AN ACOUSTIC PHONETIC ANALYSIS OF DIPHTHONGS IN NINGBO CHINESE

Fang Hu

Institute of Linguistics, CASS
fong_hwu@yahoo.com

ABSTRACT

This paper describes the acoustic phonetic properties of diphthongs in Ningbo Chinese. Data from 20 speakers indicate that (1) falling diphthongs have both onset and offset steady states while rising diphthongs only have steady states on the offset element; (2) both falling and rising diphthongs begin from an onset frequency area close to their target vowels, but only the normal-length rising diphthongs reach the offset target, and falling and short rising diphthongs stop at somewhere before reaching the target; (3) diphthongs can be well characterized by the F2 rate of change as far the falling diphthongs are concerned, whereas data lack consistency when rising diphthongs are also taken into account. Results show that the temporal organization within diphthongs, formant patterns, and formant rate of change all contribute to the characterization of Ningbo diphthongs.

1. INTRODUCTION

Ningbo Chinese is a variety of Wu dialects in the south-eastern costal region of China. In addition to a rich inventory of vowels [8 4 4 [illegible]0 [illegible] 4 0], Ningbo has nine diphthongs [0 8 0 0 4 80 84 8 4 4 4 0 0 0[illegible]]. Diphthongs [0 8 0 0 4 80 8 4 0[illegible]] occur in the (C)D position, [84 4 4] occur in the (C)D< position, and [0 0] occurs in both (C)D and (C)D< positions (D stands for Diphthongs). In Chinese literature, distinction is generally made between rising and falling diphthongs. The former refers to diphthongs beginning with a high vowel [8 0 4] and ending with the syllabic nucleus; the latter refers to diphthongs beginning with the syllabic nucleus and ending with a high vowel. Chao ([1]: 65-66) pointed out that only falling diphthongs are true diphthongs in Wu dialects. In the present study, attention is paid to see if there are any acoustic correlates with such a falling-rising patterning.

2. METHODOLOGY

Meaningful monosyllabic words were used as test words. All test words have a high level tone and a zero initial, i.e., without an initial consonant. 20 Ningbo native speakers, 10 male and 10 female, were recorded during my fieldwork in the Ningbo city in the summer of 2001, using a Sony PCM-R700 Digital Audio Recorder and a Shure SM-58 Microphone. All subjects were born and raised up in Ningbo and did not have a history of speech or hearing problems. Each test word was placed in a carrier sentence [[illegible]4 8 4 3 4<__ 5 0 < 3 0 0 9[illegible]8[illegible]] "I will read __ for you (to) listen". And the test monosyllabic words were randomized and the speakers were instructed to read the word list in a natural way with a moderate speech tempo. Five repetitions of the word list for each subject were recorded.

Speech data were digitized at a sampling rate of 10K Hz per second. Data analysis was carried out on the Kay CSL4400 speech analysis software. First, for each diphthong, formant trajectory was performed, using the pitch-synchronous method and generally 10 coefficients for male and 12 coefficients for female speakers, to see if there is steady state for the target elements in the diphthong. Second, durations of the first and second target element and the transition in a diphthong were measured directly from the speech waveform based on the formant movement information from the formant trajectory. Third, in order to get the first two formant frequency values of target elements in a diphthong, if the target element has steady state, a pitch-synchronous LPC analysis was performed at the mid-point of that element; if the target element doesn't have any steady state, the point where a high-vowel element has the lowest F1 value or a non-high-vowel element has the highest F1 value was chosen to perform a LPC analysis. And wideband spectrograms were always produced simultaneously for reference. Finally, the formant change range between two elements in a diphthong and the rate of change were then calculated to determine the dynamic properties of diphthongs.

