DIPHTHONG ACOUSTICS AND ARTICULATORY UNIFORMITY

Fang Hu

University of CASS and the Institute of Linguistics, Chinese Academy of Social Sciences
hufang@hotmail.com

ABSTRACT

This paper examines acoustics and articulation in the production of diphthongs in Ningbo Chinese. The acoustic study is based on audio recordings from 20 speakers, and the articulatory data is based on EMA recordings from 6 speakers. Formant data indicate that there is a typological difference between falling and rising diphthongs. Falling diphthongs seem to have a dynamic spectral target, while rising diphthongs are sequences of two spectral targets. Articulatory kinematics reveals an articulatory uniformity in diphthong production. In conclusion, the observed acoustic difference between falling and rising diphthongs should be attributed to the nonlinear articulatory-to-acoustic relations.
Keywords: diphthong production, falling diphthong, rising diphthong, acoustics, articulatory uniformity.

1. INTRODUCTION

The Chinese language is well-known for its simple syllabic structure and complex tonal system. The complexity of vowels is essential to the understanding of syllables in Chinese dialects and related minority languages. A Chinese syllable is composed of an initial consonant (C), a glide (G), a vowel nucleus (V), and a coda (Co), which could be represented as CGVCo, where only the nucleus V is obligatory while the other elements are all optional. There is a consensus that Chinese syllable doesn't allow consonant clusters. That is, the glide G is treated as a vowel, rather than a consonant. However, there are different analyses of Chinese vowel elements and different proposals for syllable structure [1, 2, 3, 4, 5, 6, 7, 8]. It is quite natural that different phonological analyses yielded non-unique solutions [9]. The fact is that GV forms a rising diphthong, VCo forms a falling diphthong if Co is a vocalic element, and GVCo forms a triphthong, no matter G and vocalic Co is treated as a vowel or an approximant.

This paper focuses on controversy regarding the nature of diphthongs. There are two kinds of viewpoints. Some scholars argued that diphthongs are an articulatory event with a dynamic target; that is, a diphthong is a single phonological vowel with a complex phonetic nucleus [10, 11, 12]. The others simply treated diphthongs as being composed of two articulatory events, transiting from a static target to the other; that is, a diphthong is a sequence of two monophthongal vowels [13, 14]. This paper examines spectral properties and articulatory kinematics of diphthong production in Ningbo Chinese, a language with both falling and rising diphthongs, and aims to provide fine-grained phonetic details to the understanding of diphthong production in general and the understanding of syllable structure of Chinese in particular.

2. METHODOLOGY

In acoustic studies, 20 average adult speakers, 10 male and 10 female were recorded. All speakers were born and raised up locally and have no reported speech and hearing disorders. Monosyllabic words containing target vowels or diphthongs were used as test words. Each test word X was placed in the middle of a carrier frame. Audio sound were recorded directly into a laptop through a DMX 6 Fire USB soundcard with a Shure SM 86 or 58 microphone. Audio recordings were conducted in quite rooms during the fieldwork trips. The sampling rate is 11,025 Hz or 22,050 Hz, 16 bit, and five repetitions were recorded. Monophthongs were annotated as one segment; diphthongs are typically composed of two steady states and a transition, and were annotated as three segments, i.e., an onset, an offset, and the transition connecting them. The lowest three formants were extracted from the spectrogram in the mid-point of each target segment.

6 native adult speakers were recorded for the articulatory study by using the Carstens 2D AG100 or AG200 system. Sensors were attached at the midline on the upper and lower lip, the gum ridge at the lower teeth (jaw), and the three points of the tongue, namely the tongue tip (TT), tongue mid (TM), and tongue dorsum (TD). In addition, two receiver coils were attached to the bridge of the nose and the gum ridge at the upper teeth respectively, serving as the two reference points. The original 500 Hz EMA data were down sampled to 250 Hz and smoothed with the sensor on the lower lip being smoothed with a filter that has a cut-off frequency at 40 Hz and the rest of the sensors with a filter that has a cut-off frequency at 25 Hz. The positions of the tongue, jaw, and lip transducers at the target point in a diphthong element were extracted from

the EMA data by using the tangential velocity minima criterion [15, 16]. The criterion was applied with reference to the sampled principal lingual point, that is, TM for diphthong elements in [ai] and [ia], and TD for diphthong elements in [au] and [ua].

