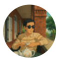
Supralaryngeal Kinematics of Implosives in Central Vietnamese: An EMA Study

Conference Paper · May 2025

5 authors, including:

Paul McGuire
National Tsing Hua University

Kye Shibata
National Tsing Hua University



Feng-fan Hsieh
National Tsing Hua University

Yueh-Chin Chang
National Tsing Hua University

Accepted to Interspeech 2025

Supralaryngeal Kinematics of Implosives in Central Vietnamese: An EMA Study

Paul McGuire, Kye Shibata, Thanh Viet Cao, Feng-fan Hsieh, Yueh-chin Chang

National Tsing Hua University, Taiwan

graemepaulmcguire@gmail.com, kye.shibata@gmail.com, kaoviet@gapp.nthu.edu.tw, ffhsieh@mx.nthu.edu.tw, ycchang@mx.nthu.edu.tw

Abstract

This study, using electromagnetic articulography (EMA), examines the supralaryngeal kinematics of bilabial voiced implosives in Central Vietnamese, focusing on lip aperture trajectories and gestural timing during stop formation and release. Results indicate that implosives exhibited a higher peak velocity away from closure, whereas their voiceless plosive counterparts were found to have a longer gestural plateau. To investigate whether these differences are inherently tied to implosivity or simply a consequence of voicing, we analysed voiced and voiceless bilabial plosives in Taiwanese Southern Min. The Southern Min data confirmed that voiced plosives did not exhibit the same rapid lip aperture movement, supporting the conclusion that, in the Central Vietnamese data, higher peak velocity away from the stop closure is a defining supralaryngeal feature of implosives.

Index Terms: implosive consonants, supralaryngeal kinematics, Vietnamese phonetics

1. Introduction and background

Implosive consonants (also termed 'suction stops') are typically described as stops produced with a glottalic ingressive airstream, where larynx lowering during closure generates a pressure differential [1, 2]. This mechanism is assumed to reduce intraoral pressure, leading to ingressive airflow. However, the extent to which such airflow is a defining characteristic remains debated. Ladefoged [3] describes the action of the larynx in implosive production as that of a "leaky piston," noting that the leakage is often significant enough that the airstream is not truly ingressive. Similarly, Clements and Osu [4] argue that implosives should not be defined by the presence of ingressive airflow or a reduction in intraoral pressure but rather by the absence of an increase in pressure. This distinction categorises them as sonorants rather than obstruents phonologically. A study by Nihalani [5], using a modified pneumotachographic setup, reported that Sindhi implosives exhibit measurable ingressive airflow, whereas Hausa implosives do not, suggesting that the presence of ingressive airflow may vary across languages.

While implosives are often analysed in terms of their laryngeal mechanisms, larynx lowering is not exclusive to implosives; it is also associated with voiced stops more generally. This complicates efforts to distinguish implosives from voiced plosives based solely on laryngeal behaviour. Oh et al. [6] recently proposed that the distinction between voiced plosives and voiced implosives lies not in the presence of larynx lowering itself but in its timing—specifically, that larynx lowering must occur after oral closure to generate the necessary suction for implosives.

If implosives cannot be fully distinguished from voiced plosives based on their laryngeal properties alone, their supralaryngeal articulation must also be considered. However, while much of the literature on implosives has focused on their aerodynamic properties, particularly airflow and pressure differentials, their supralaryngeal kinematics remain underexplored. A more detailed investigation of supralaryngeal dynamics is therefore necessary to provide a fuller understanding of the articulatory properties of implosives.

This study examines the spatial and temporal properties of lip movements associated with bilabial implosives in Central Vietnamese (CVT). To investigate whether kinematic differences are inherently tied to implosivity or simply a secondary effect of voicing, we compare them to those of voiced and voiceless bilabial plosives in Taiwanese Southern Min (TSM), which serves as a control. Generalised additive mixed models (GAMMs) were used to compare lip aperture trajectories over time, while linear mixed-effects regression models were used to analyse key temporal variables within a gestural framework, including gestural plateau duration, peak velocity toward closure, peak velocity away from closure, and gesture amplitude.