3. RESULTS

3.1. Duration results

Table 1: Mean durations (in millisecond) and SDs of Ningbo diphthongs

Diph.	Female speakers			
	1st Ele.	Trans.	2nd Ele.	Total
[0 8]	80 (23)	77 (20)	58 (18)	215 (53)
[0 0]	82 (24)	77 (20)	56 (18)	215 (55)
[4]	80 (21)	80 (24)	65 (19)	224 (57)
[80]	0	116 (24)	116 (31)	232 (52)
[84]	0	71 (16)	64 (17)	135 (28)
[8 4]	0	127 (29)	122 (42)	249 (66)
[0 0]	0	110 (21)	126 (31)	236 (49)
[0 0 (ᕙ]	0	76 (21)	68 (17)	144 (35)
[0]	0	119 (23)	133 (43)	252 (60)
[4 4]	0	75 (18)	73 (19)	148 (31)
	Male Speakers			
[0 8]	64 (16)	85 (36)	62 (20)	210 (52)
[0 0]	66 (17)	76 (24)	59 (21)	200 (49)
[4]	64 (15)	78 (26)	59 (14)	202 (43)
[80]	0	132 (24)	114 (38)	245 (55)
[84]	0	76 (17)	62 (19)	137 (32)
[8 4]	0	127 (32)	120 (46)	247 (69)
[0 0]	0	112 (27)	135 (44)	248 (66)
[0 0 (ᕙ]	0	77 (18)	64 (17)	141 (30)
[0]	0	113 (27)	127 (45)	240 (61)
[4 4]	0	83 (22)	70 (22)	152 (39)

The detailed mean duration results, with Standard Deviations in parentheses right after the means, are summarized in Table 1. As can be seen from the table, both target elements of a falling diphthong have steady states, whereas only the second target element of a rising diphthong has steady state, and in general, the duration of the first target element is greater than that of the second target element, although the difference may not be significant, especially for the male speakers. Generally, the duration of transition is the greatest for female speakers, and for male speakers, the duration of transition is similar to that of the first element while greater than that of the second element. Duration results of rising diphthongs exhibit a totally different picture. There is no steady state on the first element. Only the second element, i.e. the syllable nuclear vowel element, has steady states. Regarding the temporal relationship between the transition and the second element of a rising diphthong, data lacks consistency. In [0 0 0], the transition has a comparatively smaller duration, whereas in [8 4 80], the duration of the transition is either similar to or a little bit greater than that of the nuclear vowel element. Concerning the three short diphthongs [84 0 0 (ᕙ 4 4], the transitions always have a relatively greater duration, which suggests that the duration of the nuclear vowel element is more flexible to compress than that of the transition when the duration of the whole syllable is compressed. This is well illustrated by a comparison between the normal-length [0 0] and short [0 0 (ᕙ].

3.2. Spectral results

Frequencies of the first two formants of the two target elements of Ningbo diphthongs were summarized in Table 2.

Table 2: Means (in Hz) and SDs of frequencies of the first two formants of Ningbo diphthongs

Diph.	Female speakers			
	1st Ele. F1	1st Ele. F2	2nd Ele. F1	2nd Ele. F2
[0 8]	995 (74)	1878 (154)	609 (90)	2387 (233)
[0 0]	929 (76)	1458 (117)	543 (101)	1092 (103)
[4]	764 (55)	1417 (102)	542 (82)	2102 (123)
[80]	415 (95)	2647 (114)	1032 (104)	1839 (105)
[84]	358 (76)	2699 (135)	669 (61)	2218 (234)
[8 4]	319 (56)	2388 (155)	497 (88)	1155 (114)
[0 0]	470 (58)	925 (100)	1049 (82)	1645 (102)
[0 0 (ᕙ]	483 (77)	933 (113)	847 (110)	1481 (127)
[0]	428 (51)	933 (103)	719 (75)	2254 (129)
[4 4]	353 (73)	2208 (152)	663 (73)	1725 (230)
	Male Speakers			
[0 8]	789 (86)	1529 (107)	434 (40)	2069 (256)
[0 0]	710 (94)	1220 (103)	422 (39)	907 (80)
[4]	628 (52)	1237 (111)	420 (45)	1829 (259)
[80]	358 (63)	2159 (164)	814 (105)	1495 (128)
[84]	309 (62)	2220 (156)	541 (50)	1894 (175)
[8 4]	309 (44)	2043 (133)	450 (40)	1053 (69)
[0 0]	434 (53)	851 (68)	810 (77)	1306 (79)
[0 0 (ᕙ]	430 (73)	810 (73)	704 (120)	1188 (113)
[0]	407 (51)	865 (84)	585 (71)	1833 (196)
[4 4]	311 (52)	1889 (111)	541 (67)	1527 (156)

