THE L~N MERGER IN SOUTHWESTERN MANDARIN: AN ARTICULATORY STUDY

Jing Huang, Kye Shibata, Feng-fan Hsieh, Yueh-chin Chang, Mark Tiede
National Tsing Hua University, Haskins Laboratories
xiaokuidaren94@163.com, kye.shibata@gmail.com, ffhsieh@mx.nthu.edu.tw, ycchang@mx.nthu.edu.tw, tiede@haskins.yale.edu

ABSTRACT

This work is an articulatory study of the l~n merger in Southwestern Mandarin. The material includes canonical /l/ and /n/ produced in word-initial position. Results from four speakers show systematically different patterns of Tongue Tip and Body EMA sensors at target for these sounds, even though no consistent difference is found in their formant values. Corresponding coronal ultrasound imaging of /l/ production exhibits both concave and convex tongue shapes within and across speakers. Comparisons of the "lateral angles" between the fitted midsagittal spline and the parasagittal EMA sensor show no significant difference between /l/ and /n/. Taken together, our results suggest that the articulatory reflex of the l~n merger could be due to ongoing loss of posture differentiation at the tongue blade. Additional language-specific features of /l/ production will also be discussed.

Keywords: EMA, ultrasound, lateral, nasal, Southwestern Mandarin

1. INTRODUCTION

The lateral approximant involves the coordination of multiple gestures. In different varieties of English, /l/ production has been reported to involve the following characteristics: tongue tip raising, tongue middle lowering, tongue dorsum retraction, para-sagittal tongue lowering, and jaw-lowering (see [20] for a recent survey). [20]'s EMA results indicate that the lateral channel is formed by tilting the tongue to the left/right side of the oral cavity in Australian English (cf. [4]), that is, asymmetrical lowering of the sides of the tongue.

The lateral approximant is different in Mandarin, at least phonologically speaking. Firstly, /l/ is not possible in coda position in Mandarin. The well-known light /l/ vs. dark /l/ distinction in English (e.g., [5], [11]) is absent in Mandarin. Secondly and more interestingly, it has long been noted that the so-called l~n merger exists in many Southern Sinitic languages, especially in those languages spoken along the Yangtze River, be they Mandarin or non-Mandarin ([8], [3], [10]). To take a famous example, the l~n merger has been completed in Hong Kong Cantonese ([15]). The alveolar nasal onset is merged with the lateral approximant across the broad, resulting in a gap of the alveolar nasal onset. To this end, it is fair to say that the phonology of the lateral is different in Sinitic languages if compared with the more well-known cases in English. The first research question of this study, then, is to explore if there is any language-specific feature of /l/ production, given that the lateral approximant has distinct phonological patterning. The second research question is to explore if there is any phonetic underpinning for the l~n merger. Southwestern Mandarin (SWM) was chosen as the target language to address these research questions. This is primarily because the l~n merger remains a sound change in progress ([17], [21], [22]) in SWM. It is also beneficial if (gain/loss of) lateralization can be investigated through the lens of a phonetically gradient sound change such as the l~n merger.

2. EXPERIMENT

Seven native SWM speakers in their twenties participated in this study. All of them were born and raised in western Hubei, speaking SWM as their dominant language, and have no reported history of speech and/or hearing disorders. The data from four of them (M01~M04) are reported.

The recording materials are comprised of 30 disyllabic words. Here we analyzed and reported six of them, namely, /la.la/, /li.li/, /lu.lu/, /na.na/, /ni.ni/, and /nu.nu/. These are (possible) nicknames. Note also that the stimuli are based on prescriptive transcription. The participants were asked to read a randomized list of the target words from a computer screen at a normal speech rate in a sound-proof recording booth at National Tsing Hua University. The target words were embedded in the carrier phrase, "$p^h e$ A, *pu* $p^h e$ B", meaning "(I) pat A; do not pat B." Seven repetitions were collected for each token, and the target words in A were analyzed (as they are under focus). We used Carstens's AG501 and a Micro system from Articulate Instruments Ltd. to collect articulatory data. Acoustic data were recorded simultaneously with a sampling rate of 24 kHz. The EMA data were down-sampled to 250 Hz. The ultrasound data were collected using a transducer with a 92° field of view (FOV), set at a depth of 120mm. The frame rate was 65 f.p.s.

