Interspeech 2024
1-5 September 2024, Kos, Greece

Impact of the tonal factor on diphthong realizations in Standard Mandarin with Generalized Additive Mixed Models

Chenyu Li[1], *Jalal Al-Tamimi*[1]

[1]Université Paris Cité, CNRS, LLF, France

chenyu.li@etu.u-paris.fr, jalal.al-tamimi@u-paris.fr

Abstract

This study examines the effects of lexical tones on diphthong realizations in Standard Mandarin. We investigated the two falling diphthongs /ai/ and /au/ from a Standard Mandarin reading text corpus. A set of GAMMs models was employed to test whether and how tones and f0 influence the diphthong realizations. The results show the vowel height (reflected by F1) to differ with respect to tones: with a high tone, the diphthongs tend to be realized as more closed, and as more open with a low tone; with a rising tone, they tend to have a typical diphthongized realization, with a dynamical pattern of F1 contour; and to be monophthongized with a falling tone. The interaction between f0 and F1 extensively confirmed in monophthongs across languages, is equally applicable to /ai/ and /au/: f0 is negatively correlated with F1. The results show the universality of the tonal effect on vowel realization in different diphthongs and imply physiological factors behind it.

Index Terms: diphthong, Standard Mandarin, tone, f0, Generalized Additive Mixed Models

1. Introduction

The relationship between segmental and suprasegmental features has been the subject of research in a few cross-linguistic articulatory and/or acoustic studies, with the aim to shed light on how segmental features, as the height or backness of the vowels, interacts with suprasegmental information, as stress or tones [1]. For instance, it has been shown that the intrinsic fundamental frequency (f0) is attested in many languages, with accumulating evidences, including tonal [2, 3, 4, 5], or languages without lexical tones, such as most Indo-European languages [6], which suggests a general tendency for high vowels to have higher f0 and vice-versa [7]. Regarding the tonal impact on vowel realization, which is referred to as the tonal-segmental interaction within tonal languages, different studies have evaluated it from an articulatory [8, 9, 10] and from an acoustic [2, 11] view points. In general, a higher f0 will cause the vowel, especially /a/ to be realized as higher and more front. The tonal effect on high vowels such as /u/ and /i/, often has an “inverse effect”, which is possibly due to different mechanisms of larynx-vocal tract linkage [7] or due to the phonological control by the speaker in the high vowel range [8]. Experimental research has unified these two opposing influences into a co-function between the vertical movement of the larynx and the movement of the vocal tract, e.g., the co-function of the jaw, hyoid, laryngeal cartilage, and cervical spine [12, 13, 14, 7, 15].

Most of the research looking at the possible correlation between f0 and vowels examined monophthongal cases. However, when considering a dynamic situation, e.g., diphthongs, this correlation remains unexplored, except in a handful of studies. For instance, an indirect link between the pitch and diphthongs in perception was proposed by [16], whereby German speakers tend to perceive the closing diphthongs as a more falling pitch sequence than the opening diphthongs, which is opposite to their intrinsic f0. The relationship between f0 curves and diphthongs in the case of the interaction between complex tones and complex segments was evaluated for the first time in Cantonese by [3]. However, the researchers failed to identify a systematic relationship between the two.

In line with the research reported above, our previous work showed that Standard Mandarin exhibited a correlation between tone (f0) and the vowel realization in the diphthong /ai/ [17], which provided partial correlations with predictions from previous research, on a small scale dataset with limited generalizations. However, it remains unclear how systematic are the impacts of tonal factors on other falling diphthongs. For instance, Standard Mandarin has an /ai/ and an /au/ falling diphthongs. We ask the question whether the patterns observed on /ai/ by [17] are restricted to this diphthong or whether using another falling diphthong /au/ will exhibit a common mechanism that allows to control such segmental-tonal interaction.

The present study has a two-fold purpose. One aim is to explore realization differences between the tonal categories across the different falling diphthongs in Standard Mandarin. The second aim is to use multi-dimensional modelling via Generalized Additive Mixed-effects Models (GAMMs) to allow for a fine-grained evaluation of the tonal impact in terms of the interaction between f0 and F1 (first formant). In Standard Mandarin, there are four lexical tones (with Chao digits): tone 1 (high tone - 55), tone 2 (rising tone - 35), tone 3 (low tone - 21 in natural speech) and tone 4 (falling tone - 51) [18].

