Dynamic properties of rising diphthongs in Hefei Mandarin

Zhiheng Qian[1], Huifang Kong[2], and Yifan Yang[3†]

[1] School of Foreign Languages, Shanghai Jiao Tong University
n1vnhil@sjtu.edu.cn
[2] School of Foreign Languages, Anhui Jianzhu University
dorahf@163.com
[3] School of Foreign Languages, Shanghai Jiao Tong University
yifanyang@sjtu.edu.cn

Abstract. This paper investigates the dynamic properties of seven rising diphthongs in Hefei Mandarin among the younger generation, employing and comparing two methods of quantifying the entire formant trajectories. This study focuses on two questions to validate the findings reported in previous research. The first question examines how the formant trajectories of the rising diphthongs differ based on their positions in the carrier sentence. The second question explores whether the rising diphthongs are onset or offset-dominant. To address these questions, we employ generalized additive mixed modeling and the four-parameter logistic function. These methods have not been widely used in the study of diphthongs in Chinese dialects, and therefore, our paper also serves to demonstrate their utility in this area of research. The results show that [ja], [jɔ], [jɯ], and [wi] show significant differences with regard to their positions in the carrier sentence; both [jɯ] and [wa] exhibit an onset-dominant pattern in all or some contexts under investigation.

Keywords: vowel · diphthong · GAMM · parameters · Hefei Mandarin

1 Introduction

This paper probes into the dynamic properties of diphthongs in Hefei Mandarin among the younger generation, employing and comparing two methods of quantifying the entire formant trajectories, namely, generalized additive mixed models (GAMM) [1] and the four-parameter logistic function (also known as the Boltzmann function) [2]. The results obtained in this preliminary report will provide a basis for further investigations into the articulatory properties of Hefei diphthongs.

Hefei Mandarin is a variety of Jianghuai Mandarin, spoken in Hefei, the capital of Anhui Province. Hefei Mandarin has a large vowel inventory, but the

† Corresponding author. The authors equally contributed to the paper.

exact vowels in the inventory and their properties remain controversial in the literature, particularly the diphthongs. Diphthongs can be defined as vowels that involve changes in quality within the same syllable [3]. Though traditional descriptive works list *finals* as a single unit, we can identify 14 diphthongs in Li [4] ([ja, wa, jɛ, wɛ, we, ɥe, jɔ, jɯ, jã, wã, wæ̃, jĩ, ɥĩ]), all of which are rising diphthongs. In contrast, Kong et al. [5] analyze 9 diphthongs ([eɪ, jɯ, jɔ, wi, wɛ, wa, jã, wæ̃, wã]), excluding [ja] and [jɛ] while including an additional falling diphthong [eɪ]. Considering the environment of distribution, the prenuclear glide in [ja] and [jɛ] is arguably the formant transition of the nuclear vowel [5]. Overall, the inventory and properties of Hefei diphthongs remain contentious, especially the status of [ja] and [jɛ].

The acoustic study of Hefei diphthongs is generally limited, despite various studies on the properties of diphthongs in other Chinese dialects, including Ningbo [6], Hakka [7], Suzhou [8], Taiyuan [9], Hangzhou [10], Xiangxiang [11], Chengdu [12], among others. In addition to Kong et al. [5], Wan [13] provides a detailed discussion and analysis on Hefei diphthongs, focusing on [ja, wa, jɛ, wɛ, we, ɥe, jɔ, jɯ], the non-nasalized series in Li [4]. Similar to prior research on diphthongs in other Chinese dialects [6–12], Wan analyzes Hefei diphthongs by examining pre-selected discrete time points of formants, and investigates the temporal structure of Hefei diphthongs. Wan concludes that the formants of the diphthongs do not have a significant difference with regard to their positions in the carrier sentence. The onglide in Hefei diphthongs is short, and all rising diphthongs are offset dominant, namely, the sonorant part exhibits greater stability [13].

This paper probes into the dynamic properties of [ja, jɛ, jɔ, jɯ, wa, wɛ, wi] in Hefei Mandarin among the younger generation. Building on the findings of previous research, our study focuses on two questions. The first question examines how the formant trajectories of the rising diphthongs differ based on their positions in the carrier sentence, aiming to validate Wan's [13] previous findings. The second question explores whether the rising diphthongs are onset or offset-dominant. Although Wan's [13] study suggests that all rising diphthongs in Hefei are offset dominant and have short onglides with rapid transitions, our perception indicates that at least [jɯ] has a clear and stable onglide. To address these questions, we employ GAMM [1] and the four-parameter logistic function, each offering unique benefits in quantifying dynamic linguistic data [2, 14–17]. These methods have not been widely used in the study of diphthongs in Chinese dialects, and therefore, our paper also serves to demonstrate their utility in this area of research. For the rest of the paper, §2 introduces the method; §3 presents the results; §4 discusses the findings and concludes the paper.

