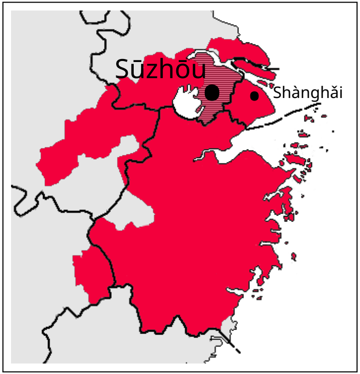
Background

Sūzhōu 苏州 Chinese

A **Wú Chinese** variety, closely related to Shanghainese

» Estimated 2–3 million speakers[1]

» Less well-described compared to some of its neighbors



[1] Yan, 1988; You, 2015; Zhengzhang, 1988.

Labial fricative vowels

Sūzhōu Chinese's vowel system includes both canonical [u] and two **labial "fricative vowels"**[2] (摩擦化元音)

- Vowels with extra supralaryngeal constriction which generates fricative noise and/or dampens formant structure
- **Bilabially compressed** or **labiodental** rather than rounded

More common than one might expect; likely underreported (perceptually similar to [u])

- Chinese dialects, especially Wú, Mandarin[3]
- Tibeto-Burman languages[4]
- Grassfields Bantu and Bantu A languages (Cameroon, Gabon)[5]

---

[2] Ling, 2009.

[3] Chen and Gussenhoven, 2015; Yuan, Ling, Shen, and Shi, 2019; Zhu, 2004.

[4] Chirkova, Wang, Chen, Amelot, and Antolík, 2015; Dell, 1981; Edmondson, Esling, and Ziwo, 2017.

[5] Connell, 2007; Faytak, 2017; Medjo Mvé, 1997; Olson and Meynadier, 2015.

Examples from Sūzhōu Chinese

Bilabially compressed [ɯ$^{\beta}$], which must occur after bilabial stops

- Frequently realized with trilling, i.e. [ʙ̩]
- Contrasts with [u]: 补 [pɯ$^{\beta}$] ‘mend’ $\neq$ 播 [pu] ‘broadcast’

Labiodental [ɯ$^{\text{v}}$], which must occur after labiodental fricatives

- Noticeable labiodental frication carries on through vowel
- In complementary distribution with [ɯ$^{\beta}$] (and [u])

[u] [ɯ$^{\beta}$] [ɯ$^{\text{v}}$]

$\rightarrow \rightarrow \rightarrow$ Increasingly small lip aperture $\rightarrow \rightarrow \rightarrow$

Study question

Since labial activity in [ɯ^{β}], [ɯ^{v}] systematically differs from [u], does this condition a systematic difference in **lingual activity**?

Study question

Since labial activity in [ɯ^{β}], [ɯ^{v}] systematically differs from [u], does this condition a systematic difference in **lingual activity**?

Vowels with simultaneous lingual and labial contributions exhibit **trading relations**, both in regular production and under perturbation[6]

» Inverse relationship between labial and lingual activity

» High back rounded [u] most often investigated, however

[6] Ménard, Perrier, Aubin, Savariaux, and Thibeault, 2008; Perkell, Matthies, Svirsky, and Jordan, 1993; Savariaux, Perrier, and Orliaguet, 1995.

Predictions

From **trading relations**:

Less lingual activity (bunching, backing) during the more heavily lip-constricted vowels [ɯᵝ], [ɯᵛ] compared to [u]

[7] Shadle, Nam, Katsika, Tiede, and Whalen, 2017.
[8] Demolin, 1992; Olson and Meynadier, 2015.

Predictions

From **trading relations**:

Less lingual activity (bunching, backing) during the more heavily lip-constricted vowels [ɯᵝ], [ɯᵛ] compared to [u]

Aerodynamic properties of consonants similar to [ɯᵝ], [ɯᵛ] make **different predictions** for [ɯᵝ]

» Active lowering of the tongue dorsum during labiodental fricatives[7]: congruent

» Bilabial trills favored near high back vowels such as [u][8]: not congruent

[7] Shadle et al., 2017.