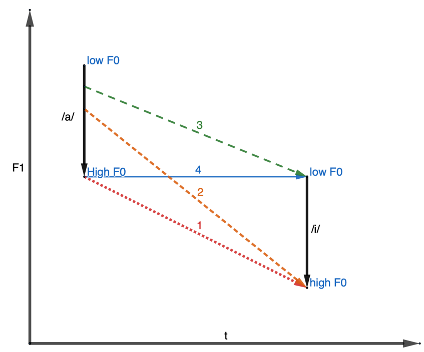


Figure 1: *Schematic diagram of F1 contour with four different tones (1, 2, 3 and 4), with low and high f0, in Standard Mandarin.*

10.21437/Interspeech.2024-303

Following previous research, we use F1 as the acoustic correlate of vowel height due to the inverse relationship between F1 and the degree of vocal tract constriction [19]. Figure 1 shows the predicted F1 contours over time for the four tones realized in the diphthongs /ai/ and /au/. Firstly, we hypothesize that the vowel realization will be affected by a similar tonal effect: a high tone (Figure 1, tone 1) will make the diphthongs to be produced as more closed, with the lowest F1 contour than that of other tones; a low tone (Figure 1, tone 3) will result in a more open realization, with the highest F1 contour than that of other tones; a rising tone (Figure 1, tone 2) will strengthen the dynamic feature of the diphthongs, where the F1 contour will be the steepest; and finally, a falling tone (Figure 1, tone 4) will weaken such feature and leads to monophthongization, shown by a relatively flatter F1 contour. As for the relation between f0 and vowel realization, we then hypothesize that the negative correlation between f0 and F1 found in simple vowel will be equally applicable to diphthongs.

2. Model construction

2.1. Corpus and Data processing

Our corpus was composed of recordings from 10 male speakers obtained from an open-source Standard Mandarin reading text corpus *AISHELL-1*, published by Beijing Shell Shell Technology Co., Ltd. [20]. The data were automatically segmented and aligned at the syllable (character) and the phoneme levels, using the *Montreal Forced Aligner* (MFA) [21]. Only lexical tones (1 to 4) in Standard Mandarin are included in the lexicon. The word tokenization information was integrated into the TextGrid files after alignment, using the tokenizer *Jieba* [22] and was checked manually. We selected the occurrences that are not the first syllable of the sentence and from monosyllabic words or disyllabic words. We chose syllables that were either onsetless (i.e., started with the diphthong) or excluded glides (i.e., started with any consonant). This yielded a data set of 2697 occurrences: 1555 for /ai/ and 1142 for /au/. For each occurrence, tonal information of the target syllable, in addition to the tone preceding the target tone, were obtained to account for the tonal coarticulation between the syllable and the preceding one [23]). We also obtained lexical information such as the full word information, the word length and the position in the word, in addition to sentence information, e.g., relative position in the sentence. Finally, we coded the speaker anonymized identity to model these as random effects.

The acoustic information (f0 and F1) for each occurrence were automatically measured using the automatic script of *Praat* (version 6.3.02) [24]. Formant frequencies were automatically obtained using the *Burg* method (time step = 5 ms, window length = 25 ms, pre-emphasis from 30 Hz, maximum frequency = 5000 Hz for /ai/ and 4400 Hz for /au/, and a maximum of 5 formants). The formant information of all the /ai/ and /au/ items was verified manually. The f0 measurements were by the two-passes method, with auto-correlation (following [25], implemented in [26]); f0 estimations are speaker-dependent based on their range that prevents errors in extraction. The f0 data estimated on a diphthong by diphthong basis are smoothed with a bandwidth of 10 Hz by the smooth function of Praat (following [27], implemented in [26]). For each occurrence, we obtained 11 time-normalized intervals, at 10% intervals for F1 and f0 measurements. Overall, across the two measures (f0, F1), we ended up having 59334 data points (2 measures * 2697 occurrences * 11 measurement points).

2.2. Model specification

We used Generalized Additive Mixed Models (GAMMs) [28] to capture the dynamic patterns of /ai/ and /au/ by modelling the non-linear relation between predictors and an outcome varying in the time domain. We performed two sets of modelling to evaluate how the tone affects these two dynamic sequences from global and fine grained perspectives. Modelling with GAMMs was done using the packages *mgcv* [28] (version 1.8-42, for modelling) and *itsadug* [29] (version 2.4, for visualization) in R computing language (version 4.2.2) [30]. For model specification, we followed recommendations from [31, 32, 33, 34, 35].