Due to the space limit, the discussion was focused on the two pairs of falling versus rising diphthongs, i.e., [ai] versus [ia] and [au] versus [ua], for comparison's sake.

3. RESULTS

3.1. Specteral properties

Chinese dialects have rich inventories of diphthongs. There are usually both falling and rising diphthongs. Northern Wu dialects are famous for the richness in monophthongs and the lack of falling diphthongs [17, 18]. The Ningbo dialect is a typical northern Wu. There are 12 monophthongs [ɿ ʯ i y ʏ e ø ɛ a ɔ o u] including the two apical vowels, 3 falling diphthongs [ai au œy], 6 rising diphthongs [ia ie io yo ua uaʔ uɛ], and 1 triphthong [uai] [19].

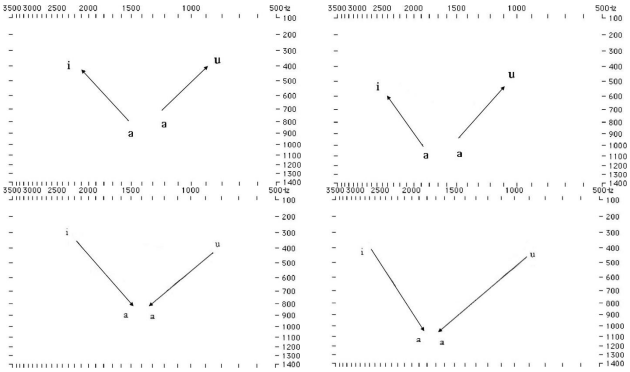

Figure 1: The paired falling diphthongs [ai au] (upper) and rising diphthongs [ia ua] (lower): averaged formant data in Hertz from 10 males (left) and 10 females (right).

Figure 1 plotted the paired falling diphthongs [ai au] (upper) and rising diphthongs [ia ua] (lower) in a two-dimensional acoustic space (F2 against F1) with the origin of the axes to the top right of the plot. The arrow connected diphthong onset to offset, which were averaged from 5 repetitions across 10 male (left) or female (right) speakers. The scaling relation of the axes has been converted to the bark to better approximate the perceptual distances. And the scale on the ordinate is double that on the abscissa to give appropriate prominence to F1 and make the plots more in accord with auditory judgments of vowels. However, the values along the axes still correspond to the original values in Hertz. The F1/F2 vowel plane establishes a good correlation with linguistic vowel features such as height and backness, namely F1 with vowel height and F2 with vowel backness.

Figure 2 compares diphthong elements of falling diphthongs with their monophthongal counterparts [i a u]. Confidence ellipses with radii of two standard deviations were drawn along axes oriented along the principal components of each vowel cluster. The descriptive statistics of the 2-sigma vowel ellipse predicts about 86% of the distribution of data points, and thus provides a straightforward way to examine the variability of diphthong elements vis-à-vis their monophthongal counterparts. It is apparent from the figure that diphthong onsets behave differently to diphthong offsets. And data are consistent across male and female speakers. On one hand, as compared with the monophthongal [a], [a] in [ai] becomes more front and [a] in [au] more back, due to the coarticulatory effect of the offset [-i] and [-u] respectively. But on the other hand, both diphthong elements and the monophthong [a] are all located in the height of low vowel, and their ellipses overlap with each other to certain extent. Diphthong offsets are different. Both [-i] in [ai] and [-u] in [au] don't reach their target high vowel positions: [-i] in [ai] stops at mid-high [e] or even mid-low [ɛ] region, and accordingly [-u] in [au] stops at mid-high [o] or even mid-low [ɔ] region. The ellipse for [-i] in [ai] even doesn't overlap with the monophthongal [i]; the ellipses for [-u] in [au] and the monophthongal [u] overlap marginally. And the apparently bigger ellipses for diphthong offsets, especially in female speakers, also suggest greater variability for diphthong offsets than for their monophthongal counterparts. In summary, formant data suggest that the onsets are better controlled than the offsets in the production of falling diphthongs [ai au]. It seems that in a falling diphthong, only the onset has a spectral target, and the offset could be viewed as a consequence of articulatory movement and is constrained by diphthong dynamics.

