

The Heterogeneity Existing in Formant Patterns of Falling Diphthongs in Mandarin Chinese

WANG Ping*

Abstract

We conducted a statistical study on the formant patterns of falling diphthongs [ai], [ei], [au], and [ou] in Mandarin Chinese pronounced by 102 Beijing native speakers from the *Chinese speech database*. The results suggest that the internal performance of falling diphthongs in Chinese Mandarin wasn't completely homogeneous, and there were variants and variations to different degrees. Among them, both [ai] and [au] contained three formant patterns, [ou] contained two formant patterns, [ei] contained only one formant pattern. The monophthongization existing in [ai], [au], [ou] was mainly represented by the young native speakers.

1. Introduction

The compound vowel of Mandarin Chinese is composed of two or three monophthongs. Its pronunciation is characterized by the gradual sliding of one vowel into another. The position of the tongue and the shape of the lips gradually change, naturally flowing into each position and forming a whole diphthong. In terms of loudness of the pronunciation, the main vowels (rimes) have the widest mouth, most sonority, and longest duration, while the other vowels (medias or endings) are relatively short in duration and blurred in quality. According to the positions of the main vowels, the diphthongs are divided into two categories: falling diphthongs and rising diphthongs.

Domestic and foreign scholars have conducted many studies on the diphthongs in Mandarin Chinese. Wu & Cao (1979) measured the dynamic curves

*School of Literature, Nankai University, China

of the diphthongs of Mandarin Chinese and expressed them in their Formant Chart. Cao & Yang (1984) used Chinese characters with zero-initials pronounced by four male and female Beijing native speakers as experimental materials, and conducted a detailed quantitative analysis on the dynamic characteristics of thirteen compound vowels (diphthongs and triphthongs). Yang (1986) proposed an exponential dynamic model of compound vowels, which can describe the quality changes of thirteen compound vowels of Mandarin Chinese as dynamic movement between several target values. Based on the acoustic parameters of the zero-initial syllables of 15 people in the Mandarin Chinese Acoustic Parameters Database, Cao (1994) focused his analysis on the frequency characteristics of [au] and [ou] observing changes with time and the similarities and differences in their related structures, providing corresponding statistical parameters and related time-frequency covariant patterns. He, Ningji (1985) conducted a perceptual experiment to prove that there was a compensation relationship between time threshold of the diphthongs' perception and change rates of the sliding sections of the formants. Ren (1986) considered that the main feature of the compound vowel is determined by the second formant (F2). Therefore, taking F2 as his research object, he proposed the Truncation Model of the compound vowel. Eric (2001) conducted measurements and statistical analysis on the pole positions of the compound vowels of Mandarin Chinese pronounced by 20 Beijing native speakers (10 males and 10 females) using an acoustic method. Lehiste & Peterson (1961), Holbrook & Faibanke (1962) and Gay (1968) conducted acoustic analyses on the diphthongs of American English using a sonagraph. They divided diphthongs into three parts: onset, transition and offset, of which the onset and offset are usually long and stable, and their pronunciation relatively "in place".

The above research provides important reference points for us to understand the performance and properties of compound vowels. At the same time, we also find that there are divergences in the formant patterns of falling diphthongs in Mandarin Chinese. The focus is mainly on whether there is a stable part of the main vowel (rime) in a falling diphthong. Cao & Yang (1984) and Wu & Lin (1989:99) believe that the falling diphthongs of Mandarin Chinese are dynamic vowels whose formant frequencies change continuously with time, of which there are no stable parts. In contrast, He, Ningji (1985) suggested that the fall-

ing diphthong produced in isolation should be divided into three parts: the front stable section, the rear stable section, and the middle transitional section. Ren (1986) proposed that the main vowels of the falling diphthongs [ai] and [au] don't have to have steady sections, but are unstable and arbitrary, exhibiting possibly steady or quasi-steady states. The author believes that the reason for the above differences is probably due to the choice of speakers, among whom individual differences inevitably exist. In addition, the above research is mainly based on the case study of a few speakers, rather than statistical analysis of a large sample. Therefore, the universality and representative nature of the results of previous research have yet to be proven.

In response to the differences and shortcomings in previous studies, we conducted measurements, classifications, and calculations on the formant trajectories of falling diphthongs [ai], [au], [ou] and [ei], produced as the finals of monosyllables pronounced by 102 Beijing native speakers from the *Chinese Speech Database*¹⁾. The results are expected to fully reveal the formant patterns of falling diphthongs in Mandarin Chinese from the perspective of large sample statistics, and give some supplementation and explanation to previous uncertainties regarding the formant patterns of falling diphthongs.

2. Procedures for acoustic experiments and statistical analysis

2.1 Selection of speech samples

In the *Chinese Speech Database*, we selected 4 monosyllabic Chinese characters whose finals are the falling diphthongs [ai], [au], [ou] and [ei] respectively, and their initials are mainly zero-initials. In order to ensure the stability of the sound, all the tones are Yin-Ping (tone1). The monosyllabic Chinese characters are “哀”[ai⁵⁵], “熬”[au⁵⁵], “欧”[ou⁵⁵] and “杯”[pei⁵⁵]. There are 102 samples for each Chinese character (102 speakers for each character). A total of

1) The *Chinese Speech Database* was completed by Robert Sanders and Shi Feng. Currently, the database contains data for 102 Beijing native speakers. The gender distribution of the speakers: 48 female speakers and 54 male speakers. The age distribution of the speakers: under 20, 12; 20–29, 57; 30–39, 10; 40–49, 15; 50–59, 5; over 60, 3. We conducted two collections of sound materials in 2005 and 2006 respectively. The first collection included all age groups, and the second collection mainly included the under 20 and 20–29 age groups. Therefore, the number of young people in the database is larger than other age groups. This is more conducive to predicting the future direction of sound change according to the performance of young people.