The first model was to evaluate the interaction between the diphthong realization on vowel height using F1 dimension and the tonal unit: the categorical predictor was the `tone` with four levels and the outcome was `F1` value. We used `time` as a continuous predictor with 11 normalized intervals, represented via a *smooth* as a non-linear variable, to track the dynamic pattern during the diphthong realization. We integrated full `word` information and `speaker` ID, all adjusted by `pos` (position in the utterance) and `leftTonalTarget` (the tonal offset of the preceding syllable, which was defined as H for tone 1 and tone 2, and L for tone 3 and tone 4, following [36]). `tone` was set as an *ordered* categorical predictor to reduce the Type I error and increase power in the model [33]. Also, the *smooth* terms were adjusted for the variable levels of tones with *by* interaction.

The second model, aiming to uncover the linguistic and physiological mechanisms of tonal effect on diphthong realization, was generated from a fine grained perspective. The phonological item `tone` in the first model was replaced by the combination of `f0` and `duration`, since the tones are acoustically realized by the f0 contour over time [18] and the duration varies with different tones [37]. This acoustic combination representing `tone` was modelled via multidimensional *smooth*s respectively since they were continuous predictors. In the model we examined the relationship between the multi-dimensional continuous predictors `f0`, `duration` and `time` and their interactions using the `ti` tensor, which computed the main effects and the interactions between *smooths*. The outcome in the second model was also F1. The random effects were `word` and `speaker`, adjusted by `leftTonalTarget`.

For the two models, we verified the auto-correlation levels of our models and obtained an Auto-Regressive GAMMs, which was done using the ρ parameter to reduce auto-correlation in the temporal dimension. We evaluated the structure of the models using the function `gam.check` to choose our optimal models.

3. Results

3.1. First model

The purpose of the first model is to evaluate any possible influences of tones as phonological categories on the vowel height realization for the diphthongs /ai/ and /au/. Figures 2 and 3 illustrate the predictions obtained from the first model as a function of the four tones.

From the two figures, we can find that all F1 contours are falling from interval 0.2 to 10, indicating that /ai/ and /au/ are vocalic sequences moving from a low to a high vowel. During the realization of /ai/, F1 dropped by a total of 150Hz (with tone 2); 100Hz for /au/ (with tone 2), which indicates that the vowel distance between targets of /ai/ is greater than the distance between targets of /au/. This can be explained with the physical limitation of the vocal tract space, that is, the vocal tract begins

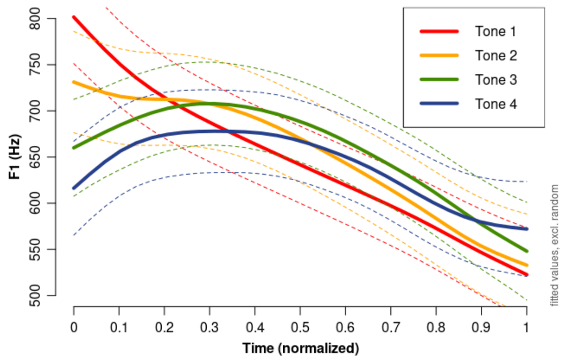


Figure 2: *Predicted F1 contours (y-axis) over time (x-axis) of /ai/. Dashed lines represent confidence intervals.*

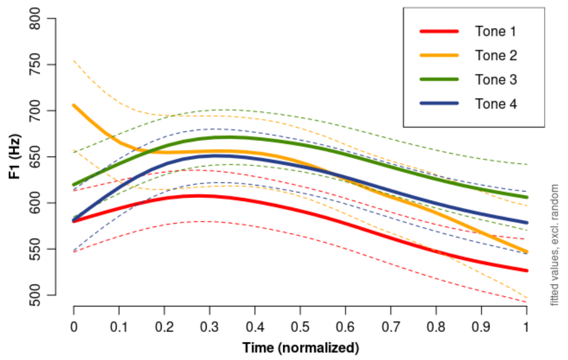


Figure 3: *Predicted F1 contours (y-axis) over time (x-axis) of /au/. Dashed lines represent confidence intervals.*

to curve downward at the soft palate (i.e., the position close to /u/), which may cause the space of tongue movement for /au/ to be weakened, which is also consistent with [8].