Let's examine the falling diphthongs first. Figure 1 shows three falling diphthongs plotted in an acoustic F1/F2 plane, F2 against F1, with the origin on the right-top corner. Elements of diphthongs are used as the starting and ending points respectively and the arrows indicate the movements. It is generally agreed that diphthongs can be viewed as a movement from an onset vowel element to an offset vowel element. However, whether the diphthong begins (should begin) from and ends (should end) at the exact element target and to what extent the diphthong elements vary with the original vowels are controversial, because in the production of diphthongs, target undershoots may take place due to the relative short duration allocated for a particular target element. In Figure 2 the diphthong element is compared with the corresponding target vowel to examine this phenomenon. The vowel element [ ] in the diphthong [4] is compared with [ ], because it is rather a back vowel element [ ], as shown in the figure, than a rounded mid-low front vowel following the traditional transcription (e.g., [1]).

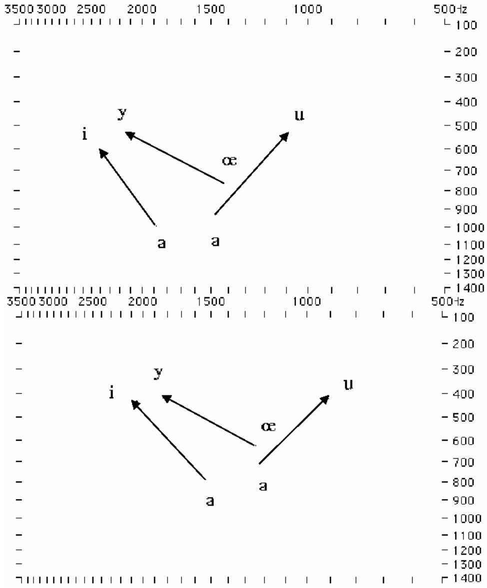

Figure 1. Ningbo falling diphthongs. Top: females' data; Bottom: males' data.

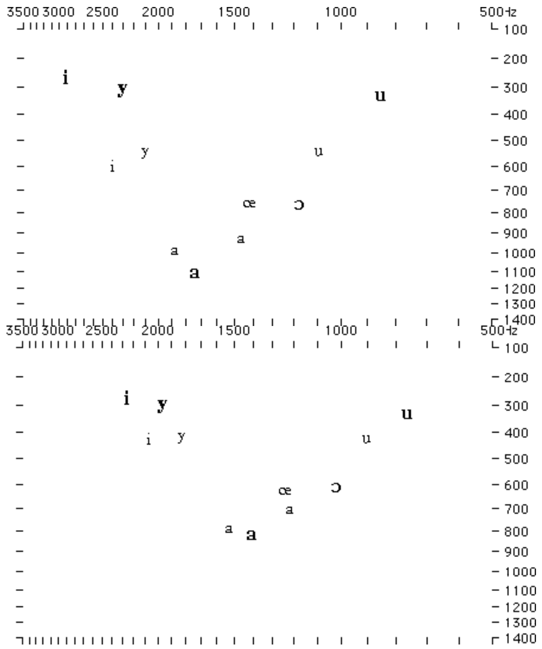

Figure 2. Elements of Ningbo falling diphthongs (with small-size IPA symbols) compared with target vowels (with bold big-size IPA symbols). Top: females' data; Bottom: males' data.

