A new model for tonal-segmental interaction in Standard Mandarin

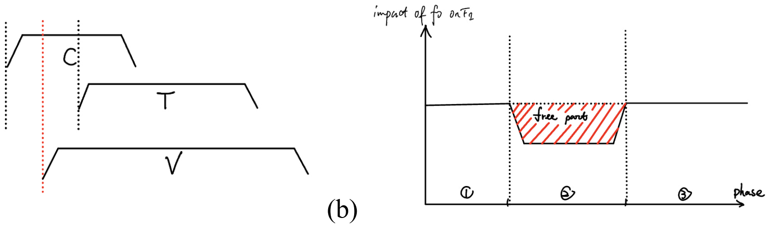
Chenyu Li[1], Jalal Al-Tamimi[1]
[1] *Université Paris Cité, LLF, CNRS*
chenyu.li@etu.u-paris.fr; jalal.al-tamimi@u-paris.fr

Introduction This paper introduces a new model of "dual mechanism" for tonal-segmental interaction in Standard Mandarin, tested with the diphthongs /ai/ and /au/ data from a reading corpus. The interaction between tones (contours of *f*0) and vowels has been extensively studied, with accumulating evidence of intrinsic *f*0 (e.g., [1]) and the tonal effect on vocalic realization of simple vowels (e.g., [2,3]) or of diphthongs [4,5,6]. This interaction has been verified as a common and inherent phenomenon. However, existing research on simple vowels all used data taken from the midpoint of production, and the dynamic changes of *f*0-F1 interaction over time are not clear. [7] treated tones as a consonantal gesture in terms of gestural coordination within the Articulatory Phonology framework [8]. In a CV syllable in Mandarin, the tonal gesture (T) tends to show an *in-phase* coordination with the vowel gesture (V), and an *anti-phase* coordination with the consonant gesture (C) leading to a *C-centre* effect on their temporal coordination [7] (Fig. 1a). We thus hypothesize that the vowel realization has three phases (Fig. 1b). In Phase 1, a purely physiological linear relationship between *f*0 and F1 is observed due to the tonal part not being yet activated, with *f*0 being controlled only by the vowel gesture; in Phase 2, due to the increased impact of the tonal gesture, we hypothesize that the impact of *f*0 on F1 is assumed to be gradually reduced or even reversed due to the joint control on *f*0 by tonal and vowel gestures; in Phase 3, the part where the release from the target of the tonal gesture starts, the control on *f*0 by the tonal gesture will fade gradually and *f*0 will display a "purely physiological" interaction with F1.

Methods The data was composed of recordings of 20 participants (10 females and 10 males) from a reading corpus *AISHELL-1* [9]. We analyzed the samples of /ai/ and /au/ from the non-initial mono/disyllabic words without glides. The data were automatically segmented and aligned at the syllable and the phoneme levels, using the *Montreal Forced Aligner* [10]. Segmental boundaries, word information, acoustic information (*f*0 and F1), tonal information, segmental information, and sentence information were automatically measured and extracted in *Praat* [11]. This yielded a data set of about 3000 occurrences for each diphthong. For each occurrence, we obtained 11 time-normalized intervals. Using GAMMs [12], we modelled the interaction between the two continuous predictors: *f*0 and *duration*, with the outcome being F1. We used *time* as a continuous predictor, represented via a *smooth* as a non-linear variable, to track the dynamic patterns during the diphthong realization. *Speaker* and *word* were considered as random *smooths* adjusted by the preceding tone.

Results We used the *plot_diff* function to quantify the mutual influence of *f*0 on F1 over time. We set the high pitch at the 80% interval of *f*0 and the low pitch at 20%. The output is the difference of F1 between two *f*0 values, which can represent the impact of *f*0 on F1 (Fig. 2, /ai/ top; /au/ bottom). Fig. 2a, 2b, 2c, and 2d show that the F1-difference in the low-high range of *f*0 is negative in Phase 1 and approaches/crosses the 0 value in Phase 2 for all cases confirming our predictions formulated above. In Phase 3, the F1-difference comes back to negative in the case of /au/ (Fig. 2c and 2d) and /ai/ for male speakers (Fig. 2a).

Conclusion The modelling results are generally consistent with the hypothesis for the two diphthongs: this "three-phase" pattern can be explained by the "dual-mechanism" control within the tonal effect. *f*0 impact on F1 shows a negative F1-difference in Phases 1 and 3, indicating an inherent negative correlation; the impact declines or reverses its direction in Phase 2, due to the "decoupling" of the laryngeal-supralaryngeal interaction due to the joint control of tonal and vowel gestures in this phase. The positive correlation in Phase 3 for /ai/ in the female case indicates a potential for a gender-specific difference in diphthong realization: female speakers tend to have more dynamic and less monophthongized realizations of diphthongs, found previously in [13,14], which needs further exploration in the future.