Table 1. Gender and Age Distribution of speakers

	Under 20	20–29	30–39	40–49	Over 50	Total
Male	2	31	4	7	4	48
Female	8	28	6	8	4	54
Total	10	59	10	15	8	102

$102 \times 4 = 408$ valid samples were obtained.

2.2 *Extraction of acoustic parameters in the acoustic experiments*

We extracted the formant data of the vowels using Praat, including the first three formants: F1, F2 and F3. Specifically, the pre-written script was run on Praat to extract the acoustic data of the sound samples in normalized time, with the formant curve of the final equally divided into 10 parts regardless of its length (the formant curve doesn't include the transition between consonant and vowel). Finally, the formant data of 11 measurement points were obtained, and the acoustic unit used was the Hertz (Hz) value.

2.3 *Statistical analysis of formant data*

Using SPSS 22.0 as a statistical tool, we made statistical calculations on formant data. Before statistics, Stem-and-Leaf Plot was adopted to check whether there were outliers in the overall data. In order to ensure the overall objectivity of the data, we corrected the outliers caused by measurement errors, and eliminated the outliers caused by individual differences. The physiological differences between males and females could make the formant values significantly distinct, so we calculated the formant data of males and females separately.

3. **Statistical Analysis of Falling Diphthongs [ai], [au], [ou] and [ei]**

The data of 11 measurement points of each falling diphthong (each measurement point includes three dimensions of F1, F2 and F3) was obtained, so the data of four falling diphthongs of 102 speakers adds up to $11 \times 102 \times 3 = 3366$. All the data obtained by the acoustic analysis was then input into SPSS and classified according to the actual performance of formant trajectories (the classifica-

tion results will be described in detail below)²⁾. And the average values of male and female speakers for each category of each falling diphthong was calculated. Finally, the formant pattern of each vowel was plotted on the formant chart, which takes time (t) (normalization) as the horizontal axis and frequency (Hz) as the vertical axis.

3.1 Statistical representation of the formant patterns of [ai]

3.1.1 Acoustic performance of three formant patterns of [ai]

According to the actual performance of 102 speakers, the formants of [ai] can be summarized into three types (see Fig. 1a- Fig. 1c)

Figs. 1a-1c show that there are three formant trends of [ai] in Mandarin Chinese. ① Fig. 1a, [ai] consists of only one part, that is, the transition segment between the two target vowels [a] and [i], hereafter referred to as [ai] 1; ② Fig. 1b, [ai] consists of two parts, that is, the stable segment of [a] + the transition segment of [ai], hereafter referred to as [ai] 2; ③ Fig. 1c, the formant of [ai] is stable and close to a straight line, it is inclined to be perceived as the monophthong [æ], hereafter referred to as [ai] 3.

In order to represent the tendency of monophthongization of [ai] 3 more intuitively, we put [ai] 1 and [ai] 3 in the same formant chart to see their contrastive performance. At the same time, the average values of the three peripheral vowels [a], [i], and [u] (the average values for males and females were calculated separately) are also shown in Fig. 2 as reference points for the acoustic position.

Fig. 2 shows that [ai] 1 has a dynamic trajectory from the main vowel [a] to the ending [i] for both males and females. The ending [i] of [ai] 1 doesn't arrive in the position of the monophthong [i] (a phenomenon known as undershoot). The data distribution of [ai] 3 is relatively concentrated, and the sound value is closer to monophthong [æ] from the relative position.

2) The classification mainly refers to the two standards of acoustics and perception. Regarding the determination of the stable section of the diphthong, the acoustic formant curves (F1, F2 and F3) are relatively stable, and the individual monophthong sections can be judged based on perception. Although the frequencies of the stable sections are not completely stable (there will always be some degree of fluctuation), compared with the real transition section, the fluctuation doesn't manifest itself as a regular rise or fall, but an irregular high and low change, and the magnitude of change is much smaller than that of the real transition section.

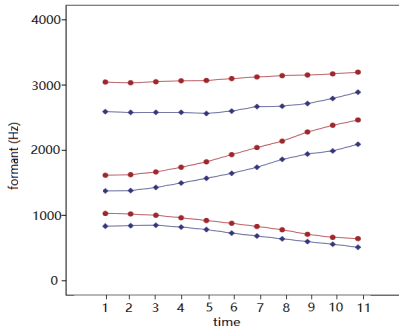


Fig. 1a

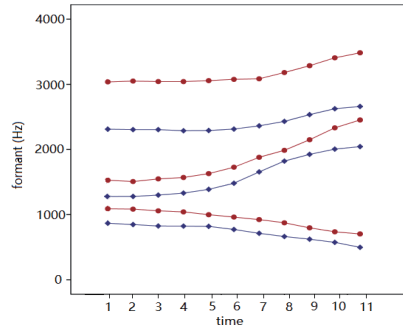


Fig. 1b

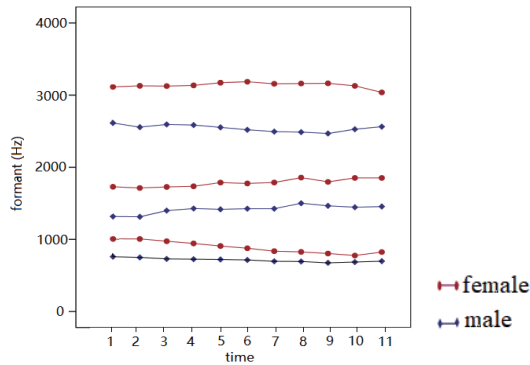


Fig. 1c

Fig. 1. Three kinds of formant trends of [ai]