The two figures show the similarity of the F1 contour patterns to those shown in Figure 1. The results of the first model show that different overall heights of tones, i.e., tonal registers, will influence the overall range of the vowel quality, making the diphthongs be realized in different levels of aperture. With tone 1 (high level tone), the curve of F1 (red) is generally lower than that in other tones, signalling a more closed realization of the two diphthongs /ai/ and /au/. The curve (green) with tone 3 (low falling tone), on the contrary, is generally higher than that with other tones, which indicates that the two diphthongs would be realized as the most open in this case. The curves with tone 3 have a relatively flat pattern, showing a partial monophthongization. The results with the high tone 1 and low tone 3 echoes previous findings in the monophthong case [8].

As for the complex tonal contours, i.e., tone 2 (rising tone) and tone 4 (falling tone), the predictions by the first model confirmed that the tonal influence of the complex tones will be continuous and non-fixed during the diphthong realization. The curve of F1 (orange) with tone 2 (rising tone) starts relatively higher than that with other tones (around 750 Hz for /ai/ and 700 Hz for /au/) and ends relatively lower (around 550 Hz in both cases), indicating a minimal monophthongization,: the diphthongs /ai/ and /au/ are produced as the most “diphthongized” with tone 2, compared with the other tones. Here, the effect of tone 2 is pulling the tongue downward in the portion of the first target /a/ because of the relatively low f0, and is pushing the tongue upward in the portion of the second target /i/ or /u/ because of the relatively high f0. The effect of tone 4 (falling tone) is the opposite. The curve of F1 with tone 4 (blue) starts relatively low (650 Hz) and ends relatively high (600 Hz), showing a flat pattern which is indicative of a stronger monophthongization. Here, the tone 4 influences the diphthong realization by pushing the tongue upward at the first target /a/ because of the relatively high f0, and pulling the tongue downward at the second target /i/ and /u/ because of the relatively low f0.

Inspecting the *smooth* terms via the function `summary()`, it was evident that the estimates associated with the *smooths* adjusted by the different tones, are significantly different ($p < 0.05$) from each other in /ai/ case, except for tone 3 in /au/ in comparison to tone 1 as the reference level.

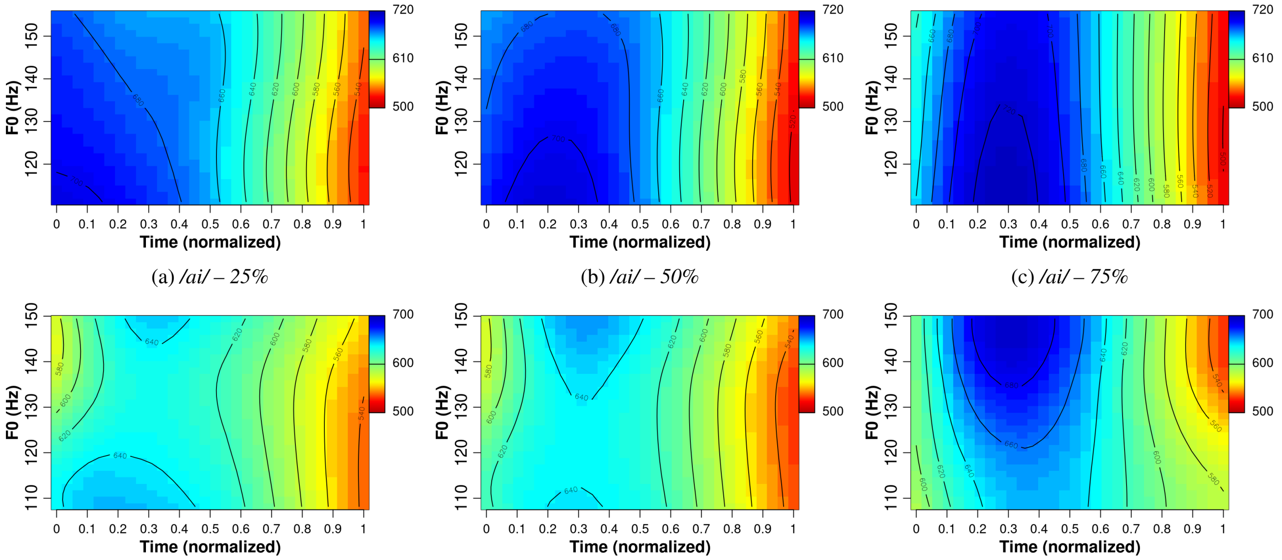
The predictions of the first model generally confirm the first hypothesis: the tonal effect on vowel realization is observed both on the diphthongs /ai/ and /au/ in this study. More concretely, in terms of tonal effect on vowel height, both diphthongs show a similar pattern: different tonal heights and different tendencies of complex tones will lead to similar influences on vowel height or level of monophthongization of /ai/ and /au/.