(a) (b)

Fig. 1. (a) Gestural organization of CV syllables in Mandarin. (b) Predicted *f*0 impact on F1 in three Phases.

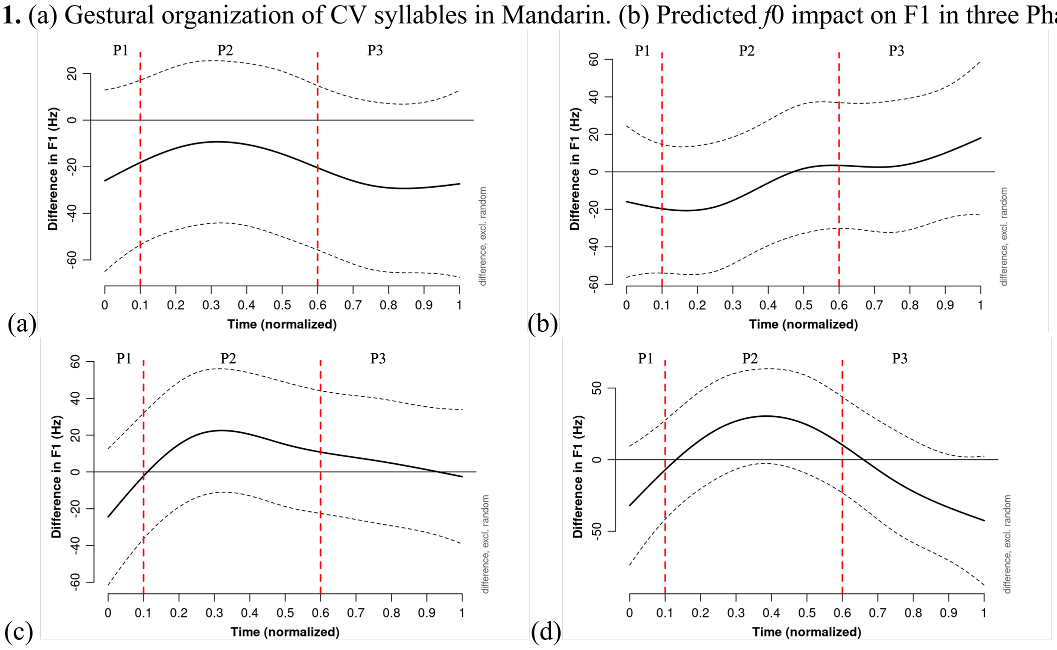


Fig. 2. The F1-difference between high pitch and low pitch. The thick red dashed vertical lines show the limits of three phases (P1, P2, P3). (a): /ai/ male. (b): /ai/ female. (c): /au/ male. (d): /au/ female.

References

[1] Chen, W.-R., Whalen, D. H., & Tiede, M. K. (2021). A dual mechanism for intrinsic f0. *Journal of Phonetics*, 87, 101063.

[2] Erickson, D., Iwata, R., Endo, M., & Fujino, A. (2004). Effect of tone height on jaw and tongue articulation in Mandarin Chinese. In *International symposium on tonal aspects of languages: With emphasis on tone languages.*

[3] Shaw, J. A., Chen, W.-r., Proctor, M. I., & Derrick, D. (2016). Influences of tone on vowel articulation in Mandarin Chinese. *Journal of Speech, Language, and Hearing Research*, 59(6), S1566–S1574.

[4] Li, C., Al-Tamimi, J., & Wu, Y. (2023). Tone as a factor influencing the dynamics of diphthong realizations in Standard Mandarin. In *Radek Skarnitzl & Jan Volín (Eds.), Proceedings of the 20th International Congress of Phonetic Sciences.*

[5] Li, C., & Al-Tamimi, J. (forthcoming). The impact of the tonal factor on diphthong realizations in Standard Mandarin: Explanations within the Articulatory Phonology framework modelled via Generalized Additive Mixed Models.

[6] Li, C., & Al-Tamimi, J. (under review). Impact of the tonal factor on diphthong realizations in Standard Mandarin with Generalized Additive Mixed Models.

[7] Gao, M. (2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics)*, Yale University, CT.

[8] Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2-4), 140–155.

[9] Bu, H., Du, J., Na, X., Wu, B., & Zheng, H. (2017). AIShell-1: An Open-Source Mandarin Speech Corpus and A Speech Recognition Baseline. In *Oriental COCOSDA 2017*.

[10] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Proc. Interspeech 2017 (pp. 498–502)*.

[11] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int., 5(9), 341–345.*

[12] Wood, S. N. (2006). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

[13] Thomas, J. (2011). Gender and /ai/ monophthongization in African American English. In *Annual Meeting of the Berkeley Linguistics Society* (pp. 449-463).

[14] Simpson, A. P. (2002). Gender-specific articulatory–acoustic relations in vowel sequences. *Journal of Phonetics,* 30(3), 417-435.