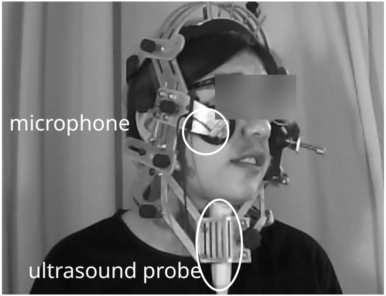
[8] Demolin, 1992; Olson and Meynadier, 2015.

Methods

Location and participants

Lingual ultrasound and **audio** recorded in a quiet room in Gūsū district, Sūzhōu (苏州市姑苏区)

- 15 native Sūzhōu Chinese speakers (13 F), all long-term residents of an urban district of Suzhou[9]
- See appendix for details of ultrasound setup



[9] Gūsū 姑苏区, Hǔqiū 虎丘区, or Wúzhōng 吴中区

Stimuli

Stimuli (after Ling[10]) presented in frame sentence using OpenSesame[11]

Vowel	As in Ling	Stimulus	Gloss
[u]	o	疤 [pu]^{44}	‘scar’
[ɯ$^{\beta}$]	u	播 [pɯ$^{\beta}$]^{44}	‘spread, sow’
[ɯ^{v}]	u	夫 [fɯ^{v}]^{44}	‘husband’
[i]	ı	边 [pi]^{44}	‘side’
[æ]	æ	包 [pæ]^{44}	‘package’

[10] Ling, 2009.

[11] Mathôt, Schreij, and Theeuwes, 2012.

Analysis

Single frame at midpoint of target vowels [u], [ɯ$^{\beta}$], [ɯ$^{\gamma}$] selected; **contours extracted** using EdgeTrak[12]

Separate **SSANOVA** for each speaker's set of frames[13]

» Polar splines re-mapped to cartesian coordinates[14]

» Sanity checking, gross overview of tongue shape and relative position

Discrete Fourier transform of contours also carried out[15]

» Straightforward comparison of **shape properties** of contours across speakers

---

[12] Li, Kambhamettu, and Stone, 2005.

[13] Davidson, 2006.

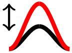
[14] Mielke, 2015.

[15] Dawson, Tiede, and Whalen, 2016.

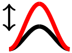
Details: discrete Fourier transform (DFT)

DFT models each tongue surface contour as a combination of sinusoidal **basis functions**

Each coefficient has **two parts** which can be taken to correspond to bunching (height) and frontness differences

- **Real parts**: frontness (**higher** = **more front**) 
- **Imaginary parts**: bunching (**higher** = **more bunched**) 

Coefficient **order**: number of peaks in function

- **First**, imaginary: degree of **simple bunching** 
- **Second**, imaginary: degree of **double bunching** 

Analysis of DFT

To assess differences in tongue contour shape, we submit each of four DFT coefficient-parts to a separate **LMER**

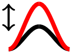
- i.e., first real, first imaginary, second real, second imaginary
- Coefficients of order > 2 do not reflect linguistically important tongue shapes

Model structure:

- coeff.-part $\sim$ vowel $+$ time* $+$ (1 $+$ vowel $+$ time | speaker)

  *time from start of experiment

Predictions in model terms [ɯ$^{\beta}$], [ɯ$^{\text{ʋ}}$] will have **smaller imaginary coefficients** than [u]

- More neutral, less bunched tongue shape
- Coef. 1, imaginary  Coef. 2, imaginary 

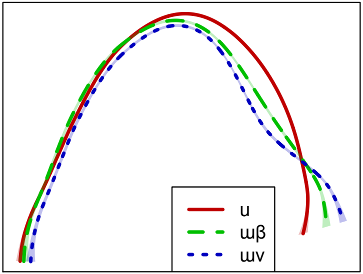
Results

SSANOVA

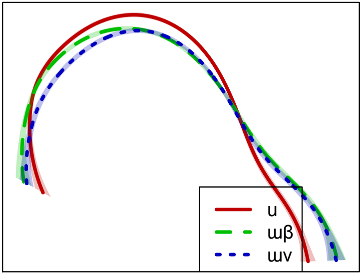
Substantial variation across speakers, but clear patterns in **tongue dorsum height** emerge