2. Methods

2.1. Participants, stimuli and recording procedures

Three native speakers of CVT and three speakers of TSM, all in their early twenties, participated in this study. The Vietnamese materials included bilabial implosives and voiceless plain explosive stops in the vowel contexts of /a/, /i/, and /u/. For Taiwanese Southern Min, the materials featured the same vowel contexts, contrasting voiced and voiceless stops. The target words for the Vietnamese data were embedded in the carrier phrase ___. *Nói âm* ___, meaning '___. Say the sound ___.', and the Taiwanese Southern Min data in the carrier phrase *Ka* ___ *la*, meaning 'Teach ___ [sentence-final particle]'. Participants read the target words in randomised order from a screen in a soundproof room. Taiwanese Southern Min speakers completed seven repetitions of each word, while Central Vietnamese speakers completed six repetitions.

Articulatory movements for the Central Vietnamese data were recorded using the Carstens AG501 electromagnetic articulograph at a sampling rate of 2,000 Hz, later down-sampled to 250 Hz. For the Taiwanese Southern Min data, recordings were made using the NDI Wave system at a sampling rate of 100 Hz. Sensors were attached to the lips, tongue, and lower incisor (for tracking jaw movement), as well as to the right and left mastoid processes and upper incisor (to correct for head movement). The primary focus of the current study is lip aperture (LA), defined as the Euclidean distance between the upper and lower lip sensors.

2.2. Articulatory measurements

For the analysis of lip trajectories over time, each token was segmented in MVIEW [7] by visually identifying the midpoint of the vowel preceding and following the bilabial closure. The vowel midpoints were determined by locating regions of steady formant activity within the vocalic portions. For the analysis of temporal variables, articulatory gestures were identified using the *findgest* algorithm in MVIEW, which detects gestural landmarks based on a velocity threshold of 20%.

2.3. Statistical analysis

Dynamic articulatory data were time-normalised, z-scored within speakers, and analysed using generalised additive mixed models (GAMMs) in R, based on the recommendations of Wieling et al. [8]. The GAMMs were fitted using the `bam()` function [9] in R to model articulatory trajectories over normalised time.

For the gestural analysis, outliers were removed using the interquartile range (IQR) method within each speaker and vowel group, following a 1.5 × IQR criterion. Linear mixed-effects models were used to analyse differences in four gestural variables: peak velocity towards and away from closure, gestural plateau, and amplitude. Each model included stop type as a fixed effect and vowel as a random intercept, with the following model structure applied to all variables:

$$\text{gestural variable} \sim \text{type} + (1|\text{vowel}) \tag{1}$$

The inclusion of a random intercept for speaker resulted in a singular fit, likely due to the data being z-scored within speakers, which reduced between-speaker variability.

Post hoc pairwise comparisons were performed using the `emmeans` package [10], with Tukey-adjusted p-values to control for multiple comparisons.

3. Results

3.1. GAMM analysis

Figures 1–3 show the GAMM smooth and difference plots for the upper and lower lips in the Z (vertical) and X (horizontal) axes across the three syllable pairs (explosive vs. implosive) in the three vowel contexts (/a/, /i/, and /u/). The smooth plots on the left display the estimated articulatory trajectories over time, with shaded confidence intervals. The difference plots on the right illustrate the pointwise differences between the two models at each time step, highlighting regions where articulatory trajectories diverge.

No statistically significant differences were found in the lip trajectories between implosive and explosive stops. Given the nature of bilabial closures, where full lip contact constrains articulatory variability, this result is not entirely unexpected. Unlike coronal stops, where the tongue has more flexibility in articulation, bilabial stops allow for less spatial variability due to the fixed nature of lip closure.

The primary motivation for this analysis was to determine whether differences in the forcefulness of lip closure could be detected through EMA by examining the distance between upper and lower lip sensors. If one stop type involved a tighter or more forceful closure, this might have been reflected in subtle differences in lip aperture. However, no such differences were observed in the present data.

While GAMMs provide a useful spatial representation of articulatory movement, they may be less sensitive to fine-grained temporal distinctions such as velocity differences at specific points in the movement trajectory. Given this, the following gestural analysis examines temporal kinematic variables—gestural plateau, amplitude, and peak velocity—to provide a more precise investigation of the temporal aspects of implosive production.