3.2. Second model

The results of the first model, especially those with tone 2 and tone 4, imply that the impact of the tonal factor is continuous and changing during the vowel realization. The second model gave a more intuitive evidence of the influence of f0 on vowel height and how it contributes to documenting the impact of tone of diphthong realization.

In the second model, we consider the combination of f0, time step and duration, using a more fine-grained model via multidimensional modelling of the impact of tones on the realizations of the diphthong /ai/ and /au/. Recall that in our model, we had three continuous predictors modelled as *smooth* terms as a function of our dependent variable F1. However, it was not possible to easily visualise a 4-dimensional space encompassing the impact of normalized time, f0 and duration on F1 frequencies in each of /ai/ or /au/ diphthongs. To do so, we modelled this interaction in a 3-dimensional space using the function `fvisgam` from *itsadug* R package [29]. Each heatmap figure shows the dependent variable, i.e., F1 value of /ai/ or /au/, which is determined by two different continuous predictors on two axes: the normalized time on the x-axis and the f0 on the y-axis. F1 frequency changes are presented on the z-axis with variable colors on the heatmap, whereby a higher frequency value of F1 is denoted by the blue end of the color scale, and a lower frequency value of F1 is denoted by the red end of the color scale. To visualise the impact of `duration` on the second model as a fourth dimension, we selected three intervals to quantify the distribution of the diphthong at the three quartiles, i.e., 25%, 50% and 75% points of the duration, leading to respectively 110 ms, 130 ms, and 170 ms for both /ai/ and /au/ cases.

Figure 4 shows the second model's predictions for `F1` value that co-varies with both `f0` and the `normalized time`. The results point to an overall lowering of F1 frequencies throughout the realization of the diphthongs /ai/ and /au/, which is expected due to these being dynamic vowel sequences. Our results highlight the continuous change of F1 under the effect of f0 over time. Figures 4a, 4b, and 4c demonstrate the predictions for /ai/. The contour lines (which show the same value of F1) and color differences show that when f0 rises, F1 falls especially



(a) */ai/ – 25%* (b) */ai/ – 50%* (c) */ai/ – 75%*

(d) */au/ – 25%* (e) */au/ – 50%* (f) */au/ – 75%*

Figure 4: *F1 with different f0 (y-axis) over normalized time (x-axis), predicted by the second model. Each sub-figure represents a duration interval. Blue coloured regions indicate an increase in F1 frequencies; red indicates a decrease.*

towards the beginning (intervals 0 - 0.2) and during the transition parts (intervals 0.4 - 0.9). In the offglide transition portion (intervals 0.4 - 0.9), especially at the 25% and 50% of *duration* (Figures 4a, and 4b), we can observe a slight positive correlation between f0 and F1 within the middle range of f0 around 130 Hz. Figures 4d, 4e, and 4f show similar patterns for /au/: when f0 rises, F1 falls, especially at the beginning (intervals 0 - 0.2), within the 25% and 50% of the *duration*) and in the target and transition parts (intervals 0.4 - 0.9). A positive correlation between f0 and F1 is seen in /au/ located around 140 Hz at 25% and 50% of the *duration* (Figures 4d, 4e).

The second model confirms that the negative correlation between f0 and F1 is stronger in the dynamic parts, like the transition part (intervals 0.4 - 0.9) between targets than in the target of the diphthongs, especially the first target since it is syllabic for a falling diphthong. This result unifies studies of tonal impact on different categories of vowels. The linkage between the larynx and vocal tract (tongue, jaw) considered as a non-phonological physiological factor, causes monophthongal and diphthongal vowel realization to be modulated by the tone at a macro phonological level.

4. Summary and discussion

In this paper, we were interested in the realization of the two diphthongs /ai/ and /au/ in Standard Mandarin with different tones. We wanted to evaluate if the two diphthongs would show similar patterns in terms of the impact of the tonal factor on diphthong realizations, on the one hand, and of the mechanism behind this impact including the effect of f0, on the other hand.