From Figure 2 we can see that the vowel element [0] in the diphthong [0 8] is slightly more front than the single vowel [0] and the [0] in [0 0] is quite back due to the coarticulatory effect of the following front and back vowel element [8] and [0] respectively. However, they are both in the low vowel position, and keep a certain distance with the mid-low vowel [ ] in both female and male speakers. The vowel element [ ] in [4] is close to [ ] and has the same vowel height, yet somewhat centralized, and can be better represented with the rounded mid-low central IPA vowel symbol [ɞ]. Although target undershoots take place, falling diphthongs in Ningbo, roughly speaking, begin from an onset near to its target position. Unlike the onset element data, offset elements of Ningbo falling diphthongs are quite far apart from the target position. Diphthong elements [8 4 0] just reached the vowel positions of [4 4] respectively or lower.

Seven rising diphthongs were plotted in Figure 3. To distinguish between the normal-length and short [0 0], the latter was labeled as [0 0] in the figure. From the figure we can see that diphthongs [84 80] have a similar onset ([84] is slightly more peripheral), whereas [8] in [8 4] is closer to [4]. And diphthongs [0 0 0 0 <0] also have a similar onset position, although [0] in [0] is a little higher.

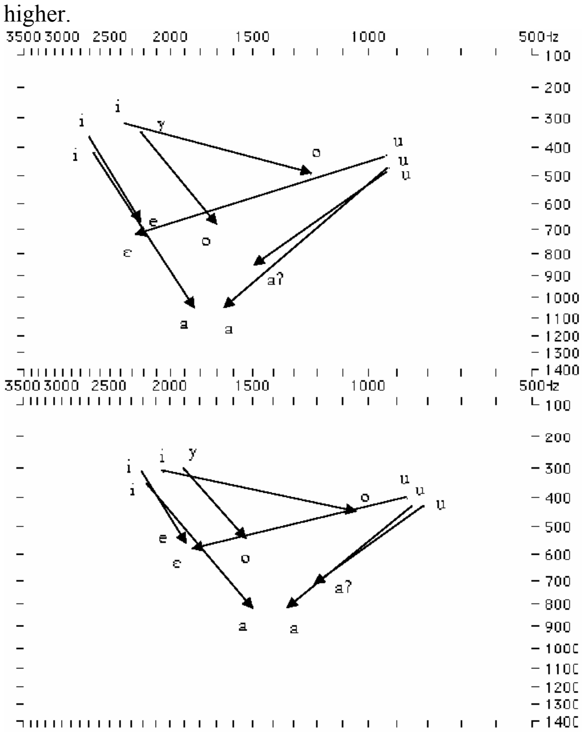

Figure 3. Ningbo rising diphthongs. Top: females' data; Bottom: males' data.

To examine the relationship between rising diphthong elements and the corresponding target vowels, they were plotted in Figure 4. In order to make necessary distinction, the second element in diphthongs [0 0 0 0] was labeled with [ ] respectively, and the second element in [4 4] was labeled with [4]. And the first element [8] and [0] in different diphthongs was denoted by the following vowel element. From Figure 4 we can see that [8] in [84] and [80] is slightly lower than the target vowel, and [8] in [8 4] is very close to the target vowel [4], which indicates the labialisation effect due to the following rounded vowel element. And [4] in [4 4] is somewhat lower than the target [4]. However, they all are in the high vowel region and keep quite distant to the mid-high vowel [4]. The high back vowel element [0] in [0 0 0 0 0] is also lower than the target vowel. Moreover, it approaches the position of the mid-high low vowel [4], especially in the males' data, which suggests a greater articulatory undershoot. However, it may partially be attributed to the fact that vowels distribute closer to each other in the back dimension than in the front dimension. As for the second element of rising diphthongs, diphthongs differ from one another. Generally, second elements of the normal-length diphthongs are very close to their targets, except that the [4] in [8 4] is shifted leftward to a central position to a greater extend. And [0] in [80] and [0 0] distributes slightly more front and back to the target vowel respectively. In short diphthongs, the second element seldom reaches its target position. [4] in [4 4] occupies a central position in the acoustic vowel plane. [0 ( ] in [0 0 ( ] is significantly higher from its target low vowel position. And it is interesting that [4] in [84] is much lower than the target vowel [4] and is even very close to the vowel [ ]. However it is not a case of "target overshoot". It is still the case of target undershoot, because the element is shifting toward a central position.

