

VOWELS AND DIPHTHONGS IN THE CHANGDE MANDARIN CHINESE

Zhenrui Zhang¹, Fang Hu^{2,1}

¹Department of Linguistics, Graduate School, University of Chinese Academy of Social Sciences

²Institute of Linguistics, Chinese Academy of Social Sciences

1661829094@qq.com, hufang@cass.org.cn

ABSTRACT

This is an acoustic phonetic analysis of vowels and diphthongs in the Changde Mandarin Chinese dialect. 10 speakers, 5 male and 5 female, provided speech data. The first three formants for each vowel and diphthong element were measured and spectral dynamic data were calculated for each diphthong. Results suggest that monophthongs and falling diphthongs are single-event articulations, whereas rising diphthongs are sequences of two articulation events.

Keywords: vowel (monophthong), falling diphthong, rising diphthong, the Changde Mandarin dialect.

1. INTRODUCTION

Syllables are straightforward in Chinese dialects, as each syllable is a separate written unit. And Chinese syllables have a simple structure, which could basically be represented as CGVC (initial consonant, on-glide, vowel nucleus, and coda), but there is controversy regarding the hierarchy of structure ([1], [2]). The essential issue concerns the phonetics and phonology of vowels and diphthongs. There is a long debate on the definition of diphthongs in the literature. Some linguists view a diphthong as a single vowel with a phonetically complex nucleus ([3], [4], [5]), while others treat a diphthong as a sequence of two vowels or a combination of one vowel and one semivowel ([6], [7]). Even for English, for instance, there is no consensus regarding what a diphthong is ([8], [9], [10]). And various solutions were proposed for structuring Chinese syllables ([11], [12], [13], [14], [15], [16], [1]).

Recent researches from Chinese dialects renewed the issue of diphthongs. Accumulative evidence from acoustic as well as lingual kinematic data showed that falling diphthongs in Chinese dialects are composed of a dynamic target, while rising diphthongs are composed of two targets ([17], [18], [19], [20], [21]). For instance, [ai] is a single dynamic articulatory event, while [ia] is a sequence of [i] and [a].

Changde is located in northwest Hunan. It's basically an area of the Xiang dialects family; but the Changde dialect belongs to the southwestern Mandarin dialects family ([22]). Previous studies on

the Changde dialect are mainly dialectological works ([23]) and little has been done on the basis of acoustic phonetic measurements. As a Mandarin dialect, Changde is typical in that it has a limited number of monophthongs and a rich inventory of diphthongs and triphthongs. There are 8 monophthongs [ɿ i y e a o u u], 4 falling diphthongs [ai ei au əu], 8 rising diphthongs [ia ie ua ue uo ya ye yo], and 6 triphthongs [iau iəu uai uei yai yei] in open syllables. There are less vowel distinctions in nasal-ending syllables: 5 monophthongs [a e i u y ə], and 4 rising diphthongs [ia ua ya yo].

This paper is an acoustic phonetic description of Changde vowels. On the basis of fine-grained phonetic details, it aims to give an acoustic-evidence-based vowel phonology of the Changde dialect in particular and to give a better understanding of syllable complexity in Mandarin Chinese dialects in general. Due to the space limit, this paper focuses on vowels and diphthongs in open syllables.

2. METHODOLOGY

10 native speakers, 5 males and 5 females, provided speech data during the fieldwork trip in 2018. None of them reported any speech or hearing disorders.

Meaningful monosyllabic words containing the target vowels in all environments were used as test words. Target words are onset-less syllables or syllables with a bilabial initial stop [p]. Each test word was placed in a carrier sentence [X, təu13 X kə21 ni21 th̃iŋ55] “X, read X for you to listen”. The speakers were instructed to read the randomized word list naturally in a normal speech rate.

The 16-bit audio sounds were recorded through an AKG microphone directly into a laptop PC with a Lexicon I-O 22 sound card. The sampling rate is 11,025 Hz. Five repetitions were recorded.