Fig. 1 shows the layout of the EMA sensors. In addition to the Tongue Tip (TT), Tongue Body (TB), and Tongue Dorsum (TD) sensors along the mid-sagittal line, a parasagittal (PL) sensor is attached to the tongue blade on the speaker's lefthand side.

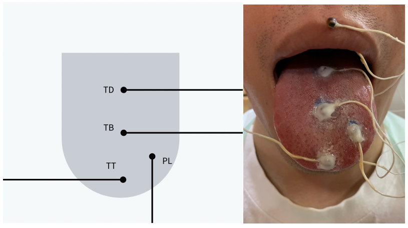

Figure 1: The layout of the EMA sensors

The EMA data were analyzed with the help of Mview [13] and ultrasound images were annotated and analyzed by the AAA software and GetContours [14]. Praat [2] was used for acoustical analysis.

3. RESULTS

The data were first transcribed by a trained phonetician. The "error rate" for /l/ is 13%, while the "error rate" for /n/ is 39%, whereby "error" here means a deviance from the prescriptive pronunciation (e.g., /li/ was transcribed as [ni]). That is to say, the target for the merger is /l/, at least in this study.

3.1. The lateral approximant in SWM

3.1.1. Visualization of the EMA data

In order to characterize the tongue posture in the coronal view, a "pseudo" (virtual) sensor, vTB, on the mid-sagittal surface of the tongue was calculated (cf. [20]). Its position is estimated by taking the point along the fitted spline connecting the mid-sagittal sensors with the shortest Euclidean distance to the PL sensor in the x- and z-dimensions. Fig. 2 illustrates the temporal changes of /l/ in both mid-sagittal and coronal planes. It seems that the TT raising is the most robust lingual movement in /l/ production. We also found variations in the coronal tongue shape: the PL could be raised, lowered, or nearly flat when compared to the vTB position, as shown in Fig. 2.

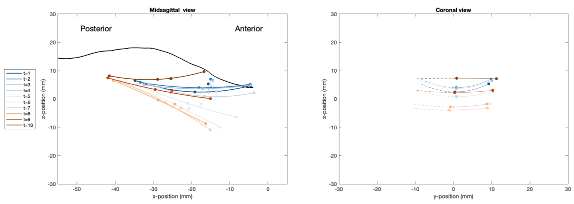

Figure 2: Temporal changes of lingual configurations of [la] of M01 (The speaker is facing right; blue line = starting point and red line = endpoint; midsagittal view on the left and the coronal view on the right side).

3.1.2. Components in /l/ production

It has been claimed in [11] that /l/-production involves a consonant-like coronal component and an intrinsic vowel-like dorsal component (i.e., TT raising and TD retraction) in English. Our results suggest that SWM patterns alike in this regard. We can see in Fig. 3 that the lateral approximant has two gestures: tongue tip raising (TTz) and tongue dorsum fronting (TDx). No significant vertical movement of the TD sensor is observed (TDz), however. The illustrations for /lu/ and /li/were omitted due to space limitations, but the observation still holds.

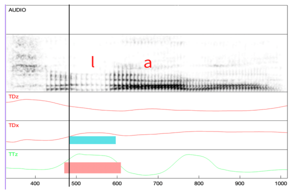

Figure 3: Temporal layout of the horizontal movement of Tongue Dorsum (TDx) sensor and vertical movement of Tongue Tip (TTz) sensor for the word-initial /l/.

3.1.3. Ultrasound results of /l/ in the coronal plane

Following [5] and subsequent work, we present the coronal tongue shapes of /l/ production. We can see from Fig. 4 that there is inter-speaker variation. Both grooved and domed coronal tongue shapes are found (M01 and M02, respectively), all else being equal. The image in Fig. 4 seems to show tilting to the speaker-right side (although it's tilted left in the image due to our POV) of the oral cavity, probably replicating the central claim made in [20], according to which the lateral channel is formed by tilting the tongue to the left/right side. The ultrasound images for /lu/ and /li/were omitted due to space limitations, but the observation still holds. It is equally worth noting that our speakers consistently produced a specific coronal tongue shape (concave or convex) across the tokens in the experiment.

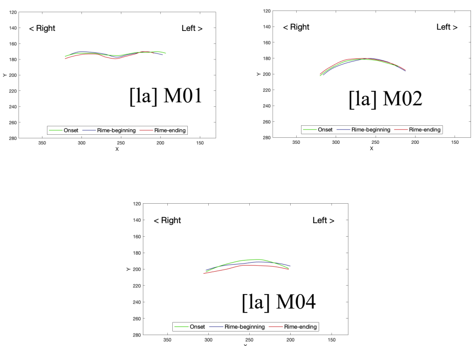

Figure 4: Ultrasound lingual configuration in coronal view of the onset in [la] for M01, M02, and M04.

