LINGUAL ARTICULATION OF THE SŪZHŌU CHINESE LABIAL FRICATIVE VOWELS

Matthew Faytak[1], Jennifer Kuo[1], and Shunjie Wang[1]

[1]University of California, Los Angeles
faytak@ucla.edu, jenniferkuo2018@ucla.edu

ABSTRACT

In this paper, we describe lingual activity for three back vowels with different labial constrictions in Suzhou Chinese: a rounded high back vowel [u] and two more unusual vowels, vertically compressed [ɯ$^{\beta}$] and labiodental [ɯ$^{\text{v}}$]. The latter two vowels are so-called *fricative vowels*, consistently produced with slight fricative noise; [ɯ$^{\beta}$] also sporadically exhibits bilabial trilling. Smoothing-spline ANOVA models of tongue surface contours extracted from ultrasound recordings suggest that [ɯ$^{\beta}$] and [ɯ$^{\text{v}}$] have a tongue position lower and fronter than [u]. Linear mixed effects modeling of contour shape parameters also suggests that [u] has a less flat and more complex, back-raised tongue shape than [ɯ$^{\beta}$] and [ɯ$^{\text{v}}$]. We relate the lowered tongue body of [ɯ$^{\beta}$] and [ɯ$^{\text{v}}$] to the trading relations between lingual and labial activity that characterize rounded vowels, as well as to aeroacoustic properties of labial fricatives and bilabial trills.

Keywords: Ultrasound, Wu Chinese, trading relation, fricative vowels

1. INTRODUCTION

The labial component of vowel articulation is known to involve several different types of constriction, including in-rounding, out-rounding, and vertical bilabial compression [15, 17]. While usage of different labial constriction types most often systematically varies with the backness of the vowel—front rounded vowels tending to be out-rounded and back rounded vowels tending to be in-rounded—different behaviors can be found cross-linguistically [17, 26]. A distinction between out-rounding and compression also minimally contrasts vowel categories in a handful of languages, notably in a pair of Swedish high vowels ranging from central to front [13, 22] and in a pair of back vowels in Shanghainese [2].

Like Shanghainese, Sūzhōu Chinese exhibits a three-way contrast in labial activity in a set of back vowels: rounding in [u] that varies between in-rounding and out-rounding; vertical lip compression in a vowel with a similar but distinct acoustic quality that we transcribe as [ɯ$^{\beta}$]; and fricative-like labiodental constriction in a vowel we transcribe as [ɯ$^{\text{v}}$]. Typical lip positions for each vowel are shown in Figure 1. The latter two vowels can be said to continue the constriction type of the consonants that obligatorily precede them: [ɯ$^{\beta}$] occurs only after bilabial stops /p, p$^{\text{h}}$, b/, and [ɯ$^{\text{v}}$] occurs only after labiodental fricatives /f, v/ [25, 16]. On distributional grounds, [u] and [ɯ$^{\beta}$] are in fact contrastive (though [u] and [ɯ$^{\text{v}}$] are allophonic): although [ɯ$^{\beta}$] is restricted in its occurrence, [u] may occur in the same environment, and minimal pairs are easily found (see Table 1).

Figure 1: Frontal lip positions at vowel midpoint for the three vowels at issue for S44.

The two vowels [ɯ$^{\beta}$] and [ɯ$^{\text{v}}$] have been called syllabic labial fricatives or labial *fricative vowels*, in part due to the acoustic consequences of their labial constrictions [25, 27, 16]. For [ɯ$^{\text{v}}$], labiodental frication from the preceding [f] or [v] onset continues through the entire vowel. Any fricative noise present in [ɯ$^{\beta}$] is more spectrally diffuse and subtle, but it is also sporadically produced with bilabial trilling. The aerodynamic conditions for trilling do not seem to be met in all tokens of [ɯ$^{\beta}$]; when this is the case, the resulting vowel quality is not easily distinguished from [u], but the position of the lips is still visually distinct from [u]. Vowels with similar constrictions to these have been attested in a handful of languages of southwestern China [7, 3, 4, 9] and Cameroon [10, 20].