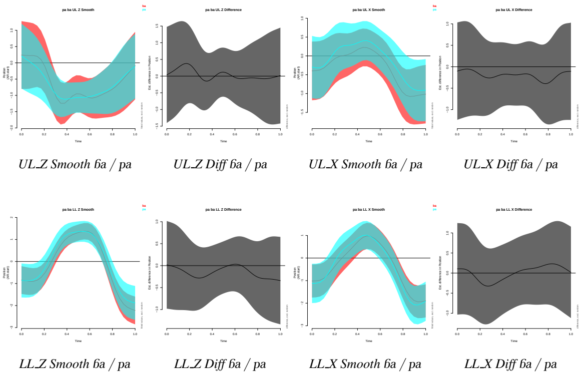


Figure 1: *GAMM smooth and difference plots for ɓa / pa*

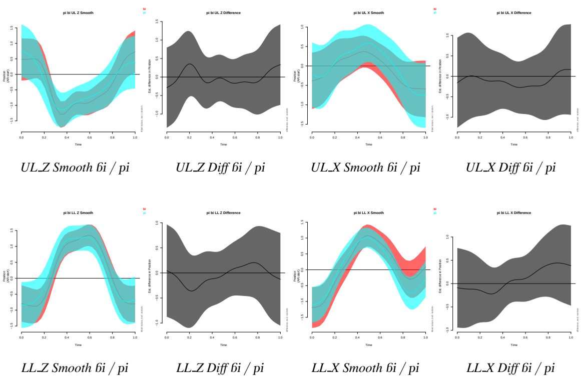


Figure 2: *GAMM smooth and difference plots for ɓi / pi*

3.2. Gestural analysis

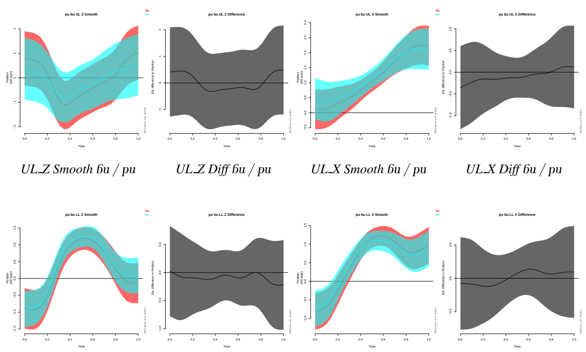
Box plots for the variables of interest are shown in Figures 4–7. Post-hoc pairwise comparisons were conducted using estimated marginal means (EMMs) with Tukey adjustments for multiple comparisons, the results of which are presented in Tables 1–4.

3.2.1. Amplitude

No significant differences in amplitude were found between the three stop types (Table 1), and no clear patterns emerged either between or within speakers of TSM and CVT or across different vowel contexts.

3.2.2. Peak velocity towards closure

Again, no significant differences were found in peak velocity towards closure (Table 2). Similarly, no consistent patterns emerged either between or within speakers of TSM and CVT or across different vowel contexts.



UL_Z Smooth ɓu / pu *UL_Z Diff ɓu / pu* *UL_X Smooth ɓu / pu* *UL_X Diff ɓu / pu*

LL_Z Smooth ɓu / pu *LL_Z Diff ɓu / pu* *LL_X Smooth ɓu / pu* *LL_X Diff ɓu / pu*

Figure 3: *GAMM smooth and difference plots for ɓu / pu*

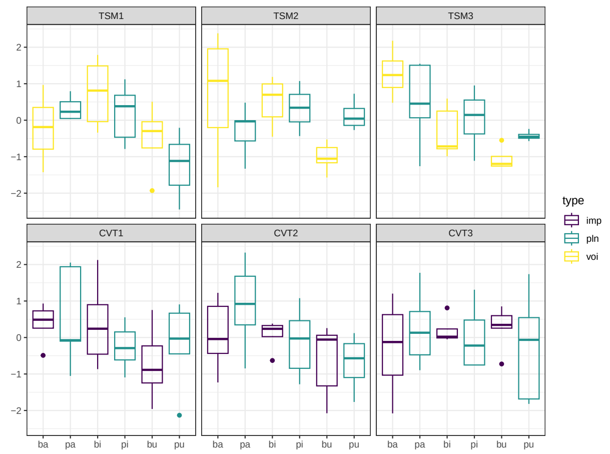


Figure 4: *Amplitude; x-axis: syllable type, y-axis: within-speaker z-scored measurement; top three panes: TSM speakers, bottom three panes: CVT speakers; stop type indicated in legend*

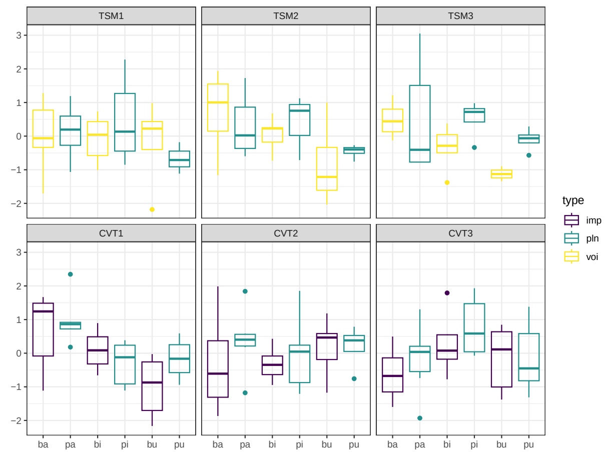


Figure 5: *Peak velocity toward closure; x-axis: syllable type, y-axis: within-speaker z-scored measurement; top three panes: TSM speakers, bottom three panes: CVT speakers; stop type indicated in legend*