3.2. The l~n merger

We turn now to the l~n merger. Note again that our results of impressionistic transcription suggest that most tokens were rendered like a lateral approximant. Nevertheless, this study classified the /l/ and /n/ groups according to prescriptive pronunciations.

3.2.1. The lateralization angle

We quantify the degree of lateralization by taking the "lateralization angle" of the tongue. This lateralization angle, θ_L, is defined as

$$(1)\ \theta_L = \tan^{-1}\left(\frac{\alpha\sqrt{d_x^2 + d_z^2}}{d_y}\right)$$

where dx, dy, and dz are the differences in x-, y-, and z-dimensional positions of the vTB and PL, respectively, and α is a coefficient which equals 1 when the PL is below the mid-sagittal spline, and -1 when above, leading to a positive angle for a lowered PL, and a negative angle for a raised PL. This is visualized in Fig. 5. Note again that θ_L is based the "vTB-PL" plane, represented as the shaded plane in the 3D view.

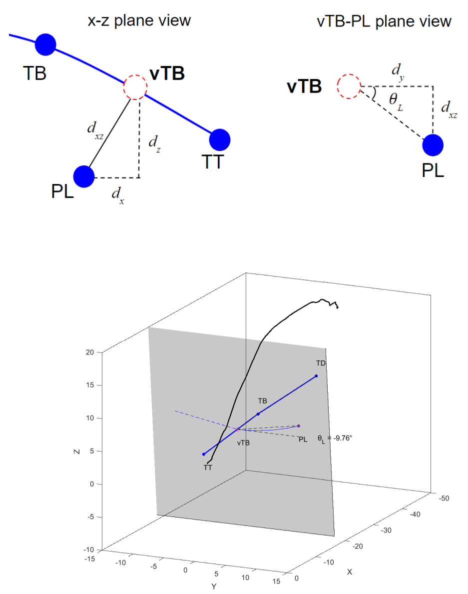

Figure 5: The visualization of the lateralization angle.

Our results show no apparent differences in θ_L between the /l/ and /n/ groups, with both groups having significant inter- and intra-speaker variation (Fig. 6). Both groups had positive lateralization, negative lateralization (or, "anti-lateralization"), and nearly flat configurations.

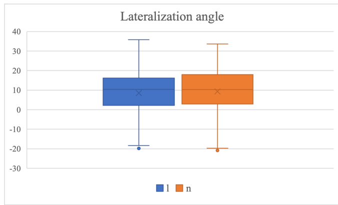

Figure 6: Lateralization angles of /l/ vs. /n/

3.2.2. GAMM results of the EMA data

The Generalized Additive Mixed Models (GAMMs) analyses for the trajectories of the EMA sensors are primarily based on the procedures and suggestions provided in [9] and [18]. The head-corrected data were z-transformed and subsequently fed into GAMM models. The models included a by-word smooth function through time to investigate articulatory changes over time and a random smooth to account for variations between all the speakers.

In Fig.7, the Tongue Tip (TT) sensor trajectories in the /l/ vs. /n/ groups were compared. Note that an x-axis in red indicates a statistically significant difference. As shown, the TTx, but not the TTz, trajectories are significantly different between the /l/ and /n/ groups.

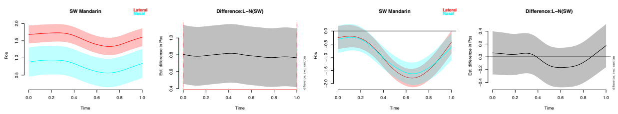

Figure 7: GAMM results of the EMA sensor trajectories of TT (TTx on the left side; TTz on the right side) of the /l/ vs. /n/ groups

The overall results for the three pairs are summarized in Table 1. The trajectories of TT, TB, TD and PL of these pairs are significantly different along either the horizontal (x) or the vertical (z) dimensions throughout the entire syllable.

Table 1: Prescriptive /l/ vs. /n/: GAMM results.