While the labial articulations of [ɯ$^{\beta}$] and [ɯ$^{\text{v}}$] have been remarked upon in various languages, to our knowledge, only Ling [16] has collected data on their *lingual* articulation; this data is somewhat limited, however, both in speaker population (n=3)

and spatial resolution (three fleshpoints tracked using EMA). We thus aim to characterize the lingual activity of [ɯ$^{\beta}$] and [ɯ$^{\upsilon}$] relative to [u] in Sūzhōu Chinese with a larger data set and a richer spatial representation using ultrasound tongue imaging. Several factors lead us to hypothesize that the tongue will exhibit a greater back-raising excursion for [u] than the other two vowels, which have more substantial lip constrictions with fewer articulatory degrees of freedom. In particular, rounded vowels such as [u] are known to exhibit trading relations in realizing their low F2 targets [21], which all three Sūzhōu Chinese vowels share [16]. A prediction of this model is that [ɯ$^{\beta}$] and [ɯ$^{\upsilon}$], having more consistent and more constricted lip settings, should have less of a need for lingual activity in service of a back vowel quality (i.e., low F2), and may be produced with a lower or centralized tongue position relative to [u].

This study can also be seen as a test of whether some lingual articulatory properties of similar consonants extend to the segments at issue. Known aerodynamic requirements for producing labiodental fricatives and bilabial trills, which may be shared by [ɯ$^{\upsilon}$] and [ɯ$^{\beta}$], respectively, make different predictions. Active lowering of the tongue dorsum has been observed during the production of labiodental fricative consonants [24], which should extend straightforwardly to [ɯ$^{\upsilon}$]. However, known lingual articulatory strategies for bilabial trills, which preferentially occur before high back vowels [8], predict tongue dorsum height similar to [u].

2. METHODS

2.1. Data collection

Participants were 15 native speakers of Sūzhōu Chinese (13 F, 2 M) who took part in the larger study described in [11]. Participants have similar residential and linguistic histories: all are long-term residents of Sūzhōu, and all report native proficiency in Sūzhōu Chinese and high or native-like proficiency in Standard Chinese.

Ultrasound video was recorded using an Echo Blaster ultrasound device equipped with a PV6.5/10/128 Z-3 microconvex probe, with a typical frame rate of 54 Hz. The ultrasound probe was held in place under the chin using an Articulate Instruments, Ltd. stabilization headset [23]. The probe angle relative to the occlusal plane varies from participant to participant given the need to accommodate differences in participant jaw and chin morphology.

Audio was recorded at a sampling rate of 44.1 kHz using a Sony ECM-77B electret condenser microphone attached to the stabilization headset's right cheekpad arm. Audio was digitized using a Focusrite Scarlett 2i2 USB audio interface, which was also configured to accept the synchronization pulse train generated by the ultrasound device for time alignment of the articulatory and acoustic signals.

Table 1: Stimuli with the simplified Chinese characters used for display. Notation for each vowel as used in [16] is provided for reference.

Vowel	As in [16]	Stimulus	Gloss
[u]	o	疤 [pu]^{44}	'scar'
[ɯ$^{\beta}$]	u	播 [pɯ$^{\beta}$]^{44}	'spread, sow'
[ɯ$^{\upsilon}$]	u	夫 [fɯ$^{\upsilon}$]^{44}	'husband'
[i]	ɪ	边 [pi]^{44}	'side'
[æ]	æ	包 [pæ]^{44}	'package'

2.2. Materials

Stimuli, shown in Table 1, were presented as simplified Chinese characters in randomized order. The stimuli were interspersed with other characters which have readings containing different consonantal onsets and vowels not relevant to the present study. Stimuli were displayed on a computer screen, and participants were instructed to produce them in a frame sentence with a reading appropriate to Sūzhōu Chinese rather than Standard Chinese. Participants produced 9–13 tokens of each stimulus.