2 Methods

2.1 Speakers

Ten native speakers of Hefei Mandarin were recruited in the study (five males and five females), aged between 18 and 22 at the time of the experiment. The speakers are college students and they were all born and raised in Hefei. All of them speak Hefei Mandarin at home with their families, but they use Standard Chinese at school.

2.2 Materials and recording procedure

The materials were prepared by the authors, one of whom is a native speaker of Hefei Mandarin. The syllables that are analyzed in this paper include [ja^{53}, jɛ^{24}, jɔ^{53}, jɯ^{53}, wa^{53}, wɛ^{53}, wi^{53}] (the superscript digits indicate tones). The tones are well controlled except for [jɛ^{24}] due to an accidental gap. It is noteworthy that we transcribe [we] in [13] as [wi] in our study, following [5]. In addition, the diphthong [ɥe] is excluded since it shows variations between [ɥe] and [y] among the younger speakers, as also reported in previous studies [13], [5]. We did not find the falling diphthong [eɪ] described in Kong et al. [5].

Each word is embedded in a carrier sentence [X, tuəʔ^{5} X tsɿ^{21}] ("X, read X"), where the first token appears in an isolated position (labeled as "A position"), while the second one is embedded in the phrase (labeled as "B position"). Each sentence is repeated 10 times in a randomized order. The data were acquired in a quiet room in Hefei. The recordings were made onto a laptop using a Zoom H6 recorder and Praat software [18]. The sampling rate used was 44,100 Hz.

2.3 Data processing and acoustic analyses

The target diphthongs in each carrier sentence were extracted from the audio files and subsequently annotated by trained research assistants. The diphthongs were labeled from their onset to offset of the formants, with reference to both the spectrogram and waveform. The lowest three formants were extracted and time-normalized automatically using the script *FormantPro* [19] in Praat. The formants were further normalized using the Lobanov method [20]. A total of 10 sentences were excluded due to disfluencies in production or errors in pronunciation, yielding 690 sentences for data analysis.

In line with previous studies [2, 21–23], the dynamic aspects of diphthongs are expressed through the range and rate of spectral change for the second formant (F2). The first formant (F1) is not considered significant in reflecting the transitions between vowels, and the third formant (F3) has minimal influence on vowel recognition. Thus, F2 is utilized as the primary measurement to characterize the diphthongs in this study. In particular, the 10% - 90% portion of each F2 trajectory was used in the analyses.

2.4 Statistics

Two statistical models were employed. First, GAMMs were constructed following the procedure in [26]. The application of GAMMs in phonetic studies has been increasingly popular recently due to their capabilities to fit non-linear models and better interpret data through fixed and random effects. In this study, the models for [jV] ([ja, jɛ, jɔ, jɯ]) and [wV] ([wa, wɛ, wi]) series were constructed separately, using the *bam* function in the *mgcv* package [24] in R [25]. In each model, the dependent variable was normalized F2, and the parametric factor was the interaction between sentence positions and diphthongs. The fixed effect included a smooth over time (trajectories) for each level of the interacting variable (sentence positions and diphthongs). The random effects included by-speaker variations in F2 trajectories. To formally test the difference between positions for each diphthong, we further re-specified the models using ordered factors. An ordered factor was created for each diphthong for positions A and B.

Second, the fitted smooth of each diphthong at A and B positions that we obtained in the modeling above was further quantified by the four-parameter logistic function (Boltzmann function) in (1) [2], using the *nls* function in R.

$$f(x) = \frac{A_1 - A_2}{1 + e^{\frac{x - x_0}{dx}}} + A_2 \tag{1}$$

Jiao [2] uses this function to quantify formant trajectories and discusses the linguistic implications of each of the four parameters. A_1 and A_2 are the lower and higher asymptotes respectively, which are defined as the ideal targets of a diphthong. dx relates to the rate of change. x_0 is the inflection point, relating to the displacement on x-axis, and it is viewed as an indicator of the proportion of the onset/offset of a diphthong. Jiao [2] argues that an offset-dominant diphthong has a x_0 value greater than 0.5, because it demonstrates leftward displacement on x-axis, thereby reducing the portion of the onset.