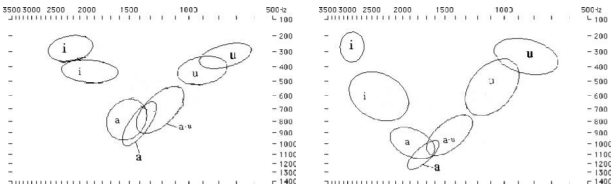

Figure 2: Comparison between diphthong elements of falling diphthongs (small IPA) and their monophthongal counterparts [i a u] (large IPA): 2-sigma confidence ellipses from 10 males (left) and 10 females (right).

Figure 3 compares diphthong elements of rising diphthongs in small IPAs with their monophthongal counterparts [i a u] in large IPAs. As shown in the figure, although there is certain coarticulatory effect, both the onset and offset in a rising diphthong have comparable vowel heights and elliptic distributions

with their monophthongal counterparts in the acoustic F1/F2 plane, suggesting that diphthong elements are as well controlled as are the corresponding monophthongs. In short, formant data suggest that both the onset and offset seem to have spectral targets in rising diphthongs [ia ua].

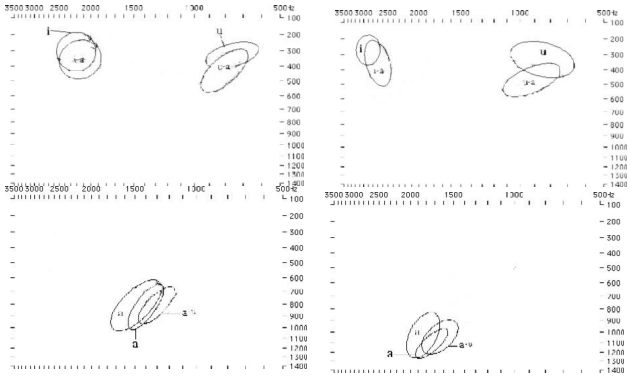

Figure 3: Comparison between diphthong elements of rising diphthongs (small IPA) and their monophthongal counterparts [i a u] (large IPA): 2-sigma confidence ellipses from 10 males (left) and 10 females (right).

3.2. Articulatory uniformity

Diphthong production could be understood as the course of spectral movement from its onset to offset. It has been shown so far that both onset and offset in [ia ua] have comparable spectral targets as their monophthongal counterparts do, whereas [ai au] seem to have a dynamic spectral target since their onset is much better controlled than their offset. A further step inquiring into the production mechanism of diphthongs is to examine the kinematics of diphthong articulation, especially the lingual articulation. Figure 4 and 5 show mean articulatory positions in millimetre (n=5) for the sampled articulators in the production of falling diphthongs [ai au] and rising diphthongs [ia ua] from 6 speakers. Diphthong elements were denoted by large IPA symbols and the three tongue points were connected by solid lines; the corresponding monophthongs were denoted by small IPA symbols and the three tongue points were connected by dot lines. Speakers are facing left in the figures.