The first model's predictions fit well with our hypothesis about how the vowel height changes for the falling diphthongs with different tones. Overall, the tone will impact the diphthong's vowel height in terms of tonal register and tonal tendency of the complex tone. Similarly to what was already identified in monophthongs [8], our results showed that if the diphthong is associated with a high tone (tone 1), it will be produced as more closed; if it is associated with a low tone (tone 3), it will be produced as more open. Results with the complex tonal contours show that if it is associated with a rising tone (tone 2), the diphthong will be realized as more diphthongized, and it will be produced with a stronger monophthongization with a falling tone (tone 4). Such results imply that the tonal impact on diphthong realization is continuous and changing as it is being realized. The predictions of the second model provided direct evidence for this implication that f0 and F1 follow generally a negative correlation. The predictions of the models show a good fit to the data and allowed to verify the hypotheses with respect to effects of tones and f0, independently of each other.

The positive correlation observed in the second model can be explained by using the Articulatory Phonology framework [38] and coordination between tonal and segmental gestures [39, 40]. As the tonal gesture being coordinated with the vowel gesture in the syllable, coarticulation between the two gestures will occur. Since the f0-F1 interaction tends to be purely physiological [13], the existence of the tonal gesture may weaken this connection due to the phonological control of the tonal gesture on f0 contour. This leads to differences in the level of impact of f0 on F1 frequencies. For instance, negative correlation between the two tends to occur within the transitional parts (at the beginning or ending) but not at the syllabic target where the tonal gesture is activated (i.e., first component of the diphthong). Following [39], the falling tone (tone 4) and the rising tone (tone 2) are all represented by the combination of two tonal gestures: H and L. The backlog of the number of gestures will inevitably lead to an extension of the duration of the offset phase, causing the phonological control part - the positive correlation part to also appear in the middle of the f0 interval of the transitional part, because the tone 2 and tone 4 will gradually approach the middle of the f0 interval in actual pronunciation. We will attempt to incorporate the current study into Articulatory Phonology framework and explain the mechanism of coordination and coarticulation between the tones and diphthongs in further research.

5. Acknowledgements

This work was supported by the Laboratoire de Linguistique Formelle, by a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex Empirical Foundations of Linguistics - EFL. Thanks to the *China Scholarship Council* for PhD funding. We gratefully acknowledge support from the CNRS/TGIR HUMA-NUM and IN2P3 Computing Center (Lyon - France) for providing High Powering Computing and data-processing resources needed for this work.

6. References

[1] Y. Xu, "Prosody, tone and intonation," *The Routledge handbook of phonetics*, pp. 314–356, 2019.

[2] P. Wang, *A Statistical Study on the Tones and Vowels of Beijing Dialect*. PhD Thesis, Nankai University, 2007.

[3] J. Siddins, "The relationship between intrinsic f0, intrinsic pitch and lexical tone in Hong Kong Cantonese," Ph.D. dissertation, Université de Munich, 2017.

[4] J.-M. Hombert, "Consonant types, vowel height and tone in Yoruba," *UCLA Working Papers in Phonetics*, vol. 33, pp. 40–54, 1976.

[5] B. Connell, "Tone languages and the universality of intrinsic F0: evidence from Africa," *Journal of Phonetics*, vol. 30, no. 1, pp. 101–129, 2002.

[6] D. H. Whalen and A. G. Levitt, "The universality of intrinsic f0 of vowels," *Journal of phonetics*, vol. 23, no. 3, pp. 349–366, 1995.

[7] W.-R. Chen, D. H. Whalen, and M. K. Tiede, "A dual mechanism for intrinsic f0," *Journal of Phonetics*, vol. 87, p. 101063, 2021.

[8] J. A. Shaw, W.-r. Chen, M. I. Proctor, and D. Derrick, "Influences of tone on vowel articulation in Mandarin Chinese," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 6, pp. S1566–S1574, 2016.

[9] F. Hu, "Tonal effect on vowel articulation in a tone language," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, 2004.

[10] P. Hoole and F. Hu, "Tone-vowel interaction in standard Chinese," in *International symposium on tonal aspects of languages: With emphasis on tone languages*, 2004.

[11] D. Erickson, R. Iwata, M. Endo, and A. Fujino, "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*, 2004.

[12] K. Honda, "Laryngeal and extra-laryngeal mechanisms of F0 control," *Producing speech: contemporary issues*, pp. 215–232, 1995.