L vs. N	TT	TB	TD	PL
/la/ vs. /na/	TTx	TBx	TDx	PLx/z
/li/ vs. /ni/	TTz	TBx	n.s.	PLx/z
/lu/ vs. /nu/	TTx/z	TBx/z	TDx/z	PLz

3.2.3. GAMM results of the ultrasound data

Regarding the ultrasound data, the polar coordinates of the tongue contours (comprised of 42 points) were exported from the AAA software. The tongue contours (in the midpoints of the /l/'s or /n/'s) were fed into GAMMs using the *bam* function in *R* from the package *mgcv* [19] for the statistical significance analysis, with the help of [7]'s script adapted to our data by us.

Fig. 8 illustrates the fitted splines of all the tokens of /la/ vs. /na/, /li/ vs. /ni/, and /lu/ vs. /nu/ in the midpoints of the onsets across all the four speakers. The regions of significant difference, indicated by the shaded areas, were based on the outputs of the *plot_diff* function of the *itsadug* package. The GAMM results indicate that the tongue postures are not different in most parts of the tongue; the pair /li/ vs. /ni/ differ only in approximately 50% of the entire fitted splines. To this end, we may say that /l/ and /n/ are not distinguishable in terms of the tongue configuration in the midpoints of the consonantal part.

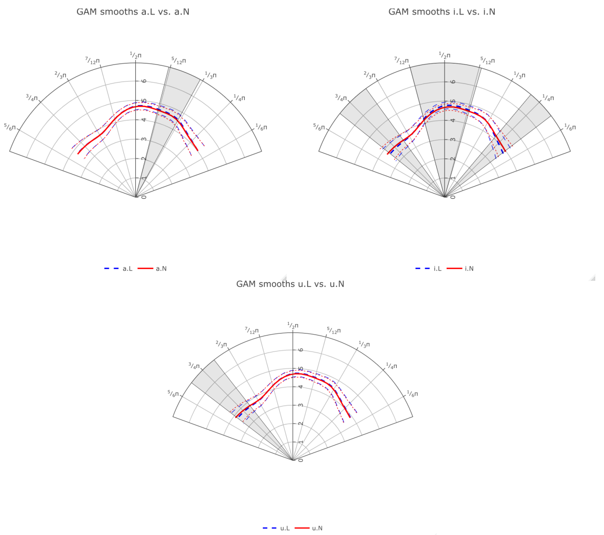

Figure 8: Fitted splines for /la/ vs. /na/, /li/ vs. /ni/, and /lu/ vs. /nu/. Gray areas indicate statistical significance. The speakers (n=4) are facing right.

3.3. The SS ANOVA results of acoustic data

Regarding the acoustic data, the formant values of the consonants were normalized using the Lobanov method ([16]). The SS ANOVA ([6]) results of the comparisons of the F1, F2, and F3 for the three pairs are provided in Table 2, where '√' means statistical significance and "0.5~1", for example, means the two trajectories are significantly different between 0.5 and 1 (consonant duration was normalized to 1).

Table 2: Pair-wise comparisons between /l/ and /n/ groups: SS ANOVA results.

L vs. N	F1	F2	F3
/la/ vs. /na/	√ 0.5~1	n.s.	n.s.
/li/ vs. /ni/	n.s.	n.s.	n.s.
/lu/ vs. /nu/	n.s.	√ 0.6~0.8	n.s.

4. DISCUSSION

Our first research question is concerned with the articulatory characteristics of /l/ in SWM. Firstly, it turned out that /l/ has two components, but they come in different flavors: TTz raising and TDx fronting (but not TD retraction, e.g., [11]). We also found that TT is formed relatively earlier than TD in /l/ production. Secondly, both EMA and ultrasound data confirm that both concave and convex tongue shapes are found in the coronal plane. This is an interesting observation because a domed coronal tongue shape has not been documented at least in /l/ production of English ([1], [12]), to our knowledge. The present results suggest that laterality might be subject to distinct phonetic implementations, especially when /l/ cannot appear in coda position in Mandarin.

Our second research question addresses a relatively understudied issue: the l~n merger. The acoustic and transcription results indicate that there is an ongoing merger, which is also confirmed by the ultrasound results. Inspired by [20], we proposed the lateralization angle of the tongue to quantify the degree of lateralization (1). The results show that there is no apparent difference between the /l/'s and /n/'s, suggesting that these two sounds are not distinguishable in terms of posture differentiation at the tongue blade. We take this finding as a potential articulatory indicator of the l~n merger. Finally, it is equally remarkable that the comparisons of the TT/TB sensor trajectories along the midsagittal planes are significantly different between the two groups. We speculate that the results in Table 1 reflect the kinematic differences between a stop ([n]) and a sonorant ([l]) in the anterior parts of the tongue. In other words, the l~n merger can be treated as a near merger, with the difference of the lateralization angle being lost first.