2.3. Analysis

The series of ultrasound frames recorded during the production of each vowel token was extracted from the ultrasound video. For each vowel token, tongue surface contours consisting of 100 sampling points were extracted from the frames in this series using EdgeTrak [14]. The frame closest to the acoustic midpoint of each target vowel was selected. Each participant's set of midpoint tongue surface contours was submitted to a smoothing-spline analysis of variance (SSANOVA) [5]; splines are generated in polar coordinates and then converted to Cartesian coordinates to avoid distortion in the tongue tip and root regions [19]. The resulting models are not directly comparable across participants due to variation in vocal tract morphology and probe orientation.

A discrete Fourier transform (DFT) was also performed on the contour for each midpoint frame for [u], [ɯ$^{\beta}$], and [ɯ$^{\upsilon}$], following the method implemented in [6]. DFT converts contour data into a smaller set of numerical coefficients representing the

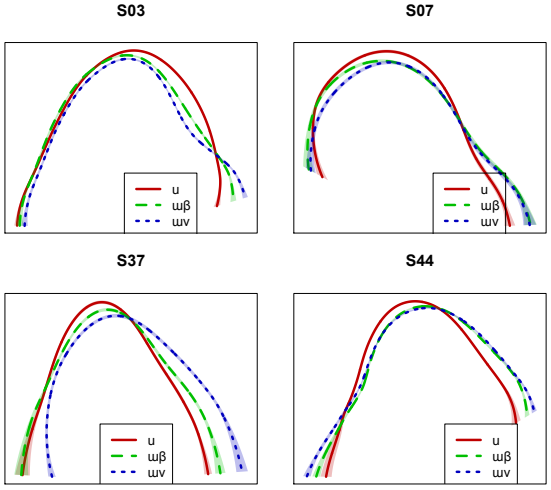
frequency and magnitude of sine and cosine functions that can be combined to model the basis data. The DFT used here is computed from the tangent angle of each point in the contour and makes no reference to a fixed coordinate system. As such, the coefficients analyzed here do not reflect contour position or rotation, but rather solely contour shape. This property is desirable for the present study, given that the direction of speaker-specific articulatory lines may differ in physical space owing to variation in vocal tract morphology.

DFT coefficients have real and imaginary portions, which correspond to the phase and magnitude of the sinusoidal basis functions, respectively [6]. Coefficients' real parts model the anterior-posterior position of the bulk of the tongue, and coefficients' imaginary parts the extent of bunching or raising. The order n of coefficients corresponds to a wavelength of $1/n$ contour arc lengths for the sinusoidal functions represented, such that higher coefficients represent greater spatial frequencies. The first coefficient thus models simple bunched tongue shapes, and the second coefficient double-bunched or saddle-like tongue shapes.

To determine whether [u], [ɯ$^{\beta}$], and [ɯ$^{\upsilon}$] systematically differ in tongue shape, the first two DFT coefficients' real and imaginary parts were each submitted to a separate linear mixed-effects model constructed using the lme4 library in R [1], with fixed effects of vowel ([u], [ɯ$^{\beta}$], and [ɯ$^{\upsilon}$], with [u] as reference level) and time elapsed since first recorded stimulus to account for any effect of speaker fatigue. The models also included a random intercept for speaker and random slopes for vowel and time (i.e., *coef* ∼ *vowel* + *time* + *(1+vowel+time|speaker)*). p-values were calculated for regression coefficients using the R car library [12], and model effects are visualized below using sjPlot in R [18].

We view the SSANOVA and DFT analyses as usefully complementary: the SSANOVA results are useful for visualizing the data in physical space and observing (potentially idiosyncratic) differences in the position of a given contour shape that cannot be assessed by the DFT method employed here, whereas the DFT allows us to make a comparison across speakers, albeit in more general terms of tongue shape complexity. DFT also functions as a "low-pass spatial filter" [6] and may as such factor out higher-frequency sinusoidal components corresponding to details of contour shape that are salient in the SSANOVA results (but linguistically irrelevant). The DFT coefficients and subject-specific SSANOVA splines are as such not expected to suggest precisely the same inter-category differences,

Figure 2: SSANOVA splines for four representative speakers; anterior is right.



but we hypothesize that both will point to a larger back-raising excursion for rounded [u] than for the other two vowels.