The target vowels were labelled in praat 6.0.29 ([24]). The diphthongs were annotated as being composed of an onset, an offset, and the transition connecting them. The lowest four formant frequencies were extracted from the midpoint of the steady state for the onset or offset segment. If there is no steady state, first formant (F1) minimum was taken as the target point for high vowels and F1 maximum was taken as the target point for low vowels. The duration of each segment was measured. And range

and rate of the second formant (F2) change were further calculated for each diphthong.

3. RESULTS

3.1. Monophthongs

Figure 1 shows the 8 Changde monophthongs [ɿ i y ɛ a ɔ u u] in the acoustic vowel plane. The acoustic vowel space is determined by using the first formant (F1) as ordinate and second formant (F2) as abscissa with the origin of the coordinate to the top right. Coordinates are bark-scaled ([25]), but still labelled in Hertz. Each two-sigma ellipse, which intuitively manifests the vowel distribution, is based on 30 sampled data points.

Figure 1: 2-sigma ellipses for the Changde monophthongs in male (left) and female speakers (right).

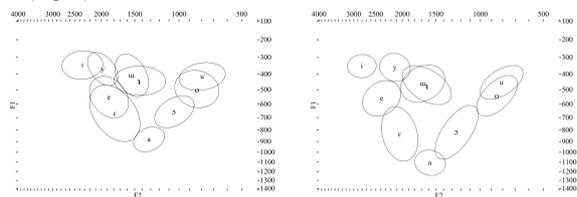


Table 1: Means and one-way ANOVAs for F1, F2, and F3 (in Hz) of [ɿ u].

	Male		Female	
	ɿ	u	ɿ	u
F1	422	391	441	453
	F=11.263; P=0.887		F=2.627; P=0.673	
F2	1428	1534	1637	1685
	F=10.996; P=0.0015		F=26.505; P=0.11	
F3	2913	2242	3376	2947
	F=79.592; P<.0001		F=83.025; P<.0001	

Changde monophthongs show a triangular distribution in both male (left) and female (right) speakers. There are 3 levels of height and 3 levels of backness: [ɿ y ɿ u u] are high vowels, [e o] are mid vowels, and the low vowel [a] does not distinguish in backness; [ɿ y ɛ] are front vowels, [ɿ a] are central vowels, [o u u] are back vowels. First, mid vowels [e o] are diphthongized. As is denoted in the figure, [e o] are actually [eɛ oɔ] respectively. Second, except for the high vowels, Changde monophthongs have a predictable relationship between backness and lip rounding: front vowels are unrounded and back vowels are rounded. Both front and back high vowels distinguish in lip rounding: [ɿ u] vs [y u] respectively. Third, the apical vowel [ɿ] occupies a central high position in the acoustic F1/F2 vowel space. This is an acoustic characteristic of apical vowels that is commonly observed in Chinese dialects ([26]). Fourth, the ellipses for the apical vowel [ɿ] and unrounded back vowel [u] heavily overlap with each

other. As shown in Table 1, they can nevertheless be differentiated in F3, although having similar F1 and F2.