Although there are inter-speaker variations, the onset in falling diphthongs [ai au] has articulatory positions comparable to those for the monophthong [a] in general, in terms of both lingual articulation and jaw position. However, there are a few exceptions. For instance, [a] in [ai au] has a higher jaw position than the monophthong [a] in female speaker 1. Unlike the onset, the offset in [ai au] apparently doesn't reach the articulatory positions for the monophthong [i] or [u], in terms of either lingual articulation or jaw position. The articulatory data are consistent with formant data discussed in Section A. It seems that in the production of falling diphthongs [ai au] in Ningbo Wu, the onset has both articulatory and acoustic targets, whereas the offset is probably a dynamic consequence of diphthong movement, as it has no specific articulatory or acoustic targets.

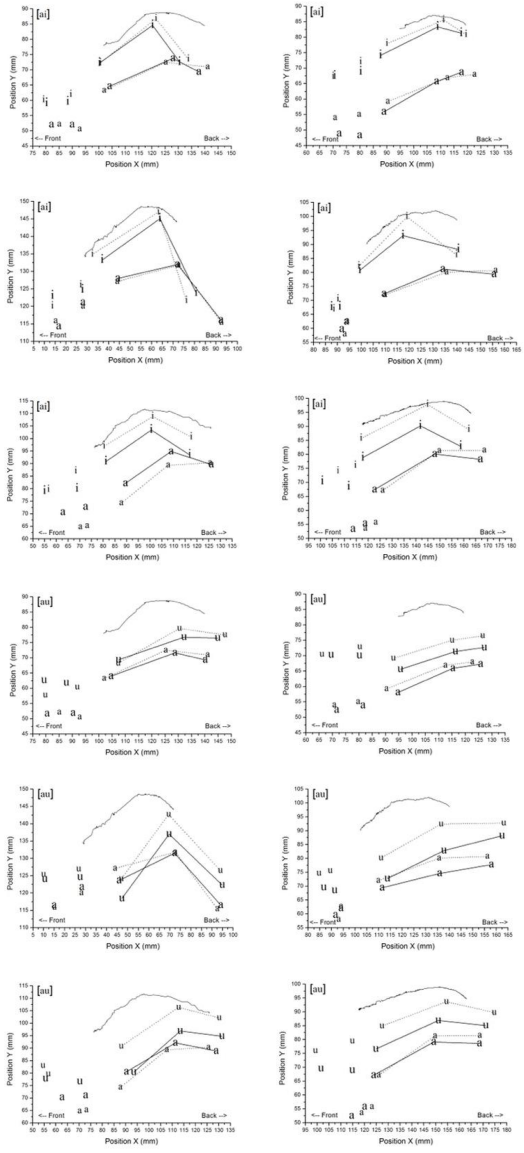

Figure 4: Mean lower lip, jaw, and three points on the tongue in mm (from left to right) in the production of falling diphthongs [ai au].

The articulation of the rising diphthongs [ia] demonstrates a similar pattern to that of falling diphthongs. First, [i] in [ia] has generally comparable articulatory positions to the monophthong [i] in terms of both lingual articulation and jaw position. This corroborates the formant data, suggesting a good control for the onset target. But what is not expected is that [a] in [ia] has apparently different articulatory positions with the

monophthong [a] in most speakers. The speakers don't seem to employ a specific articulatory strategy for the realization of the offset spectral target in the production of rising diphthong. Instead, there is articulatory uniformity regarding both falling and rising diphthongs.