3. RESULTS

3.1. SSANOVA

SSANOVA splines for [u], [ɯ$^{\beta}$], and [ɯ$^{\upsilon}$] are shown for representative speakers in Figure 2. It is clear even from this subset of the fifteen participants data that there is substantial variation in the position of the tongue relative to the probe origin and in the portion of the tongue root and blade visible for tracking. Some shared patterns are visible: [u] generally has a higher tongue dorsum and lower anterodorsal region and blade compared to [ɯ$^{\beta}$] and (often to a greater extent) [ɯ$^{\upsilon}$]. There is also some degree of speaker idiosyncrasy present in the lingual articulation of the contrasts among [u], [ɯ$^{\beta}$], and [ɯ$^{\upsilon}$]. For some speakers, e.g. S44, the difference between the splines for [u] and the other vowels can readily be interpreted as involving a distinction in backness instead of height. Differences in tongue root position are most frequently not significant.

3.2. DFT analysis results

Results of linear mixed-effects regression on the first two DFT coefficients' imaginary parts are shown in Table 2 and Figure 3 by coefficient and part. Models for the coefficients' real parts are not shown, as the main effect of vowel also failed to reach significance in these models. This last result indicates that the phase of the first- and second-order sinusoidal basis functions used to model the shape data is not

Table 2: Partial summaries for models with effects significant at $\alpha = 0.05$ (estimates, standard errors, standard deviation of random effects, and p-value for effects).

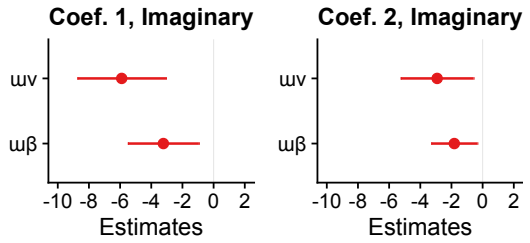
a. Coefficient 1, imaginary part

	Est.	SE	Rnd. SD	$p(>\chi^2)$
vowel ɯ$^{\beta}$	-3.22	1.16	2.80	<**0.001**
ɯ$^{\nu}$	-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$^{\nu}$	-2.92	1.19	3.94	
time	0.052	0.043	0.092	0.23

Figure 3: Estimates for fixed effects of vowel for two DFT parameters, with [u] as reference.



affected by vowel category, suggesting no substantial differences in backness of tongue dorsum raising among [u], [ɯ$^{\beta}$], and [ɯ$^{\nu}$]. The fixed effect of time did not reach significance in any of the four models.

A small but significant main effect of vowel was obtained in the model for the first coefficient's imaginary part (χ^2 = 16.60, p = 0.00025, Table 2a); taking [u] as reference, [ɯ$^{\beta}$] had a lower imaginary part of first coefficient by 3.22 $\pm$ 1.16 (SE), and [ɯ$^{\nu}$] by 5.90 $\pm$ 1.45 (SE). A significant but less robust main effect of vowel was also obtained in the model for the second coefficient's imaginary part (χ^2 = 7.012, p = 0.03, Table 2b): again taking [u] as reference, [ɯ$^{\beta}$] had a lower imaginary part of second coefficient by 1.82 $\pm$ 0.76 (SE), and [ɯ$^{\nu}$] by 2.92 $\pm$ 1.19 (SE). These effects indicate an appreciable difference in the magnitude of the sinusoidal basis functions used to model the shape data, suggesting that [u] is more bunched (in a back-raising direction, given the SSANOVA data) than [ɯ$^{\beta}$], and [ɯ$^{\beta}$] more than [ɯ$^{\nu}$]. All significant fixed effects of vowel show substantial variation among speakers: the random effects associated with each have a large standard deviation.

4. DISCUSSION

Our results reinforce the general description in [16] with a more spatially detailed representation of tongue shape and a larger speaker population. SSANOVA model results indicate that [u] exhibits a slightly higher tongue dorsum position than [ɯ$^{\beta}$] and [ɯ$^{\nu}$]. Slight advancement and raising of the tongue blade can also be observed for [ɯ$^{\beta}$] and [ɯ$^{\nu}$] relative to [u]. The DFT model results suggest that across speakers, [u] has a somewhat more complex tongue shape with a greater degree of back raising than [ɯ$^{\beta}$] and [ɯ$^{\nu}$], requiring higher-magnitude sinusoidal functions at both relatively high and low spatial resolutions to model.