Table 1: *Pairwise comparisons for amplitude.*

Contrast	**Estimate**	**SE**	**df**	**t-ratio**	**p-value**
ɓ - p	-0.0120	0.164	176	-0.073	0.9971
ɓ - b	-0.0304	0.195	176	-0.156	0.9867
p - b	-0.0184	0.174	176	-0.106	0.9939

Table 2: *Pairwise comparisons for peak velocity towards closure.*

Contrast	**Estimate**	**SE**	**df**	**t-ratio**	**p-value**
ɓ - p	-0.2931	0.168	176	-1.748	0.1906
ɓ - b	0.0196	0.199	176	0.098	0.9947
p - b	0.3127	0.178	176	1.758	0.1869

3.2.3. Peak velocity away from closure

Results indicated that implosives exhibited significantly higher peak velocity away from closure compared to both voiced and voiceless plosives (Table 3). This effect appears reduced in the context of the vowel /u/, which may be due to the articulatory demands of lip rounding.

Table 3: *Pairwise comparisons for peak velocity away from closure*

Contrast	**Estimate**	**SE**	**df**	**t-ratio**	**p-value**
ɓ - p	0.600	0.0758	176	7.922	< 0.0001
ɓ - b	0.431	0.0898	176	4.799	< 0.0001
p - b	-0.169	0.0804	176	-2.107	0.0913

3.2.4. Gesture plateau duration

The results showed that voiceless plosives had a significantly longer gestural plateau compared to both implosives and voiced plosives, while the latter two did not differ significantly (Table 4).

3.2.5. Summary of results

A summary of the results is shown in Figure 8, where compact letter displays (CLDs) indicate statistical groupings. This was generated using the `multcomp` package [11], with Sidak adjustments applied to control for multiple comparisons. As shown in the legend, darker colours represent higher relative Z-scores for stop type within each variable.

Peak velocity away from closure was significantly higher in voiced implosives. Voiced plosives had a slightly higher peak velocity away from closure, though this difference was not significant. Voiceless plosives had a slightly higher peak velocity towards closure than the other stop types, but this also did not reach significance. Plateau duration was significantly longer in voiceless plosives than in the other two stop types. Finally, no significant differences were found between stop types for amplitude.

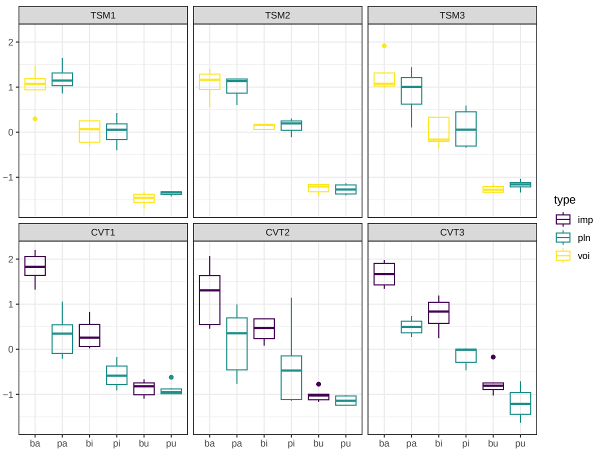


Figure 6: *Peak velocity away from closure; x-axis: syllable type, y-axis: within-speaker z-scored measurement; top three panes: TSM speakers, bottom three panes: CVT speakers; stop type indicated in legend*