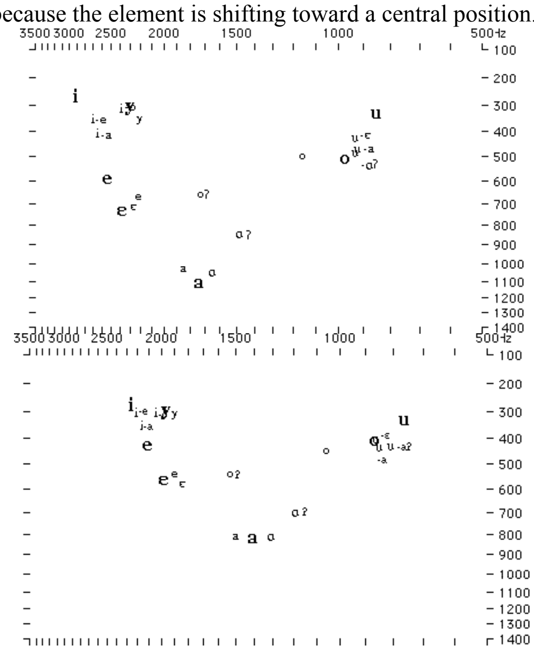

Figure 4. Elements of Ningbo rising diphthongs (with small-size IPA symbols) compared with target vowels (with bold big-size IPA symbols). Top: females' data; Bottom: males' data.

3.3. Dynamic aspects

The dynamic aspect of diphthong movement is observed by means of the measurement of the range and rate of F2 change in the present study, since it is well documented in the literature that F2 transition rate of change is one of the most important acoustic properties of diphthongs (e.g., [2], [3], [4]). In Table 3, a negative number denotes the movement with an opposite direction, i.e., from a relatively front position to a more back position. As far the falling diphthongs are concerned, different diphthongs are characterized by their own distinctive F2 rate of change, namely [4] has the greatest F2 range of change and the fastest F2 rate of change, [0 8] has a median F2 range of change and a median F2 rate of change, and [0 0] has the smallest F2 range of change and the slowest F2 rate of change. However, if rising diphthongs are also taken into account, the data lack consistency and it can hardly be generalized that diphthongs can be characterized by their F2 rate of change: first, diphthongs of female speakers have different ordering with those of male speakers; second, some diphthongs have very similar values with each other, for example, [4 4 0 0 0 8 84 80 0 0 ( ] have values increasing from 6.44 to 7.21 Hz/ms in females' data, and [0 0 0 0 84 4 4 0 0 (80] have values increasing from 4.05 to 5.03 Hz/ms in males' data; third, a greater F2 range of change doesn't always correlate to a faster F2 rate of change, for example, in females' data, [0 0] has an F2 range of change as great as 721 Hz and [4] has a smaller F2 range of change of 685 Hz, whereas the F2 rate of change of [4], 8.56 Hz/ms, is faster than that of [0 0], 6.55 Hz/ms. Nevertheless, some generalizations can still be reached. Diphthongs [0 8 4 4] occupy the top three position in terms of F2 rate of change in both female and male speakers' data. They all have relatively greater movements across the front-back axis. It can be observed from the table that the ordering of the short diphthongs [0 0 (84 4 4] is rather arbitrary if females' data are compared with males'. What can be pointed out is that the short [0 0 ( ] always has a slightly faster F2 rate of change than its normal-length counterpart [0 0]: 7.21 Hz/ms to 6.55 Hz/ms in females' data, and 4.91 Hz/ms to 4.05 Hz/ms in males' data. It can probably be inferred that the short diphthong has a relatively faster F2 rate of change. Note that the short diphthong [84] even has a similar F2 rate of change with the normal-length [80] in females' data, although the latter has a much greater F2 range of change. This is due to the fact that in short syllables, diphthongs have shorter duration to accomplish the movement. If the top three diphthongs [0 8 4 4], which is consistent in both female and male speakers' data, and the short diphthongs discussed above are excluded, the left four diphthongs [0 8 0 0 80 0 0], two pairs of falling and rising diphthongs, display an interesting ordering of F2 rate of change. For the female speakers, rising diphthongs always have faster F2 rate of change than their falling counterparts, but vice versa for the male speakers. However, the difference between the falling and rising diphthongs may be not significant, especially the [80] and [0 8] in females' data and [0 0] and [0 0] in males' data.