From the nature of the formants, [ai] can be classified into two categories: ① of diphthongs [ai] 1 and [ai] 2, [ai] 1 has a higher frequency of occurrence and constitutes the main part, while [ai] 2 appears less frequently, so it is a variant; ② monophthong [ai] 3 has changed to the monophthong [æ], and can be regarded as an independent category.

3.1.2 Parameters of F2 in the three formant patterns of [ai]

Lehiste and Peterson (1961: 268–277), Gay (1968: 1570–1573, 1970: 65–88), Kent and Moll (1972:278–300), and Manrique (1979:194–206) investigated the change rate of the second formant of the diphthong under different

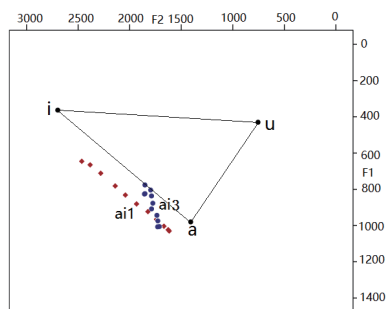


Fig. 2a. Female

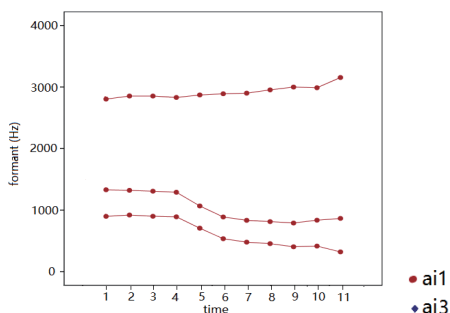


Fig. 2b. Male

Fig. 2 The formant chart of [ai]

Table 2. Parameters of F2 in three formant patterns of [ai]³⁾

	[ai] 1	[ai] 2	[ai] 3
Frequency span($\triangle$ Hz)	782	846	130
Total duration(ms)	294	303	304
Overall slope(Hz/ms)	2.66	2.79	0.43
Duration of transition(ms)	294	165	
Slope of transition (Hz/ms)	2.66	4.49	
Duration of [a] (ms)		138	
Slope of [a] (Hz/ms)		0.77	

conditions. These studies take the frequency change of F2 (acoustic level) and the tongue movement (physiological level) as two important parameters representing the characteristics of the diphthong, and confirm that the transition rate of F2 is an important acoustic parameter representing the acoustic features of the diphthong. On the basis of this, Ren's (1986) coarticulation experiment also proved that the main characteristics of the compound vowel were determined by F2, not F1 and F3. Based on this, we calculated the frequency span, duration, slope and other parameters of F2 in the three formant patterns of [ai]. The specific results are shown in Table 2.

3) Frequency span={the difference between the frequency values of the starting point and the end point among 11 measuring points}; Overall slope=Frequency span / Total duration; Slope of transition={the difference between the frequency values of the starting point and the end point

In table 2, it can be seen that for frequency span, compared with [ai] 1 and [ai] 2, [ai] 3 is the smallest, with the differences of 652Hz, 716Hz, respectively. For overall slope, [ai] 1 and [ai] 2 are relatively close, while [ai] 3 is significantly smaller than [ai] 1 and [ai] 2, with the differences of 2.23 and 2.36, respectively. For slope of transition, [ai] 2 is larger than [ai] 1, the difference between them 1.83. This is because the duration of the transition of [ai] 2 is shorter, with the tongue movement from [a] to [i] being completed in a relatively short time. By contrast, the duration of the transition of [ai] 1 is longer, with the tongue movement relatively slow. The slope of [a] in [ai] 2⁴⁾ and the overall slope of [ai] 3 are close, with a difference of 0.34. Frequency span and overall slope are two important manifestations of the monophthongization of [ai] 3.

3.2 The statistical performance of the formant patterns of [au]

3.2.1 Acoustic performance of the three formant patterns of [au]

Based on the performance of 102 speakers, the formant trends of [au] can be classified into three types (See Figs. 3a-3c).

Figures 3a-3c show the three formant trends of [au] in Mandarin Chinese. ① In Fig. 3a, [au] consists of the transition segment between the two target vowels

of the transition}/the duration of the transition. The above formulas are consistent with the measuring standard and calculating method of the “change rate of formant” proposed by Gay (1968) and He (1985). Both scholars believe that the “change rate of formant” is the most stable and important parameter in diphthong perception; the slope of [a]= {the difference between the frequency values of the starting point and the end point of [a]} / the duration of [a]; the determination of [a] and the transition are mainly based on the formant curve, intensity curve in the three-dimensional map, combined with the listening. In the following, the calculating method of each parameter in [au], [ou], and [ei] is the same.