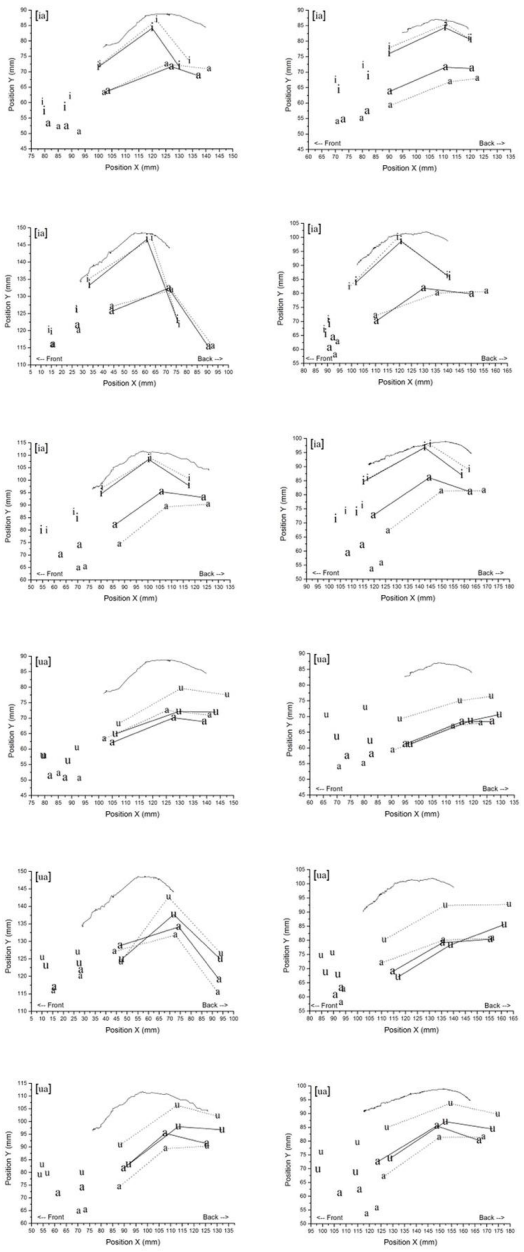

Figure 5: Mean lower lip, jaw, and three points on the tongue in mm (from left to right, n=5) in the production of rising diphthongs [ia ua].

The question is why a rising diphthong has a better offset controlled spectral offset. This could possibly be explained by the articulatory-to-acoustic relations. EMA recording sampled three tongue points in the mouth. But the constriction location for the low vowel [a] is in lower pharynx [20, 21]. So, it is reasonable to speculate that the observed variability in lingual articulation doesn't affect the realization of articulatory target in pharynx, since a general low lingual position, rather than a precisely controlled specific position, facilitates the constriction in pharynx. The fact that the offset [a] in a rising diphthong has great variability in articulation but a good controlled spectral target corroborates the quantal nature of the low vowel [a] [20, 22, 23, 24, 25]. Regarding the other rising diphthong [ua], the onset [u] in [ua] demonstrates anticipatory effect of coarticulation, namely it has apparently lower lingual and jaw positions than the monophthong [u]. As shown in Fig.3 in Section A, [u] in [ua] also have a lower distribution than the monophthong [u] in the acoustic F1/F2 vowel plane. However, they are both located in the high back region in the vowel plane. The offset [a] in [ua] behaves similar to [a] in [ia].

In summary, articulatory uniformity is observed for the production of both falling and rising diphthongs. Speakers seem to employ a uniform articulatory strategy to produce a diphthong. First, no matter in a falling or rising diphthong, diphthong onset is better controlled than diphthong offset both articulatorily and acoustically, although it subject to certain coarticulatory effect. Second, both falling and rising diphthongs are constrained by articulatory kinematics as well as the inherent sluggishness of articulators [26, 27]. Third, the difference in acoustic outputs for falling and rising diphthongs is constrained by the articulatory-to-acoustic relation and the quantal nature of the three point vowels [i a u]. As a result, falling diphthongs have a dynamic spectral target, while rising diphthongs are composed of two static spectral targets.

4. CONCLUSION

Acoustic data suggest a typological difference between falling and rising diphthongs. Falling diphthongs seem to have a dynamic spectral target, while rising diphthongs are sequences of two spectral targets. But articulatory kinematics reveals an articulatory uniformity in diphthong production, and the observed spectral difference between falling and rising diphthongs could be attributed to the articulatory-to-acoustic relations and the quantal nature of the three point vowel targets [i a u].