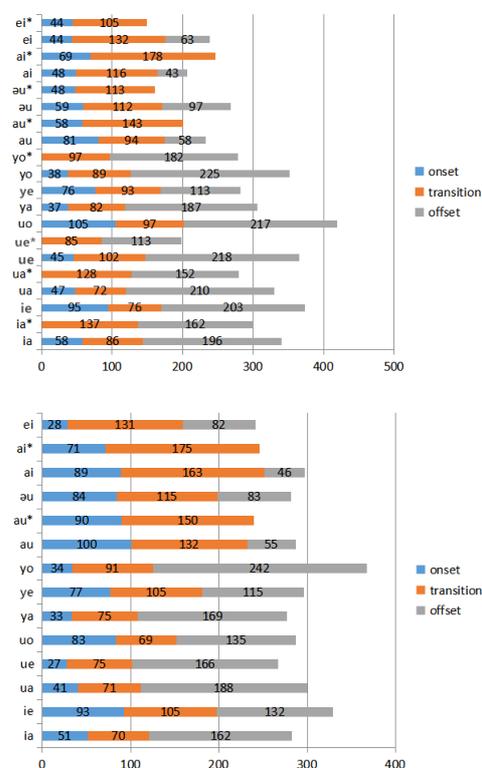
3.2. Diphthongs

As compared to monophthongs, there are less distinctions in diphthongs. The 4 falling diphthongs have a 4-way oppositions, i.e. 2-way oppositions of height in both front and back series: [ai ei] and [au əu]. The rising diphthongs have a 2-way opposition for the i-series [a e], and a 3-way opposition for the u-series and y-series [a e o]. Acoustic properties of diphthongs were characterized below in terms of temporal organization, spectral properties, and dynamic aspects.

3.2.1. Temporal organization:

Mean durations in milliseconds and in percentages of diphthong components, namely onset, transition, and offset, for the 12 diphthongs were plotted in bar charts in Figure 2 and 3 respectively.

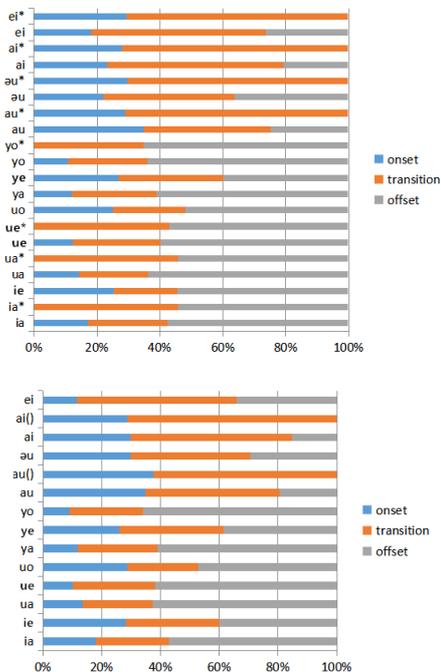
Figure 2: Mean durations in milliseconds of diphthong components in male (upper) and female speakers (lower).



First of all, not all diphthongs have three components. About 15% of falling diphthongs have no steady state for the offset in male speakers, and about 22% in female speakers. About 5% of rising diphthongs have no steady state for the onset but only

in male speakers. Furthermore, some diphthongs only have a transition. That is, the formants are always changing. The temporal organization for these diphthongs were denoted by an asterisk in the figures.

Figure 3: Mean durations in percentages of diphthong components in male (upper) and female (lower) speakers.



Second, falling and rising diphthongs exhibit different patterns of temporal organization in general. The transition is the longest component, taking about 40%-60% of total duration in falling diphthongs, while the offset tends to be the longest component, taking about 40%-60% of total duration in rising diphthongs. The rising diphthongs in Changde show a pattern of temporal organization similar to other Chinese dialects, whereas the falling diphthongs are quite different to those in other dialects such as Ningbo Wu ([17]), Longchang Mandarin ([18]), Taiyuan Jin ([19]), Cangnan southern Min ([20]), and Hui dialects ([27], [28], [29]). In these dialects, falling diphthongs have a balanced structure in temporal organization.

3.2.2. Spectral properties

Figures 4-6 compare formant patterns of the onset and offset elements of rising diphthongs with their monophthongal counterparts in the acoustic F1/F2 vowel plane. Diphthong onsets were denoted by offsets in parentheses, and vice versa.

It can be seen from the figures that both onset and offset elements in rising diphthongs have comparable distributions to their monophthongal counterparts in the acoustic vowel plane. In general, ellipses for

diphthong onsets [i u y] and offsets [a ε o] and the corresponding monophthongs extensively overlap with each other. The only exception is [u] in [ue], which exhibits a clear effect of anticipatory coarticulation so that the ellipse for [u] shifts inward to a less peripheral position in the acoustic vowel plane in both male and female speakers. However it is still a high back position and as manifested by the size of ellipse, the variabilities for diphthong element [u] in [ue] and the monophthong [u] are comparable. In summary, spectral data suggest that both onset and offset elements in rising diphthongs are well controlled as their monophthongal counterparts. In other words, rising diphthongs have two specific spectral targets.