- 4) W. Ainsworth (1972: 648–651) showed that the recognition rate would decrease when the vowel was too long (500 milliseconds) or too short (50 milliseconds) using synthetic vowels. Liang (1965:20–23) found that when the vowel length was more than 60 milliseconds, the recognition rate would remain above 95%. When the syllable length was less than 60 milliseconds, the recognition rate would decrease with the shortening of the vowel length. In this paper, the duration of [a] in [ai] 2 is 138 ms, which greatly exceeds the shortest length of auditory recognition of 50 ms–60 ms. Moreover, the slope of [a] is 0.77, which is close to the overall slope of [ai] 3, but is much smaller than the slope of the transition of [ai] 2. Therefore, both the duration and the slope can fully prove that the steady segment of [a] in [ai] 2 is real at the acoustic level, which is consistent with the results for perception. The criterion for determining the steady segment of [a] in [au] 2 is the same.

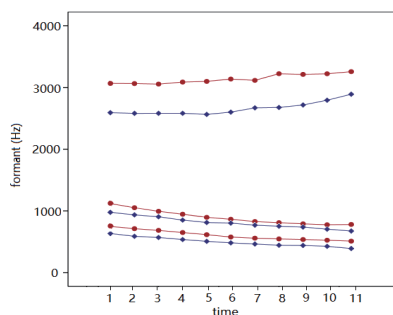


Fig. 3a

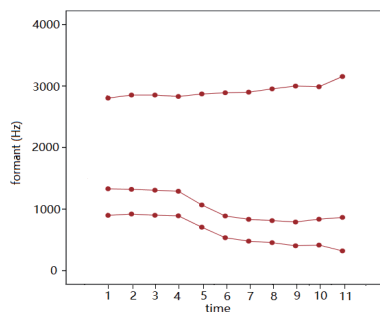


Fig. 3b

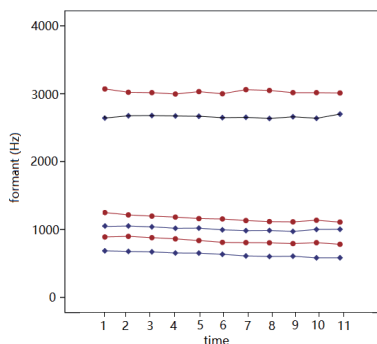


Fig. 3c

Fig. 3. Three kinds of formant trends of [au]

[a] and [u], hereafter referred to as [au] 1. ② In Fig. 3b, [au] consists of two parts, that is, the stable segment of [a] and the transition segment of [au], hereafter referred to as [au] 2. ③ In Fig. 3c, the formant of [au] is stable and close to a straight line, and it is inclined to be perceived as the monophthong [ɔ], hereafter referred to as [au] 3.

Fig. 4 shows that for both males and females, [au] 1 behaves as dynamic trajectories of tongue movement from the main vowel [a] to the ending [u]. And the main vowel [a] and the ending [u] of [au] 1 don't arrive in the positions of their respective monophthongs [a] and [u] (undershoot). The data distribution of [ai] 3 is relatively concentrated, and the sound value is closer to monophthong [ɔ] relative to the peripheral vowels [a], [i] and [u].

From the nature of the formants, [au] can be classified into two categories. ①

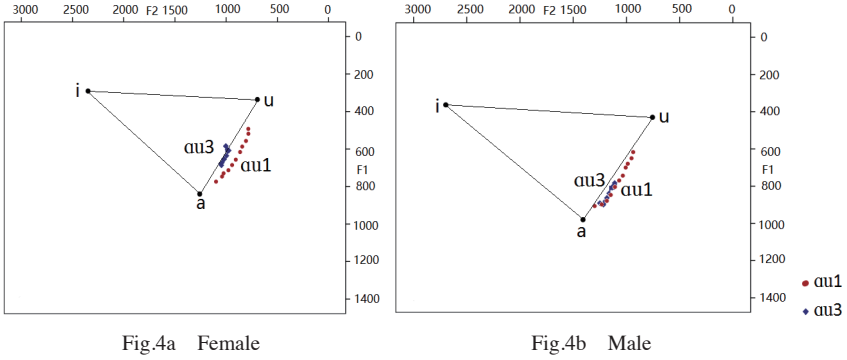


Fig. 4 The formant chart of [au]

Diphthong. Both [au] 1 and [au] 2 are dynamic diphthongs. But because the frequency of [au] 2 is relatively low, it can be regarded as a variant of [au] 1. ② Monophthong. [au] 3 has changed into monophthong [ɔ], which can be regarded as an independent category.

3.2.2 Parameters of F2 in the three formant patterns of [au]

In order to observe the three formant patterns of [au] from a quantitative point of view, we calculated the performance of various parameters of F2 (the calculation method is consistent with that of [ai]). See Table 3 for the specific results.