» **[u]** generally has a higher tongue dorsum and lower anterodorsum/blade than **[ɯᵝ]** and **[ɯᵛ]**

S03



S07

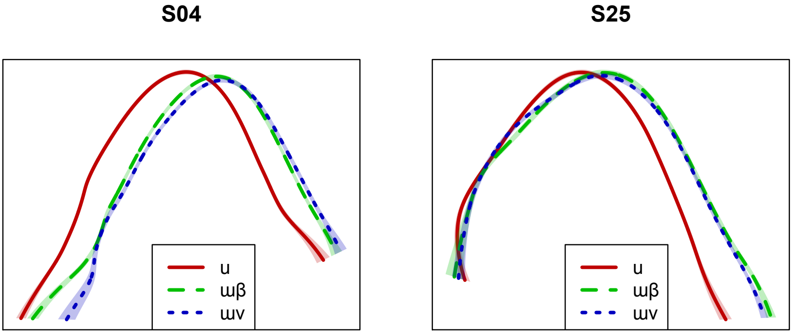


Right is front

SSANOVA

Fronting of [ɯ^{β}], [ɯ^{v}] relative to [u] is also sporadically observed

- Two speakers (S4, S25) produce **[u]** vs. **[ɯ^{β}]**, **[ɯ^{v}]** essentially as a **backness** distinction
- Several others use both strategies simultaneously (e.g. 6)

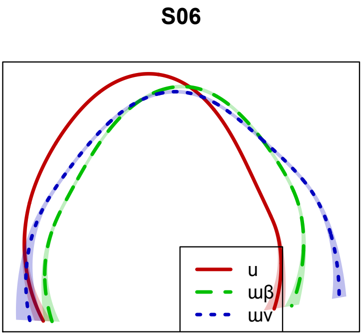


Right is front

SSANOVA

Fronting of [ɯ^{β}], [ɯ^{ᵛ}] relative to [u] is also sporadically observed

- Two speakers (S4, S25) produce **[u]** vs. **[ɯ^{β}]**, **[ɯ^{ᵛ}]** essentially as a **backness** distinction
- Several others use both strategies simultaneously (e.g. 6)



Right is front

DFT models

Effects of **vowel** for both **imaginary parts** reach significance

- First coefficient, imaginary part (p = 0.00025)
  - Substantial effect size (β = -5.9, -3.2 for [ɯ^{v}], [ɯ^{β}])
- Second coefficient, imaginary part (p = 0.03)
  - Smaller effect size (β = -2.9, -1.8 for [ɯ^{v}], [ɯ^{β}])
- Substantial interspeaker variation, in line with SSANOVA: random effects have high SD (1.66–4.38)

DFT models

Effects of **vowel** for both **imaginary parts** reach significance

- First coefficient, imaginary part (*p* = 0.00025)
  - Substantial effect size (β = -5.9, -3.2 for [ɯᵛ], [ɯᵝ])
- Second coefficient, imaginary part (*p* = 0.03)
  - Smaller effect size (β = -2.9, -1.8 for [ɯᵛ], [ɯᵝ])
- Substantial interspeaker variation, in line with SSANOVA: random effects have high SD (1.66–4.38)

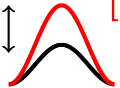
Not significant:

- Effect of **time** for any coefficient-part
- Effect of **vowel** for **real parts**

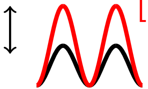
Interpreting DFT model results

The models of DFT coefficients suggest a **more neutral tongue position** for [ɯ^{v}], [ɯ^{β}] compared to [u]

- Lower **first imaginary** coefficient means [ɯ^{v}], [ɯ^{β}] are less bunched (singly) than [u]

or: 
[u] basis fcn.
[ɯ^{v}], [ɯ^{β}] basis fcns.

- Lower **second imaginary** coefficient means [ɯ^{v}, ɯ^{β}] are less doubly-bunched than [u]

or: 
[u] basis fcn.
[ɯ^{v}], [ɯ^{β}] basis fcns.