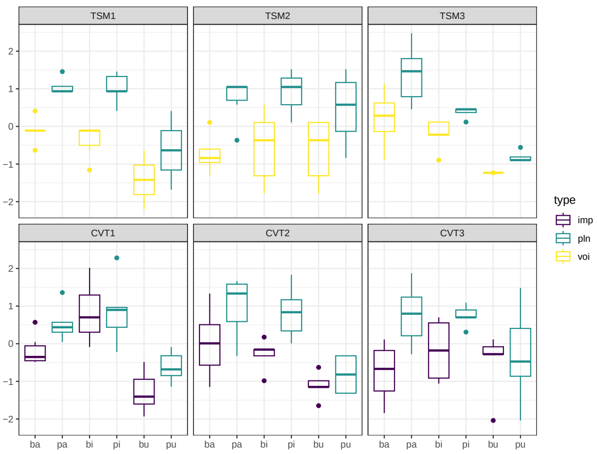


Figure 7: *Plateau duration; x-axis: syllable type, y-axis: within-speaker z-scored measurement; top three panes: TSM speakers, bottom three panes: CVT speakers; stop type indicated in legend*

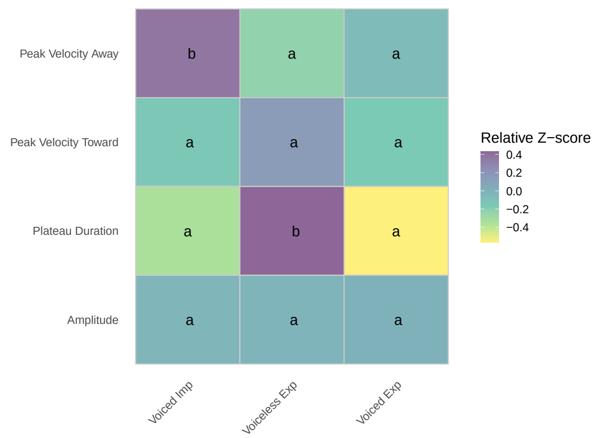


Figure 8: *Summary of results; CLD letters indicate statistically similar groups based on multiple comparisons*

Table 4: *Pairwise comparisons for gestural plateau duration.*

Contrast	Estimate	SE	df	t-ratio	p-value
ɓ - p	-0.772	0.128	176	-6.033	< 0.0001
ɓ - b	0.224	0.152	176	1.474	0.3060
p - b	0.995	0.136	176	7.337	< 0.0001

4. Discussion

The results of the GAMM analysis indicated no significant differences in lip trajectories between implosives and plosives. Similarly, amplitude did not differ across stop types. Among the temporal variables analysed, peak velocity towards closure did not differ significantly across stop types. However, gestural plateau duration was significantly longer for voiceless plosives compared to voiced implosives and voiced plosives. Crucially, the defining difference in the CVT data was that peak velocity away from closure was significantly higher for implosives.

These findings suggest that implosives in CVT involve a more rapid articulatory release compared to other stop types. While this could result from aerodynamic suction effects or local muscular adjustments at the lips, the exact mechanism remains an open question.

Future research should expand the speaker sample and include tongue and jaw movement to provide a more comprehensive articulatory analysis. Additionally, investigating alveolar implosives could determine whether the increased release velocity observed in bilabial implosives is a general property of implosives in CVT.

5. Acknowledgements

This study was partially funded by research grant 103-2410-H-007-036-MY3 from the Ministry of Science and Technology, Taiwan.

6. References

[1] K. L. Pike, “Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds.” 1972.

[2] J. C. Catford, *Fundamental Problems in Phonetics*, reprint ed. Bloomington, IN: Indiana University Press, 1977.

[3] P. Ladefoged, *Preliminaries to linguistic phonetics*. University of Chicago press, 1971.

[4] G. N. Clements and S. Osu, “Explosives, implosives and nonexplosives: the linguistic function of air pressure differences in stops,” *Laboratory phonology*, vol. 7, pp. 299–350, 2002.

[5] P. Nihalani, “Phonetic implementation of implosives,” *Language and Speech*, vol. 29, no. 3, pp. 253–262, 1986.

[6] M. Oh, D. Byrd, L. Goldstein, and S. S. Narayanan, “Vertical larynx actions and intergestural timing stability in hausa ejectives and implosives,” *Phonetica*, vol. 81, no. 6, pp. 559–597, 2024.

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

[8] M. Wieling, “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*, vol. 70, pp. 86–116, 2018.

[9] S. N. Wood, *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*, 2024, r package version 1.9-1. [Online]. Available: https://CRAN.R-project.org/package=mgcv

[10] R. V. Lenth, H. Singmann, J. Love, P. Buerkner, and M. Herve, *Package "Emmeans"*, 2018, r package version 4.0-3. [Online]. Available: http://cran.r-project.org/package=emmeans

[11] T. Hothorn, F. Bretz, and P. Westfall, "Simultaneous inference in general parametric models," *Biometrical Journal*, vol. 50, no. 3, pp. 346–363, 2008.