In table 3, we can see that for frequency span, [au] 3 is the smallest, and its

Table 3. Parameters of F2 in the three formant patterns of [au]

	[au] 1	[au] 2	[au] 3
Frequency span (Δ Hz)	339	465	91
Total duration (ms)	315	430	315
Overall slope (Hz/ms)	1.08	1.08	0.29
Duration of transition (ms)	315	295	
Slope of transition (Hz/ms)	1.08	1.50	
Duration of [a] (ms)		135	
Slope of [a] (Hz/ms)		0.17	

differences between [au] 1 and [au] 2 are 248Hz and 374Hz respectively. For total duration, [au] 2 is the longest, and [au] and [au] 3 are equal. For overall slope, [au] 1 and [au] 2 are equal, and [au] 3 is significantly less than [au] 1 and [au] 2, with a difference of 0.79. For slope of transition, [au] 2 is 0.42 larger than [au] 1, the reason being the same as that of [ai] 2. The slope of [a] in [au] 2 is close to the overall slope of [au] 3, with a difference of 0.12. Frequency span and overall slope are two important manifestations of monophthongization of [au] 3.

3.3 The statistical performance of the formant patterns of [ou]

3.3.1 Acoustic performance of the two formant patterns of [ou]

According to the actual performance of 102 speakers, the formant trends can be classified into two types (See Figs. 5a–5b).

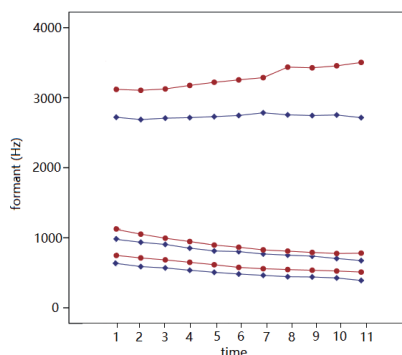


Fig. 5a

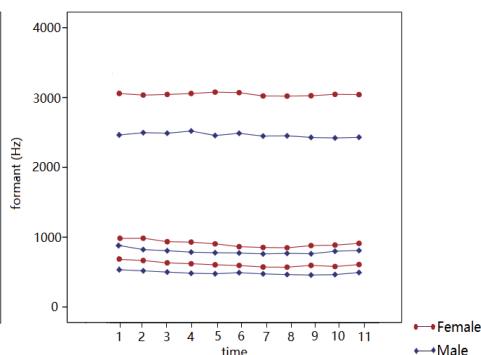


Fig. 5b

Fig. 5 Two kinds of formant trends of [ou]

Figs. 5a-5b show that there are two kinds of formant trends of [ou] in Mandarin Chinese. ① In fig. 5a, [ou] consists of a transition segment between the two target vowels [o] and [u], hereafter referred to as [ou] 1. ② In fig. 5b, the formant of [ou] is stable and close to a straight line, and it is inclined to be perceived as the monophthong [o], hereafter referred to as [ou] 2.

Fig. 6 shows that for both males and females, [ou] 1 behaves as dynamic trajectories of tongue movement from the main vowel [o] to the ending [u]. The data distribution of [ou] 2 is relatively concentrated, and the sound value is

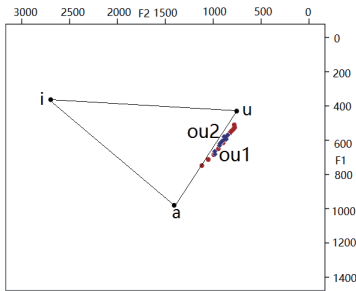


Fig.6a Female

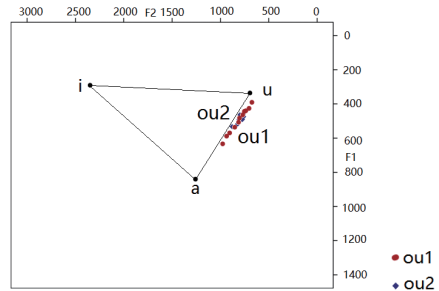


Fig.6b Male

Fig. 6 The formant chart of [ou]

closer to monophthong [o] relative to the peripheral vowels [a], [i] and [u].

3.3.2 Parameters of F2 in the two formant patterns of [ou]

The parameters of F2 in the two formant patterns of [ou] are shown in Table 4 (the calculation method is consistent with that of [ai]).

Table 4. Parameters of F2 in two formant patterns of [ou]

	[ou] 1	[ou] 2
Frequency span (Δ Hz)	320	71
Total duration (ms)	286	308
Overall slope (Hz/ms)	1.12	0.23

As can be seen in table 4, the frequency span of [ou] 1 is significantly larger than that of [ou] 2, with the difference between them being 249Hz. For total duration, [ou] 1 and [ou] 2 are similar. For overall slope, [ou] 1 is significantly larger than [ou] 2, with the difference between them being 0.89. Compared with [ou] 1, the frequency span and total duration show that [ou] 2 has become a monophthong.

3.4 The statistical performance of the formant pattern of [ei]

There was only one formant pattern of [ei], which consists of a transition segment from [e] to [i]. See Fig. 7 and Fig. 8.