Figure 4: 2-sigma ellipses for the onset and offset of [ia ie] and corresponding monophthongs in male (left) and female (right) speakers.

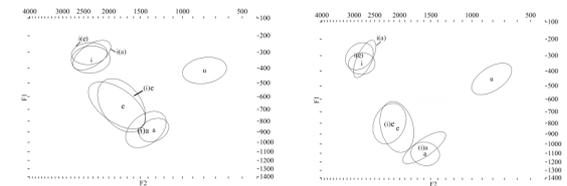


Figure 5: 2-sigma ellipses for the onset and offset of [ua ue uo] and corresponding monophthongs in male (left) and female (right) speakers.

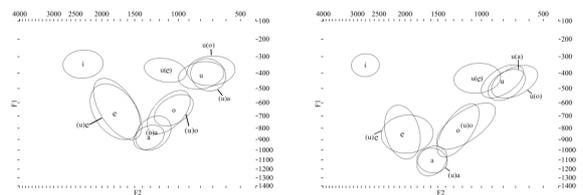
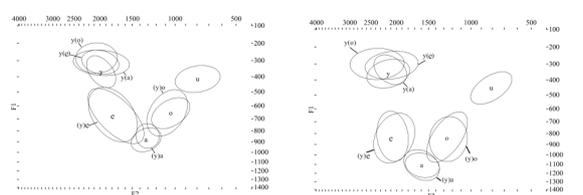


Figure 6: 2-sigma ellipses for the onset and offset of [ya ye yo] and corresponding monophthongs in male (left) and female (right) speakers.



Figures 7-8 show the ellipses for diphthong components of falling diphthongs and their monophthongal counterparts. Different to in rising diphthongs, only onsets in falling diphthongs have comparable distributions to their monophthongal counterparts, whereas the offsets in falling diphthongs are highly variable. This is well illustrated by [ai au] in Figure 7. [a]s in [ai au] are comparable

to their monophthongal counterpart, although there is certain effect of anticipatory coarticulation so that [a] in [ai] shifts to an anterior position and [a] in [au] shifts to a posterior position. But the offsets, [i] in [ai] and [u] in [au] only reach a mid-high, rather than a high position in the acoustic vowel plane. It seems that falling diphthongs do not have two spectral targets as rising diphthongs do. Rather, falling diphthongs have a dynamic spectral target.

The other two falling diphthongs [ei əu] have spectral movements shorter than [ai au] in the acoustic vowel plane. [ei əu] and [ai au] form a high versus low distinction from a dynamic perspective of diphthong production ([24]). Moreover, as mentioned earlier, the monophthong [e o] is diphthongized [eɛ oɔ] respectively. That is, [ei əu] are diphthongs with opposite directions of spectral movement to [eɛ oɔ] respectively.

Figure 7: 2-sigma ellipses for onsets and offsets of [ai ei] and corresponding monophthongs in male (left) and female (right) speakers.

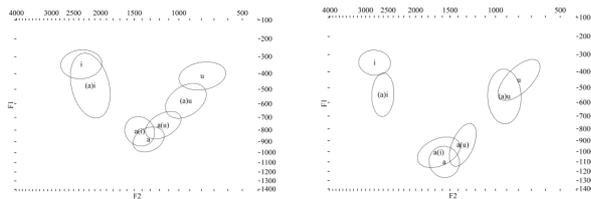
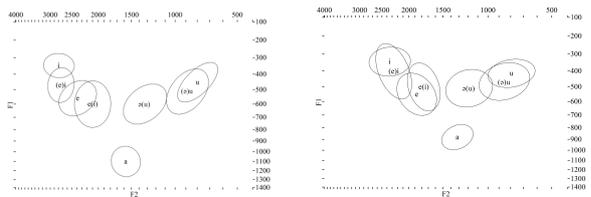


Figure 8: 2-sigma ellipses for onsets and offsets of [au əu] and corresponding monophthongs in male (left) and female (right) speakers.