  - Most likely, extra back-raising for [u] produces slight divot in tongue blade

Discussion

Summary of findings

Sūzhōu Chinese [u] exhibits a **greater back-raising excursion** and thus **higher tongue dorsum position** than [ɯ^{β}] and [ɯ^{v}]

- **Trend toward fronting** of [ɯ^{β}] and [ɯ^{v}] relative to [u] is also visible
- Shown with complementary SSANOVA and DFT analyses

In keeping with known **trading relations** between labial articulation and lingual articulation

- [ɯ^{β}] and [ɯ^{v}] have more consistent and constricted typical labial activity
- Accordingly, the tongue dorsum constriction appears to contribute less to the overall production goal

Aerodynamic considerations

Lingual articulation of [ɯᵛ] may also reflect aerodynamic requirements for **labiodental fricative noise**

- Lowered tongue dorsum ensures airflow is impeded at lips alone[16]
- Congruent with labiodental frication as a **target for production** in [ɯᵛ]

However, tongue dorsum lowering for [ɯᵝ] is at odds with reported characteristics of **bilabial trills**

- Raised tongue dorsum reduces oral cavity volume and makes trilling more likely[17]
- Bilabial trilling in Sūzhōu Chinese [ɯᵝ] could be viewed as a **side effect** of the primary articulatory goal of lip compression

---

[16] Shadle et al., 2017.

[17] Demolin, 1992.

谢谢尔笃
Thanks

Contact: faytak@ucla.edu

Download slides:

References

Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, *45*(3), 321–337.

Chirkova, K., Wang, D., Chen, Y., Amelot, A., & Antolík, T. K. (2015). Ersu. *Journal of the International Phonetic Association*, *45*(2), 187–211.

Connell, B. (2007). Mambila fricative vowels and Bantu spirantisation. *Africana Linguistica*, *13*, 7–31.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, *120*(1), 407–415.

Dawson, K. M., Tiede, M. K., & Whalen, D. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical linguistics & phonetics*, *30*(3-5), 328–344.

Dell, F. (1981). La langue bai: Phonologie et lexique. *Cahiers de linguistique–Asie orientale*, *10*(1), 143–143.

Demolin, D. (1992). *Le mangbetu: Étude phonétique et phonologique.* (Doctoral dissertation, Université Libre de Bruxelles).

Edmondson, J. A., Esling, J. H., & Ziwo, L. (2017). Nuosu Yi. *Journal of the International Phonetic Association*, *47*(1), 87–97.

Faytak, M. (2017). Sonority in some languages of the Cameroon Grassfields. In M. J. Ball & N. Müller (Eds.), *Challenging Sonority*. Equinox.

Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, *19*(6-7), 545–554.

Ling, F. (2009). *A phonetic study of the vowel system in Suzhou Chinese*. (Doctoral dissertation, City University of Hong Kong).

References

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Medjo Mvé, P. (1997). *Essai sur la phonologie panchronique des parlers fang du Gabon et ses implications historiques*. (Doctoral dissertation, Univ. Lumière Lyon II).

Ménard, L., Perrier, P., Aubin, J., Savariaux, C., & Thibeault, M. (2008). Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children. *The Journal of the Acoustical Society of America*, *124*(2), 1192–1206.

Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America*, *137*(5), 2858–2869.

Olson, K. S., & Meynadier, Y. (2015). On Medʉmba bilabial trills and vowels. In *Proceedings of ICPhS 18*.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel [u]: A pilot “motor equivalence” study. *The Journal of the Acoustical Society of America*, *93*(5), 2948–2961.

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, *98*(5), 2428–2442.

Scobbie, J. M., Wrench, A. A., & van der Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement.. In *Proceedings of the 8th international seminar on speech production* (pp. 373–376).

References

Shadle, C. H., Nam, H., Katsika, A., Tiede, M., & Whalen, D. (2017). Aeroacoustic consequences of tongue troughs in labiodentals. *The Journal of the Acoustical Society of America*, *141*(5), 3579–3579.

Sprouse, R., & Faytak, M. (2018). Ultratils. Python package, accessed 24 Nov 2018.