4. CONCLUSION AND DISCUSSION

Falling diphthongs and rising diphthongs in Ningbo differ in terms of the temporal organization: the falling diphthongs have both onset and offset steady states; the rising diphthongs only have steady states on offset elements. And in general, both the falling and rising diphthongs begin from an onset frequency area close to their target vowels, while only the normal-length rising diphthongs end at the frequency region close to the offset target, whereas the falling and short rising diphthongs stop at somewhere before reaching the offset target position. Regarding the dynamic aspects of diphthongs, diphthongs are characterized by F2 rate of change as far the falling diphthongs are concerned, whereas the data lack consistency when all rising diphthongs are taken into account. Results show that the temporal organization within diphthongs, formant patterns, and formant rate of change all contribute to the characterization of Ningbo diphthongs. The Ningbo data support the viewpoint that the acoustic phonetic character may differ from language to language, and generalizations made within one language may be not valid for other languages. Roughly speaking, in languages with a poor inventory of diphthongs, especially those only have falling diphthongs, diphthongs can be well characterized by invariant F2 rate of change (see, e.g., [2] for English and [3] for Maithili). But in languages with a rich diphthong inventory, diphthongs can hardly be characterized by a single acoustic property. Manrique ([4]) found in Spanish, some diphthongs [0 4] and [48], [48] and [8 4] have very similar F2 rate of change and thus cannot be characterized by the F2 rate of change alone, although F2 rate of change is the most stable parameter when the speaking rate is changed. Ren ([5]) suggests that the type of diphthong (falling or rising, backness, direction of movement, and size of movement) can affect the F2 rate of change. And results from the Ningbo data also suggest that the short diphthong tends to have a faster F2 rate of change than its normal-length counterpart. Finally, what should be pointed out that the dichotomy between falling and rising diphthongs in terms of different temporal organizations may also be a language-specific feature. Previous studies show that, in Beijing Mandarin, falling diphthongs [48 4 0] don't have steady states on onset element, while rising diphthong [84] has steady state on onset but not on offset element ([5]).

Table 3. F2 range (in Hz) and rate (Hz/ms) of change of Ningbo Diphthongs

DIPH.	FEMALES		MALES	
	ΔF2	Rate	ΔF2	Rate
[0 8]	510	6.62	539	6.34
[0 0]	-366	-4.75	-313	-4.12
[4]	685	8.56	592	7.59
[80]	-808	-6.97	-664	-5.03
[84]	-481	-6.77	-326	-4.29
[8 4]	-1233	-9.71	-990	-7.80
[0 0]	721	6.55	454	4.05
[0 0 (ɘ)]	548	7.21	378	4.91
[0]	1320	11.09	968	8.57
[4 4]	-483	-6.44	-362	-4.36
[4 4]	-483	-6.44	-362	-4.36

5. REFERENCES

[1] Chao, Y. R. *Studies in the modern Wu-dialects*. Tsing Hua College Research Institute, Peking, 1928.

[2] Gay, T. "Effects of speaking rate on diphthong formant movements", *J. Acoust. Soc. Amer.*, 44:1570-1573, 1968.

[3] Jha, S. K. "Acoustic analysis of the Maithili diphthongs", *J. Phonetics*, 13:107-115, 1985.

[4] Manrique, A. "Acoustic analysis of the Spanish diphthongs", *Phonetica*, 36:194-206, 1979.

[5] Ren, H. M. On the acoustic structure of diphthongal syllables. UCLA Working Papers in Phonetics, 65, 1986.