Fig. 7 clearly shows the dynamic trajectories of F1, F2 and F3 in [ei]. The

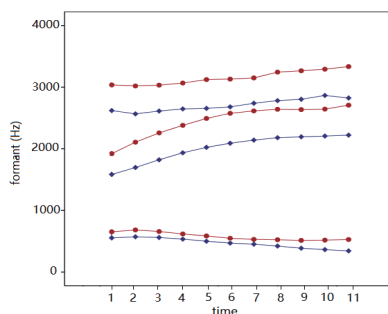


Fig. 7 The formant trend of [ei]

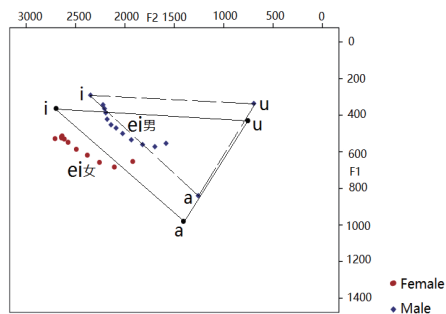


Fig. 8 The formant chart of [ei]

curve of F1 declines gradually, while by contrast, the curves of F2 and F3 rise gradually, and the increase of F2 is more obvious. Fig. 8 shows that for both males and females, the formant trajectory of [ei] changes from low to high and from back to front, corresponding to the movement of the tongue position from the main vowel [e] to the ending [i]. The frequency span of F2 of [ei] is 713Hz, the total duration is 286ms, and the overall slope is 2.49.

For frequency span, [ai] is larger than [au] and [ei] is larger than [ou]. It suggests that in the vowel triangle of Mandarin Chinese, the distribution space of the front vowels is larger than that of the back vowels. Of course, this has a close causal relationship with the physiological conditions of articulation, that is, the anterior cavity is large, and the posterior cavity is small. Martinet (1952:1–32) held that since there is asymmetry in the articulation space in the upper part of the larynx, the articulation space is used more for the front vowels than the back vowels.

4. Distribution of the Heterogeneous Formant Patterns of Falling Diphthongs [ai], [au] and [ou] for Different Groups of People

On the basis of acoustic analysis, we investigated the distribution of heterogeneous formant patterns of falling diphthongs [ai], [au] and [ou] for different groups of people.

Table 5. Distribution of the three formant patterns of [ai] for males and females

	Male	Female	Percentage of the total
[ai] 1	33 (32%)	30 (29%)	61%
[ai] 2	6 (6%)	15 (15%)	21%
[ai] 3	7 (7%)	11 (11%)	18%

Note: [ai] 1 consists of only the transitional section from the main vowel [a] to the ending [i]; [ai] 2=the stable section of [a]+the transitional section; in [ai] 3, the formant trend of [ai] is close to a straight line and the sound value is like that of the monophthong [æ].

4.1 *Distribution of the three formant patterns of [ai] for different groups of people*

4.1.1 Distribution of the three formant patterns of [ai] for males and females

We made a statistical calculation on the distribution of the three formant patterns of [ai] for males and females. The specific results are shown in Table 5.

Table 5 shows that [ai] 1 has the highest proportion of speakers at 61%; the proportions of [ai] 2 and [ai] 3 are close to each other, 21% and 18% respectively. The proportion of [ai] 1 of males is similar to that of females, with a difference of 3%. The proportions of [ai] 2 and [ai] 3 for males are 9% and 4% lower than those of females, respectively.

4.1.2 Distribution of the three formant patterns of [ai] for different age groups

Because the number of the speakers in each age group is not balanced, with the number of speakers ages 20–29 being the most (57), and the number of speakers over 50 years old being the least (8). In order to eliminate the effect caused by the imbalance of the number of speakers on the results, we have made some adjustments to the calculation method of percentage value. The adjusted results can scientifically reflect the distribution and trend of each formant pattern. Specifically, we took each age group as a whole, that is 100%, and then calculated the percentage values of the three formant patterns of [ai] for each age group separately. The calculation method for [au] and [ou] was the same as that of [ai].

Table 6 shows that [ai] 1 appears in all five age groups from under 20 to over 50. With an increase in age, the percentages of 40–49 and over 50 are higher

Table 6. Distribution of the three formant patterns of [ai] for different age groups

	Under 20	20–29	30–39	40–49	Over 50
[ai] 1	4 (33%)	33 (58%)	5 (50%)	14 (93%)	7 (88%)
[ai] 2	2 (17%)	17 (30%)	1 (10%)	0 (0%)	1 (12%)
[ai] 3	6 (50%)	7 (12%)	4 (40%)	1 (7%)	0 (0%)

Note: The numbers in the table are the actual number of people for each formant pattern, and the numbers in parentheses are their corresponding percentage values. Table 8 and Table 10 are the same.

than those for under 20, with the difference ranging from 55% to 60%; [ai] 2 appears in four of the age groups, namely under 20, 30–39, 40–49 and over 50; and [ai] 3 appears in four age groups, from under 20 to 40–49, with the highest value in under 20, and the lowest value in 40–49. The synchronic variation of [ai] in the different age groups presents the following distribution tendency: [ai] 1 mainly occurs in the middle-aged and elderly population over 40, [ai] 3 (monophthongization) is mainly in the youth population under 39, [ai] 2 has an uneven distribution in all age groups (except 40–49).

4.2 Distribution of the three formant patterns of [au] for different groups of people

4.2.1 Distribution of the three formant patterns of [au] for males and females

We have calculated the distribution of the three formant patterns for males and females respectively. The specific results are shown in Table 7.

Table 7. Distribution of the three formant patterns of [au] for males and females

	Male	Female	Percentage of the total
[au] 1	33 (32%)	32 (31%)	63%
[au] 2	0 (0%)	1 (1%)	1%
[au] 3	15 (15%)	21 (21%)	36%

Note: [au] 1 consists of only the transitional section from the main vowel [a] to the ending [u]; [au] 2=the stable section of [a]+the transitional section; for [au] 3, the formant trend of [au] is close to a straight line and the sound value is like that of the monophthong [ɔ].