Yan, Y. (1988). 吴语使用人口的再统计 [New statistics on the Wú-speaking population]. 中国语文 *[Zhongguo Yuwen]*, (4), 287–289.

You, R. (2015). Wú 吳 Dialects. In R. Sybesma (Ed.), *Encyclopedia of Chinese Language and Linguistics*. Accessed online 10 August 2017.

Yuan, D., Ling, F., Shen, R., & Shi, M. (2019). Bilabial Trill Induced by Fricative High Rounded Vowel: The Emerging of “Tb” in the Wu Dialect of Su-wan Boundary. *Journal of Chinese Linguistics*, *47*(1).

Zhengzhang, S. (1988). 关于吴语的人口 [On the Wú-speaking population]. 中国语文 *[Zhongguo Yuwen]*, (4), 289–291.

Zhu, X. (2004). 汉语元音的高顶出位 [Sound changes of high vowels in Chinese dialects]. 中国语文 *[Zhongguo Yuwen]*, (5), 440–51.

Ultrasound details

Device specs

- Telemed EchoB equipped with PV6.5/10/128 Z-3 convex probe
- Probe stabilized using Articulate Instruments headset[18]
- Frame rate $\sim$ 54 Hz

Automatic synchronization of audio and ultrasound signals

- Device's frame strobe signal combined with audio signal using a Scarlett Focusrite 2i2 USB audio interface
- Processed using **ultratils** Python package[19] and custom Python utilities

[18] Scobbie, Wrench, and van der Linden, 2008.

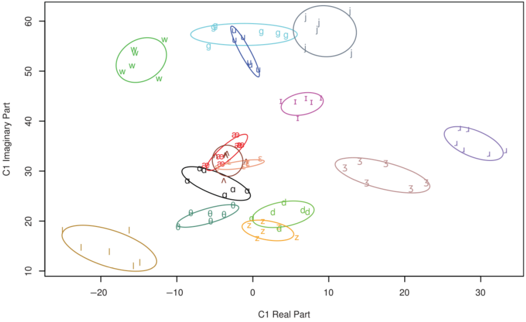
[19] Sprouse and Faytak, 2018.

More DFT details

DFT as implemented in Dawson et al., 2016 is used here

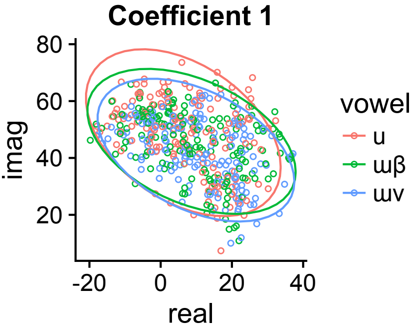
» Gives magnitude of curvature and phase **independent of size, rotation**

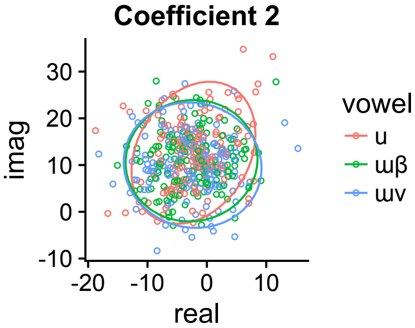
» Has been used to separate tongue contours by degree of bunchedness:



DFT coefficient-parts for all tokens

95% confidence ellipses drawn around each category





DFT models, partial summary

Significant fixed effects; [u] as baseline for “vowel” factor

a. Coefficient 1, imaginary part

	Est.	SE	Rnd. SD	$p(> \chi^2)$
vowel ɯ$^{\beta}$	-3.22	1.16	2.80	$<$**0.001**
ɯ^{v}	-5.90	1.45	4.38	
time	-0.019	0.091	0.28	0.84

b. Coefficient 2, imaginary part

	Est.	SE	Rnd. SD	$p(> \chi^2)$
vowel ɯ$^{\beta}$	-1.82	0.76	1.66	**0.03**
ɯ^{v}	-2.92	1.19	3.94	
time	0.052	0.043	0.092	0.23

DFT models, effects plots

With [u] as baseline, small to moderate effects for imaginary parts