3.2.3. Spectral dynamics:

The range and rate of second formant (F2) change capture certain dynamic aspects of formant structure in diphthong production ([31], [32]). However, F2 range and rate of change are viable for the characterization of diphthongs in languages with a simple inventory of diphthongs, but not in languages with a complex inventory of diphthongs ([33], [34], [35]). And data from Chinese dialects suggest that falling and rising diphthongs should be taken into account separately ([26]).

Figure 3 shows means of ranges in Hz and rates in Hz/ms of F2 change for the diphthongs in Changde dialect. First, there is a clear pattern for the falling

diphthongs in both male and female speakers: [ai] > [ei] > [au] > [əu] both in terms of F2 range and rate of change. Second, rising diphthongs have a much greater range and rate of F2 change than their falling counterparts in both male and female speakers, e.g. [ia] > [ai], [ie] > [ei], [ua] > [au], and [ue] or even [uo] > [əu]. Third, there is however no consistent pattern of F2 range or rate of change among rising diphthongs.

Table 2: Mean ranges in Hz and rates in Hz/ms of the F2 change for the diphthongs in Changde.

Diphthongs		Male		Female	
		ΔF2	rate	ΔF2	rate
Falling	au	218	1.85	339	2.41
	əu	156	1.38	197	1.71
	ai	751	5.11	954	5.36
	ei	400	3.42	748	4.99
Rising	ie	512	6.74	666	6.34
	ia	900	10.47	1079	15.41
	ue	611.37	5.99	925.09	12.33
	ya	394.41	4.81	523.42	6.98
	ua	532.31	7.39	715.82	10.08
	uo	380.81	3.93	457.44	6.63
	ye	253.39	2.72	271.77	2.59
	yo	978.87	11	1117.83	12.28

4. CONCLUSION

This paper gives an acoustic phonetic description of the vowels and diphthongs in the Changde dialect. And vowel phonology was discussed on the basis of fine-grained phonetic details.

First, Changde monophthongs have a 3-way distinction in height and backness. And the mid high vowels [e o] are diphthongized [eɛ oɔ] respectively. Second, falling diphthongs have a dynamic spectral target, while rising diphthongs have two spectral targets. Third, the transition is the longest component in a falling diphthong, while the offset is the longest component in a rising diphthong. In summary, falling diphthongs and monophthongs are single-event articulations, while rising diphthongs are sequences of two articulation events. And falling diphthongs are characterized by a much smaller range and rate of F2 change than rising diphthongs. Last, but not least, it should be admitted that more data from languages are needed to further explore what is universal and what is language-specific in vowel production.