3 Results

3.1 F2 trajectories at different positions

The model for [jV] series explained 84.1% of the variance in F2, while the one for [wV] series explained 87.5% of the variance in F2. The by-speaker random effect was significant. The smooths obtained for the diphthongs are visualized in Figure 1. The smooths illustrate the patterns of F2 trajectories of the diphthongs at different positions. From visual inspection, for each diphthong, the starting and ending points at positions A and B are similar.

To assess whether these diphthongs statistically differ in F2 patterns at different positions, the summary of the re-specified models with ordered factors is presented in Table 1. The results indicate that the diphthongs, except for [jɛ], [wa], and [wɛ], show significant differences regarding the F2 trajectories in different positions. To better visualize the results, the difference smooth for each diphthong, obtained from the models with ordered factors, is illustrated in Figure 2. For [jɛ], [wa], and [wɛ], the differences between trajectories at positions A and B are not significantly different from zero. This is indicated by the inclusion of the horizontal line at 0 within the 95% confidence interval (the shaded bands).

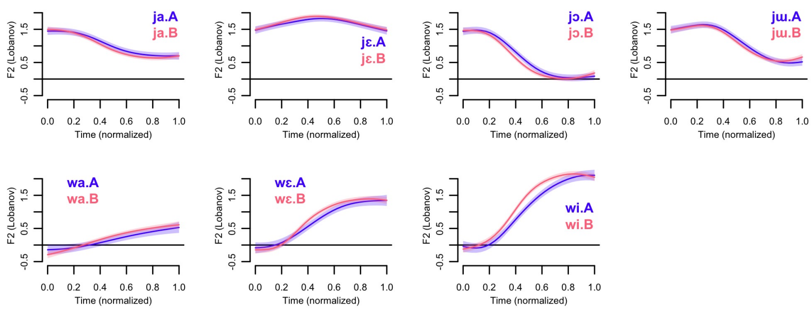

Fig. 1. GAMM fitted smooths with 95% confidence intervals illustrating normalized F2 values as a function of normalized time. The random effect term is excluded.

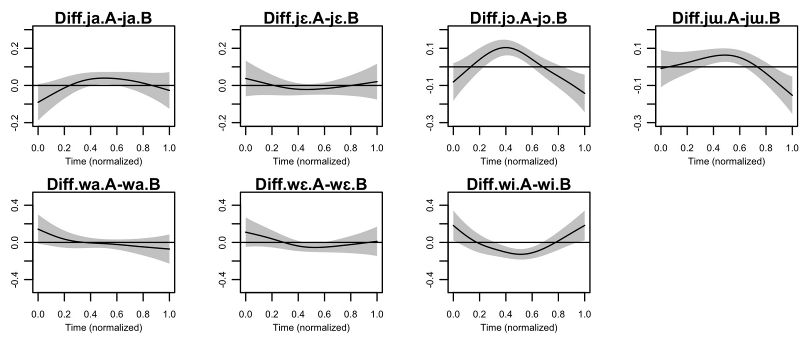

Fig. 2. Difference between trajectories at positions A and B for each diphthong. The 95% confidence intervals are indicated by shaded bands.

Table 1. Summary of smooth functions.

	edf	F	p-value		edf	F	p-value
s(Time):				s(Time):			
ja.A.Ordered	2.702	2.603	< 0.05	wa.A.Ordered	2.716	1.549	0.27
jɛ.A.Ordered	2.119	1.201	0.38	wɛ.A.Ordered	2.703	1.589	0.18
jɔ.A.Ordered	3.836	6.895	< 0.001	wi.A.Ordered	3.930	6.413	< 0.001
jɯ.A.Ordered	3.183	4.769	< 0.01				

3.2 Parameters of F2 trajectories

By quantifying the F2 trajectories using the four-parameter logistic function, the parameters A_1, A_2, x_0, and dx are provided in Table 2. The goodness-of-fits are illustrated in Figure 3. To save space, we only present the diphthong models at A position. For each diphthong, the red dots represent the fitted smooth of F2 obtained from the GAMM, and the dashed lines indicate the upper and lower boundaries of the 95% confidence interval for each trajectory. The blue dashed lines indicate the logistic curve of the trajectory. All the logistic curves, except for [jɛ], fall within the confidence interval, indicating a good predictive ability. Details will be discussed in §4.