Table 7 shows that the percentage of [au] 1 is the highest (63%), which occupies the dominant position; the percentage of [au] 3 is the second (36%); and the percentage of [au] 2 is the lowest (only 1%). For [au] 1, the number of males and females is close to each other; for [au] 3, the proportion of females is 6% higher than that of males.

4.2.2 Distribution of the three formant patterns of [au] for different age groups

Table 8. Distribution of the three formant patterns of [au] for different age groups

	Under 20	20–29	30–39	40–49	Over 50
[au] 1	7 (58%)	34 (60%)	5 (50%)	12 (80%)	7 (88%)
[au] 2	0 (0%)	0 (0%)	1 (10%)	0 (0%)	0 (0%)
[au] 3	5 (42%)	23 (40%)	4 (40%)	3 (20%)	1 (12%)

Table 8 shows that [au] 1 appears in all five age groups from under 20 to over 50. Compared with 40–49 and over 50, the percentage values for under 20, 20–29 and 30–39 are 20%–38% lower; [au] 2 only appears in the age group of 30–39; and [au] 3 occurs in all five age groups. Among them, the percentage values of under 20, 20–29 and 30–39 are relatively higher, while the percentage values of 40–49 and 50–59 are lower, with a difference of 20%–30%. The distribution of synchronic variation in all age groups presents the following regularities: the proportion of [au] 1 is higher in middle-aged and elderly people over 40, the percentage of [au] 3 (monophthongization) is higher in age groups under 39, and the frequency of occurrence of [au] 2 is very low.

4.3 *Distribution of the two formant patterns of [ou] for different groups of people*

4.3.1 Distribution of the two formant patterns of [ou] for males and females

Table 9 shows that the percentage of [ou] 1 is 71%, and the percentage of [ou] 2 is 29%. [ou] 1 is significantly higher than [ou] 2, with a difference of 42%. For [ou] 1, the percentage of females is 9% higher than that of males, while for [ou] 2, the percentage of males is slightly higher than that of females, with a difference of 3%.

Table 9. Distribution of the two formant patterns of [ou] for males and females

	Male	Female	Percentage of the total
[ou] 1	32 (31%)	41 (40%)	71%
[ou] 2	16 (16%)	13 (13%)	29%

Note: [ou] 1 consists of only the transitional section from the main vowel [o] to the ending [u]; for [ou] 2, the formant trend of [ou] is close to a straight line and the sound value is like that of the monophthong [o].

4.3.2 Distribution of the two formant patterns of [ou] for different age groups

Table 10. Distribution of the two formant patterns of [ou] for different age groups

	Under 20	20–29	30–39	40–49	Over 50
[ou] 1	6 (50%)	37 (65%)	8 (80%)	14 (93%)	8 (100%)
[ou] 2	6 (50%)	20 (35%)	2 (20%)	1 (7%)	0 (0%)

Table 10 shows that [ou] 1 appears in all five age groups from under 20 to over 50, and the percentage values gradually increase with the increase of age level, with the lowest value (under 20) and the highest value (over 50) differing by 50%; [ou] 2 only appears in four age groups, from under 20 to 40–49. Contrary to [ou] 1, the percentage values decrease gradually with the increase of age level, with the lowest value (40–49) and the highest value (under 20) differing by 43%. The above results indicate that the percentage of [ou] 1 gradually rises with the increase of age level, reaching 100% in the age group of over 50; In contrast, the percentage of [ou] 2 (monophthongization) falls with the increase of age level, and reaching 0% in the age group of over 50. From the overall percentage, [ou] 2 is significantly less than that of [ou] 1.

5. Conclusion

We conducted a statistical analysis of large samples on the dynamic formant trajectories of the falling diphthongs [ai], [au] and [ei] in Mandarin Chinese pronounced by 102 Beijing native speakers in Chinese Speech Database. From the quantitative point of view, their statistical characteristics are provided a more accurate description, when in turn provides a qualitative classification ac-

cording to their quantitative performance. The results show that the performance of the falling diphthongs in Mandarin Chinese is not uniform and homogeneous, but that there are different degrees of variation. The complexity of different vowels is different. Specifically, the performance of [ai] and [au] is the most complex, with each including three formant patterns; [ou] is moderately complex, including two formant patterns; and the performance of [ei] is the purest, with only one formant pattern.

Previous scholars have different opinions on whether there are stable segments in falling diphthongs of Chinese. This study suggests that the stable segments only appear in [ai] and [au], and only appear in a few speakers, so the proportion is relatively low. Specifically, the number of [ai] 2 is 19, accounting for 19% of the total number (102). The number of [au] 2 is one, accounting for 1% of the total number. Therefore, compared with the representative bodies of formant patterns of diphthongs consisting solely of the transition segments, they can only be regarded as marginal variants. This is similar to the results of the study made by Ren (1986), who held that there are steady segments in the main vowels /a/ of the falling diphthongs [ai] and [au] in Chinese, but they are unstable, arbitrary and not always present. However, due to the small number of speakers in Ren's experiment (only six male Beijing native speakers), no clear quantitative data were given on the degree and performance of this instability and non-mandatory. It can be said that from the perspective of large sample statistics, since we provided specific quantitative data on the frequency of the occurrence of the stable segments of /a/ in [ai] and [au] and other related parameters, this can be regarded as an improvement on and deepening of previous studies.