5. REFERENCE

- [1] Duanmu, S. 2008. *Syllable Structure: The Limits of Variation*. Oxford University Press.
- [2] Duanmu, S. 2016. Syllable and syllable structure in Chinese. In R. Sybesma (editor-in-chief), W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers, (eds.) *Encyclopedia of Chinese Language and Linguistics*. Leiden: Brill.
- [3] Malmberg, B. 1963. *Structural linguistics and human communication*, Berlin: Springer-Verlag.
- [4] Abercrombie, D. 1967. *Elements of general phonetics*, Edinburgh: Edinburgh University Press.
- [5] Catford, I. 1977. *Fundamental problems in phonetics*, Edinburgh: Edinburgh University Press.
- [6] Sweet, H. 1877. *A handbook of phonetics including a popular exposition of the principles of spelling reform*. Oxford: Clarendon Press.
- [7] Jones, D. 1922. *Outline of English phonetics (2nd Ed.)*. New York: E. P. Dutton.
- [8] Pike, K. L. 1947. On the phonemic status of English diphthongs. *Language* 23, 151-159.
- [9] Lehiste, I. & Peterson, G. E. 1961. Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America* 33, 268-277.
- [10] Holbrook, A. and Fairbanks, G. 1962. Diphthong formants and their movements. *Journal of Speech and Hearing Research* 5, 38-58.
- [11] Hartman, L. M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20, 28-42.
- [12] Luo, C. and Wang, J. 1957. *An Outline of General Phonetics* [in Chinese]. Beijing: China Science Press.
- [13] Cheng, R. L. 1966. Mandarin phonological structure. *Journal of Linguistics* 2, 135-158.
- [14] Cheng, C. C. 1973. *A synchronic phonology of Mandarin Chinese. Monograph on Linguistic Analysis*, No. 4, The Hague: Mouton.
- [15] Chao, Y.-R. 1968. *A grammar of spoken Chinese*, Berkeley: University of California Press.
- [16] You, R., Qian, N. and Gao, Z. 1980. On the phonemic system of Putonghua [in Chinese]. *Chinese Language* 5, 328-334.
- [17] Hu, F. 2013. Falling diphthongs have two targets: Acoustics and Articulation of the diphthong production in Ningbo Chinese. *Journal of the Acoustical Society of America* 134(5), Pt. 2, 4199; full paper in Chinese in *Bulletin of Linguistic Studies* 10, 12-37. Shanghai: Shanghai Lexicography Press.
- [18] Qiu, Y. & Hu F. 2013. Longchang vowels [in Chinese]. *Bulletin of Linguistic Studies* 10, 38-51. Shanghai: Shanghai Lexicography Press.
- [19] Xia, L. and Hu, F. 2016. Vowels and diphthongs in the Taiyuan Jin Chinese dialect. *Proceedings of Interspeech 2016*, San Francisco, 993-997.
- [20] Hu, F. and Ge, C. 2016. Vowels and diphthongs in Cangnan Southern Min Chinese dialect. *Proceedings of Interspeech 2016*, San Francisco, 978-982.
- [21] Hu, F. and Yue, Y. 2018. Vowels and diphthongs in Hangzhou Wu Chinese dialect. *Proceedings of Interspeech 2018*, Hyderabad, 207-211.
- [22] Chinese Academy of Social Sciences and City University of Hong Kong 2012. *Language Atlas of China (2nd edition): Chinese dialect volume*. Beijing: The Commercial Press.
- [23] Zheng, Q. 1999. *Studies on the Changde Dialect*. Changsha: Hunan Educational Press.
- [24] Boersma, P. & Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.29.
- [25] Smith, J. O. and Abel, J. S. 1999. Bark and ERB Bilinear Transforms. *IEEE Trans. Speech and Audio Proc.*, 7(6): 697-708.
- [26] Hu, F. expected 2019. *Vowels* [in Chinese]. Beijing: Foreign Language Teaching and Research Press.
- [27] Hu, F. & Zhang, M. 2014. Diphthongized Vowels in the Yi County Hui Chinese Dialect. *Proceedings of Interspeech 2014*, Singapore, 1703-1707.
- [28] Hu, F. & Zhang, M. 2015. On the Diphthongized vowels in Qimen Hui Chinese. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, Glasgow, Scotland.
- [29] Zhang, M. & Hu, F. 2017. Diphthongized vowels in the Xiuning Hui Chinese dialect. *Proceedings of O-Cocosda*, Seoul, Korea, 133-138.
- [30] Hu, F. 2017. Towards a dynamic theory of vowel production. *Journal of the Acoustical Society of America* 142 (4), Pt.2, 2579.
- [31] Gay, T. 1968. Effects of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America* 44, 1570-1573.
- [32] Gay, T. 1970. A perceptual study of American English diphthongs. *Language and Speech* 13, 65-88.
- [33] Manrique, A. 1979. Acoustic analysis of the Spanish diphthongs. *Phonetica* 36, 194-206.
- [34] Ren, H. 1986. *On the Acoustic Structure of Diphthongal Syllables*. Ph. D. dissertation, UCLA. Also in *UCLA Working Papers in Phonetics* 65.
- [35] Hu, F. 2005. *A Phonetic Study of the Vowels in Ningbo Chinese*. Ph. D. dissertation, City University of Hong Kong. [2014. A revised version published by China Social Sciences Press, Beijing.]