Taken together, the two analyses confirm the hypothesis that of the three vowels, [u] involves the greatest deformation of the tongue in a back-raising direction. As for the other vowels, [ɯ$^{\beta}$] is produced with less back-raising and [ɯ$^{\nu}$] with even less in turn. The small but consistent effects of vowel on tongue position is in keeping with known trading relations between labial articulation and lingual articulation for each vowel: the more consistent and constricted the typical labial activity for a vowel, the less the tongue appears to contribute to achieving a given set of formant frequencies. Token-by-token variation in labial activity is not taken into account in this analysis, and including this additional factor in future analyses may shed more light on the intra- and interspeaker variation in tongue dorsum height observed here.

Of note, the lingual articulation of [ɯ$^{\nu}$] can be predicted from the aerodynamic requirements for the production of similar consonants, but [ɯ$^{\beta}$] presents complications. A lowered tongue dorsum for [ɯ$^{\nu}$] relative to [u] *is* in keeping with known aerodynamic requirements for producing labiodental fricatives, per [24]: a lowered tongue body ensures that incoming airflow is primarily impeded by the constriction at the lips. However, tongue dorsum lowering for [ɯ$^{\beta}$] is at odds with the lingual articulatory characteristics that have been suggested for bilabial trills. In languages which have bilabial trill consonants, the latter overwhelmingly precede (and lingually coarticulate with) high back vowels and in fact specifically favor [u] [8]. Sūzhōu Chinese's [ɯ$^{\beta}$] may simply have different aerodynamic requirements, given that trilling only occurs sporadically: bilabial trilling in [ɯ$^{\beta}$] could be viewed as a mere side effect of the primary articulatory goal of lip compression, such that lingual and labial articulation are not fine-tuned to consistently produce trilling.

5. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P. 2018. Linear mixed-effects models using 'eigen' and s4. R Package, v. 1.1.17, accessed 6 May 2018.

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

[3] Chirkova, K., Chen, Y. 2013. Lizu. *Journal of the International Phonetic Association* 43(1), 75–86.

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

[5] 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.

[6] 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.

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

[8] Demolin, D. 1992. *Le mangbetu: Étude phonétique et phonologique.* PhD thesis Université Libre de Bruxelles.

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

[10] Faytak, M. 2017. Sonority in some languages of the Cameroon Grassfields. In: Ball, M. J., Müller, N., (eds), *Challenging Sonority*. Equinox.

[11] Faytak, M. 2018. *Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese*. PhD thesis University of California, Berkeley.

[12] Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Laboissière, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Ripley, B., Venables, W., Walker, S., Winsemius, D., Zeileis, A. 2018. Companion to applied regression. R Package, v. 3.0.0, accessed 6 May 2018.

[13] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Blackwell Oxford.

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

[15] Lindau, M. 1978. Vowel features. *Language* 54(3), 541–563.

[16] Ling, F. 2009. *A phonetic study of the vowel system in Suzhou Chinese*. PhD thesis City University of Hong Kong.

[17] Linker, W. J. 1982. *Articulatory and acoustic correlates of labial activity: a cross-linguistic study*. PhD thesis University of California, Los Angeles.

[18] Lüdecke, D. 2016. sjplot: data visualization for statistics in social science. R package v. 2.6, accessed 3 November 2018.

[19] 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.

[20] Olson, K. S., Meynadier, Y. 2015. On Meḍumba bilabial trills and vowels. *Proceedings of ICPhS 18*.

[21] 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.

[22] Riad, T. 2013. *The phonology of Swedish*. Oxford University Press.

[23] 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. *Proceedings of the 8th International Seminar on Speech Production* 373–376.

[24] 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.

[25] Wang, P. 2011. 苏州方言研究 *[Research on the Sūzhōu dialect]*. 中华书局 [Zhonghua Book Company].

[26] Zerling, J. 1992. Frontal lip shape for French and English vowels. *Journal of Phonetics* 20, 3–14.

[27] 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.