5. CONCLUSION

The present results suggest that /l/ production may involve language-specific characteristics (i.e., TD fronting in Fig. 3). More importantly, the l~n merger was confirmed not only by the transcription and acoustic results but also by the EMA and ultrasound data. In particular, the EMA data showed loss of posture differentiation at the tongue blade between /l/'s and /n/'s (see the lateralization angles in Figs. 5 and 6). In conclusion, this study contributes to understanding laterality from a novel perspective.

6. ACKNOWLEDGEMENTS

This study was supported by an NSTC research grant (MOST 109-2410-H007-061-). This study was approved by the National Tsing Hua University Research Ethics Committee (10812HS125). The first two authors are listed in alphabetical order as they contribute equally to this paper.

7. REFERENCES

[1] Berkson, K. H., de Jong, K., Lulich, S. M. 2017. Three dimensional ultrasound imaging of pre-and post-vocalic liquid consonants in American English: Preliminary observations. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5080-5084.

[2] Boersma, P. 2009. Praat: doing phonetics by computer (Version 5.1.05). http://www. praat. org/.

[3] Cheng, R., Jongman, A., & Sereno, J. A. 2022. Production and Perception Evidence of a Merger: [l] and [n] in Fuzhou Min. *Language and Speech*, 00238309221114433.

[4] Gick, B., Allen, B., Roewer-Després, F., Stavness, I. 2017. Speaking tongues are actively braced. *Journal of Speech, Language, and Hearing Research* 60(3), 494-506.

[5] Giles, S. B., Moll, K. L. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31, 206–227.

[6] Gu, C. 2014. Smoothing spline ANOVA models: R package gss. Journal of Statistical Software 58(5), 1-25.

[7] Heyne, M., Derrick, D., Al-Tamimi, J. 2019. Native Language Influence on Brass Instrument Performance: An Application of Generalized Additive Mixed Models (GAMMs) to Midsagittal Ultrasound Images of the Tongue. *Frontiers in psychology* 10. 2597–2597.

[8] Soejima, K., Imagana, H., Kiritani, S. 1990. Alternation between Stop Nasal and (Nasalized) Flap or Lateral. *Ann. Bull. RILP.* 24, 131–144.

[9] Sóskuthy, M. 2021. Evaluating Generalised Additive Mixed Modelling Strategies for Dynamic Speech Analysis. *Journal of phonetics* 84, 101017.

[10] Shi, X. J. 2015. Chengdu hua xiangyin de bihuadu—jian lun /n,l/ bufen de shizhi ji leixing [The nasality degree of sonorant in Chengdu dialect]. *Journal of Chinese Phonetics*, 1, 92-100.

[11] Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291–311.

[12] Stone, M., Lundberg, A. 1996. Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*, *99*(6), 3728-3737.

[13] Tiede, M. 2005. MVIEW: software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories*.

[14] Tiede, M., Whalen, D. H. 2015. Getcontours: An interactive tongue surface extraction tool. *Proceedings of Ultrafest VII.*

[15] To, C. K., McLeod, S., Cheung, P. S. 2015. Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical linguistics & phonetics* 29(5), 333-353.

[16] Thomas, E. R., Kendall. T. 2007. NORM: The Vowel Normalisation and Plotting Suite. Online Resource. http://lingtools.uoregon.edu/norm/about_norm1.php.

[17] Wang, L. 2004. *Hanyu Shigao [The history of Chinese]*. Beijng: Zhonghua Shuju.

[18] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86-116.

[19] Wood, S. 2019. Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Available: https://cran.r-project.org/web/packages/mgcv.pdf.

[20] Ying, J., Shaw, J. A., Carignan, C., Proctor, M., Derrick, D., Best, C. T. 2021. Evidence for active control of tongue lateralization in Australian English/l. *Journal of Phonetics* 86, 101039.

[21] Zhang, W. 2007. Alternation of [n] and [l] in Sichuan dialect, Standard Mandarin and English: A single-case study. *Leeds Working in Linguistics and Phonetics* 12, 156-173.

[22] Zhang, W., Levis, J. M. 2021. The southwestern mandarin /n/-/l/ merger: effects on production in Standard Mandarin and English. *Frontiers in Communication*, 171.