[13] K. Honda, H. Hirai, S. Masaki, and Y. Shimada, "Role of vertical larynx movement and cervical lordosis in f0 control," *Language and Speech*, vol. 42, no. 4, pp. 401–411, 1999.

[14] D. H. Whalen, B. Gick, M. Kumada, and K. Honda, "Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0," *Journal of Phonetics*, vol. 27, no. 2, pp. 125–142, 1999.

[15] S. R. Moisik, H. Lin, and J. H. Esling, "A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS)," *Journal of the International Phonetic Association*, vol. 44, no. 1, pp. 21–58, 2014.

[16] O. Niebuhr, "Intrinsic pitch in opening and closing diphthongs of German," in *Speech Prosody 2004, International Conference*, 2004.

[17] C. Li, J. Al-Tamimi, and Y. Wu, "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 (pp. 1876-1880). Guarant International.*, 2023.

[18] S. Duanmu, *The phonology of standard Chinese*. OUP Oxford, 2007.

[19] G. Fant, *Acoustic theory of speech production*. Walter de Gruyter, 1970, no. 2.

[20] H. Bu, J. Du, X. Na, B. Wu, and H. Zheng, "AIShell-1: An Open-Source Mandarin Speech Corpus and A Speech Recognition Baseline," in *Oriental COCOSDA 2017*, 2017, p. Submitted.

[21] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

[22] J. Sun. (2020) Jieba: Chinese text segmentation [python package]. https://pypi.org/project/jieba/#files.

[23] C. Yang and Y. Xu, "Crosslinguistic trends in tone change: A review of tone change studies in East and Southeast Asia," *Diachronica*, vol. 36, no. 3, pp. 417–459, 2019.

[24] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot. Int.*, vol. 5, no. 9, pp. 341–345, 2001.

[25] J. Al-Tamimi and G. Khattab, "Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops," *Journal of Phonetics*, vol. 71, pp. 306–325, 2018.

[26] J. Al-Tamimi. (2022) Praat $f0$ Accurate Estimation. https://github.com/JalalAl-Tamimi/Praat-f0-Accurate-Estimation.

[27] R. Karlin, "Expanding the gestural model of lexical tone: Evidence from two dialects of serbian," *Laboratory Phonology*, vol. 13, no. 1, 2022.

[28] S. N. Wood, *Generalized additive models: an introduction with R*. chapman and hall/CRC, 2006.

[29] J. Van Rij, M. Wieling, R. H. Baayen, and D. van Rijn, "itsadug: Interpreting time series and autocorrelated data using GAMMs," 2015.

[30] R Core Team, *R: A Language and Environment for Statistical Computing*, https://www.R-project.org/, R Foundation for Statistical Computing, Vienna, Austria, 2021.

[31] 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.

[32] M. Sóskuthy, "Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction," *arXiv preprint arXiv:1703.05339*, 2017.

[33] M. Soskuthy, "Evaluating generalised additive mixed modelling strategies for dynamic speech analysis," *Journal of Phonetics*, vol. 84, p. 101017, 2021.

[34] C. Carignan, P. Hoole, E. Kunay, M. Pouplier, A. Joseph, D. Voit, J. Frahm, and J. Harrington, "Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 11, 2020.

[35] J. Al-Tamimi and P. Palo, "Dynamics of the tongue contour in the production of guttural consonants in Levantine Arabic," in *Radek Skarnitzl & Jan Volín (Eds.), Proceedings of the 20th International Congress of Phonetic Sciences (pp. 1876-1880). Guarant International.*, 2023.

[36] Y. Xu, "Contextual tonal variations in Mandarin," *Journal of phonetics*, vol. 25, no. 1, pp. 61–83, 1997.

[37] F. Wu and M. Kenstowicz, "Duration reflexes of syllable structure in Mandarin," *Lingua*, vol. 164, pp. 87–99, 2015.

[38] C. P. Browman and L. Goldstein, "Some notes on syllable structure in articulatory phonology," *Phonetica*, vol. 45, no. 2-4, pp. 140–155, 1988.

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

[40] M. S. Lundmark, J. Frid, G. Ambrazaitis, and S. Schötz, "Word-initial consonant–vowel coordination in a lexical pitch-accent language," *Phonetica*, vol. 78, no. 5-6, pp. 515–569, 2021.