Results support a dynamic view of vowel production. That is, the fundamental unit of vowel could be dynamic. In Ningbo Chinese that has been examined, a falling diphthong has a dynamic spectral target, and should thus be treated as one vowel phoneme; a rising diphthong has two static spectral targets, and can thus be viewed as a sequence of two vowel phonemes.

7. REFERENCES

[1] Hartman, L. M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20, 28-42.

[2] Luo, C., Wang, J. 1957. *An Outline of General Phonetics* [in Chinese]. Science Press, Beijing.

[3] Cheng, R. L. 1966. Mandarin phonological structure. Journal of Linguistics, 2, 135-158.

[4] Chao, Y.-R. 1968. *A grammar of spoken Chinese*. University of California Press, Berkeley.

[5] Cheng, C. C. 1973. *A synchronic phonology of Mandarin Chinese*. Monograph on Linguistic Analysis, No. 4. Mouton.

[6] You, R., Qian, N., Gao, Z. 1980. On the phonemic system of Putonghua [in Chinese], *Chinese Language* 5, 328-334.

[7] Duanmu, S. 2008. *Syllable Structure: The Limits of Variation*. Oxford University Press.

[8] Duanmu, S. 2017. Syllable and syllable structure in Chinese. In: Sybesma, R. (editor-in-chief), Behr, W., Gu, Y., Handel, Z., Huang, C.-T. J., Myers, J. (eds.) *Encyclopedia of Chinese Language and Linguistics* Brill, 230-236.

[9] Chao, Y.-R. 1934. The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology*, Academia Sinica, 4, 363-397.

[10] Malmberg, B. 1963. *Structural linguistics and human communication*. Springer-Verlag.

[11] Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh University Press.

[12] Catford, I. 1977. *Fundamental problems in phonetics*. Edinburgh University Press.

[13] Sweet, H. 1877. *A handbook of phonetics including a popular exposition of the principles of spelling reform*. Clarendon Press, Oxford.

[14] Jones, D. 1922. *Outline of English phonetics*, 2nd Ed. E. P. Dutton, New York.

[15] Löfqvist, A., Gracco, V. L., Nye, P. W. 1993. Recording speech movements using magnetometry: One laboratory's experience. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 31, 143-162.

[16] Löfqvist, A. 1999. Interarticulator phasing, locus equations, and degree of coarticulation. *The Journal of the Acoustical Society of America* 106, 2022-2030.

[17] Chao, Y.-R. 1928. *Studies in the Modern Wu Dialects*. Peking: Tsing Hua College Research Institute.

[18] Yuan, J. et al. 1960. *An outline of Chinese dialects* [in Chinese]. Script Reformation Press, Beijing.

[19] Hu, F. 2014. *A Phonetic Study on the Vowels in Ningbo Chinese*. China Social Sciences Press.

[20] Stevens, K. N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In: David, E. E. Jr. & Denes, P. B. (eds.). *Human Communication: A Unified View*. McGraw-Hill, 51-66.

[21] Wood, S. 1979. A radiographic analysis of constriction locations for vowels. *Journal of Phonetics* 7: 25-43.

[22] Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.

[23] Perkell, J. S., Nelson, W. L. 1982. Articulatory targets and speech motor control: A study of vowel production. In Grillner, S., Persson, A., Lindblom, B. Lubker J. (eds.). *Speech Motor Control*. Pergamon, 187-204.

[24] Perkell, J. S., Nelson, W. L. 1985. Variability in production of the vowels /i/ and /a/. *The Journal of the Acoustical Society of America* 77: 1889-1895.

[25] Perkell, J. S., Cohen, M. H. 1989. An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics* 17: 123-133.

[26] Saltzman, E. L., Kelso, J. A. S. 1983. Skilled actions: A task dynamic approach. *Haskins Laboratories Status Report on Speech Research* SR-76: 3-50.

[27] Kelso, J. A. S., Saltzman, E. L., Tuller, B. 1986. The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 14: 29-59.