Table 2. Parameters of logistic functions.

		ja	jɛ	jɔ	jɯ	wa	wɛ	wi
A position	A_2	1.473	1.707	1.502	1.592	-0.222	-0.164	-0.202
	A_1	0.691	1.456	0.039	0.491	0.613	1.371	2.134
	x_0	0.438	**0.866**	0.415	**0.569**	**0.541**	0.417	0.453
	d_x	0.106	0.045	0.092	0.080	0.223	0.132	0.128
B position	A_2	1.520	1.756	1.497	1.564	-0.622	-0.251	-0.165
	A_1	0.648	1.490	0.058	0.581	0.692	1.384	2.135
	x_0	0.381	**0.856**	0.362	**0.519**	0.296	0.363	0.384
	d_x	0.103	0.046	0.083	0.068	0.267	0.104	0.102

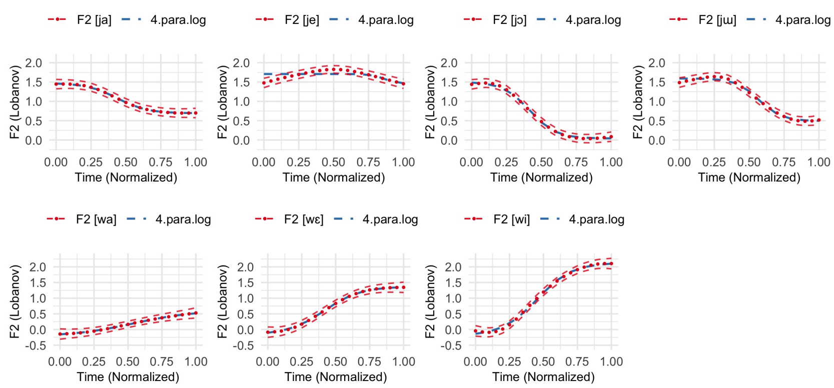

Fig. 3. Logistic models and F2 trajectories (dashed lines: 95% confidence interval)

4 Discussion and conclusion

By quantifying the entire F2 trajectories, we investigate the dynamic properties of Hefei diphthongs from novel perspectives. Regarding the first question, contrary to the conclusion in Wan's [13] study, the GAMMs indicate that [ja], [jɔ], [jɯ], and [wi] show significant differences with regard to their positions in the carrier sentence. As shown in Figure 2, the differences mainly exhibit in the transitional part of the rising diphthongs.

Regarding the second question, though previous studies suggest that all rising diphthongs in Hefei are offset dominant [13], we have made different observations. Again, in Figure 1, [jɯ] has a clear and stable onset portion, which aligns with our perception in the fieldwork. Though GAMMs offer greater advantages in formally assessing non-linear differences, the internal structure can be further quantified using the four-parameter logistic function. By following the criteria defined in [2], values of x_0 greater than 0.5 are in bold in Table 2, indicating a greater proportion of the onset part. Therefore, [jɯ] consistently shows onset-dominant patterns at both positions in the carrier sentence, with $x_0 > 0.5$, while [wa] has a parameter of $x_0 > 0.5$ only at the A position. Though [jɛ] also has an x_0 greater than 0.5, the logistic model does not accurately capture the F2 trajectory, according to Figure 3. It is worth noting that Jiao's [2] study briefly discusses the dynamic properties of Hefei diphthongs, but it is conducted on a small dataset collected from one single speaker (only one token for each diphthong). Instead, the current study provides a more informative analysis with a larger dataset.

Through these statistical models, we have observed certain dynamic properties of the formant trajectories that cannot be obtained by formant measurements at specific time points. When comparing the results of the GAMMs and the logistic models, we find that the diphthongs exhibiting significant difference at both positions in the carrier sentence tend to have a larger onset-offset difference (i.e., greater $|A_2 - A_1|$ in Table 2). Further, the results show an interesting asymmetry that most [jV] diphthongs show variations by position (3 out of 4), while most [wV] diphthongs do not show such variations (2 out of 3). This leads to speculation that [j] and [w] may differ in their intrinsic properties. Since it has been argued that the labial gesture is consonantal in English (a "C-gesture"), which is different from the tongue back gesture ([j]) [27], we speculate that [w] in Hefei could also be more consonantal than [j]. Thus, [wV] is in a consonant-vowel coupling relation, which is more stable in terms of gestural coordination [28] and may exhibit less variation in different contexts.

Nevertheless, as a short preliminary report, there are still details that need to be explored in future work. The raw duration of the diphthongs will be considered, and the onset and offset targets will be compared with relevant monophthongs, and the hypothesis regarding the intrinsic properties of [j] and [w] also needs careful examination. Additionally, the special pattern of [jɛ] warrants further investigation. These additional investigations will help to further demonstrate the utility and predictive capabilities of the statistical models employed in this paper.

Acknowledgements We thank Prof. Shuiyuan Yu and three reviewers for providing valuable suggestions. We also thank Yunjia Lv, Qianyi Yang, and Tianjian Zhu for segmenting and annotating the diphthong data. All errors are our own.

References

1. Wood, S.: Generalized additive models: An introduction with R. 1st edn. CRC Press, Boca Raton (2006).
2. Jiao, L.: A parametric study of diphthongs in Chinese dialects. Doctoral dissertation. The Hong Kong University of Science and Technology (2015).
3. Ladefoged, P., Johnson, K.: A course in phonetics. 6th edn. Wadsworth/Cengage Learning, Boston, MA (2011).
4. Li, J.: Hefeihua yindang [The phonetic files of Hefei dialect]. Shanghai Educational Press, Shanghai (1997).

5. Kong, H., Wu, S., Li, M.: Hefei Mandarin. Journal of the International Phonetic Association 53(3), 1145–1166 (2023).
6. Hu, F.: An acoustic phonetic analysis of diphthongs in ningbo Chinese. INTERSPEECH (2003).
7. Man, C.: An acoustical analysis of the vowels, diphthongs and triphthongs in Hakka Chinese. ICPhS (2007).
8. Ling, F.: The phonetic boundary between monophthongs and diphthongs in Suzhou Chinese. In: Clarke, S., Hirayama, M., Kim, K., Suh, E. Toronto Working Papers in Linguistics 28 (2008).
9. Xia, L., Hu, F.: Vowels and Diphthongs in the Taiyuan Jin Chinese Dialect. INTERSPEECH (2016).
10. Yang, Y., Hu, F.: Vowels and Diphthongs in Hangzhou Wu Chinese Dialect. INTERSPEECH (2018).
11. Zeng, T.: The Xiangxiang dialect of Chinese. Journal of the International Phonetic Association 50(2), 258-281 (2020).
12. Li, N., Hu, F.: Vowels and Diphthongs in Chengdu Mandarin. ICPhS (2023).
13. Wan, L.: Experimental research on the vowel system of Hefei dialect among the younger generation. MA thesis, Nanjing Normal University, Nanjing (2014).
14. Warren, P.: 2019. Non-linear analysis of a diphthong merger. ICPhS (2019).
15. Renwick, M., Stanley, J.: Modeling dynamic trajectories of front vowels in the American South. The Journal of the Acoustical Society of America 147(1), 579–595 (2020).
16. Kong, C., Long, X., Liu, J.: An acoustic study on age-related changes in vowel production of Chinese. Journal of Phonetics (104) 101324. (2024)
17. Steffman, J.: Vowel-initial glottalization as a prominence cue in speech perception and online processing. Laboratory Phonology 14(1) (2023).
18. Boersma, P., Weenink, D.: Praat, a system for doing phonetics by computer. (2024).
19. Xu, Y., and Gao, H.: FormantPro as a tool for speech analysis and segmentation. Revista de Estudos da Linguagem 26(4), 1435-1454 (2018).
20. Lobanov, B.: Classification of Russian vowels spoken by different speakers. The Journal of the Acoustical Society of America 49(2B), 606-608 (1971).
21. Gay, T.: Effect of speaking rate on diphthong formant movements. The Journal of the Acoustical Society of America 44(6), 1570-1573 (1968).
22. Gay, T.: A perceptual study of American English diphthongs. Language and speech 13(2), 65-88 (1970).
23. Ren, H.: On the acoustic structure of diphthongal syllables. Doctoral Dissertation, UCLA (1986).
24. Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1), 3–36 (2011).
25. R Core Team.: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2021).
26. Wieling, M.: 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 (2018).
27. Gick, B.: The articulatory basis of syllable structure: a study of English glides and liquids. Doctoral Dissertation. Yale University (1999).
28. Goldstein, L., Byrd, D., Saltzman, E. The role of vocal tract gestural action units in understanding the evolution of phonology. Action to language via the mirror neuron system. 215–49 (2006).