Wu Zongji and Lin Maocan (1989) classified diphthongs into two categories, true and pseudo, according to their nature. A true diphthong means that each of the two target vowels has a longer steady segment on the spectrum, and the transition segment between them appears to be shorter. A pseudo diphthong has two target positions. Their formant trajectories are continuously moving, and there are few stable segments. Only the main vowel of the rising diphthong may have a stable segment. For example, the diphthongs in Wa and Ali dialects in Tibetan are true diphthongs, and the diphthongs in Mandarin Chinese are pseudo diphthongs. From the statistical results of this paper, for [ai] and [au],

the pseudo diphthongs dominate the mainstream ([ai] 1 and [ai] 2, and [au] 1 and [au] 2), but they also contain a small amount of true diphthong variants; [ou] ([ou] 1) and [ei] are all pseudo diphthongs. Overall, the falling diphthongs in Mandarin Chinese (except for the monophthongized variations) give priority to pseudo diphthongs, accompanied by a small amount of true diphthongs, and the lower the tongue position of the main vowel, the greater the movement of the formant trajectory ([ai]), the higher the frequency of the real diphthongs. Of course, the conclusion of this paper is only based on Mandarin Chinese and require further proof for other languages and dialects.

On the whole, this paper summarizes the heterogeneity and categories of the formant patterns of [ai], [au], [ou] and [ei] in Mandarin Chinese using the statistical method. By investigating their distribution for different genders and different age groups, we revealed the characteristics and regularities of the synchronic variations. In addition, because the number of speakers in each age group is unbalanced (more young people aged 20–29, fewer middle-aged and elderly people aged over 50), whether this synchronic variation implies the direction and trend of diachronic change should be further explored by increasing and balancing the number of speakers for each age level in the future.

Acknowledgements

This study was supported by a grant from the Institute of Economic Research, Aoyama Gakuin University in the academic year of 2018. Sincere thanks to Dr. Jonathan Sieg for the proofreading of this article.

References

- Ainsworth, William. 1972. Duration as a cue in the recognition of Synthetic Vowels. *The Journal of Acoustic Society of America* 51: 648–651.
- Cao, Jianfen. 1994. Putonghua yunmu [ao] yu [ou] de shipin xiebian duibi fenxi [A contrastive analysis of coordinated variation of vowels [ao] and [ou] in Standard Mandarin]. *Yingyong Shengxue [Journal of Applied Acoustics]* 2: 21–27.
- Cao, Jianfen, Shun'an Yang. 1984. Beijinghua fuhe yuanyin de shiyan yanjiu [The experimental study on the diphthongs in Beijing dialect]. *Zhongguo Yuwen [Studies of the Chinese Language]* 6: 426–433.
- Eric, Zee. 2001. The Phonetic value of the vowels, diphthongs, and triphthongs in Beijing Mandarin. *The Proceeding of 5th National Conference on Modern Phonetics*, eds. by Lianghong Cai, Tongchun Zhou & Jianhua Tao, 54–60. Beijing: Tsinghua University Press.
- Gay, Thomas. 1968. Effect of speaking rate on diphthong formant movement. *The Journal of*

- Acoustic Society of America* 44 (6):1570–1573.
- Gay, Thomas. 1970. A perceptual study of American English diphthongs. *Language and Speech* 13: 65–88.
- He, Ningji. 1985. Beijinghua erhe yuanyin ganzhi zhong de shijian yinsu [Time factors of perception of diphthongs in Beijing dialect]. *Beijing Yuyin Shiyuanlu [The Experimental Study on the Sound of Beijing Dialect]*, eds. by Tao Lin & Lijia Wang, 196–224. Beijing: Peking University Press.
- Hobrook, Anthony, & Grant Faibanke. 1962. Diphthong, formants and their movements, *Journal of Speech, Language, and Hearing Research* 5 (1): 38–58.
- Ren, Hongmo 1986. On the acoustic structure of diphthongal syllables. *UCLA Working Papers in Phonetics* 65:1–100.
- Kent, Raymond. D., & K.L. Moll. 1972. Tongues body articulation during vowel and diphthong gesture. *Folia phoniatica et Logopeadica* 24: 278–300.
- Lehiste, Iise, & Gordon E. Peterson. 1961. Transitions, glides, and diphthongs. *The Journal of Acoustic Society of America* 33 (3):268–277.
- Liang, Zhi'an. 1965. Danyuanyin Shibielü yu yinjie changduan de guanxi [Vowel identification in relation to syllable duration]. *Shengxue Xuebao [Acta Acustica]* 1965. 1: 20–23.
- Manrique, André. 1979. Acoustic analysis of the Spanish diphthongs, *Phonetica*, 36 (3): 194–206.
- Martinet, André. 1952. Function, structure, and sound change. *Word* 8: 1–32.
- Wu, Zongji, & Jianfen Cao. 1979. Shiyán yuyinxue zhishi jianghua [Introduction to Experimental Phonetics]. *Zhongguo Yuwen [Studies of the Chinese Language]* 1979. 4: 314–320.
- Wu, Zongji, & Maocan Lin. 1989. *Shiyán Yuyinxue Gaiyao [Summary of Experimental Phonetics]*. Beijing: Higher Education Press.
- Yang, Shun'an. 1986. Fuhe yuanyin de zhishushi dongtai moxing Jiqi zai hecheng zhong de yingyong [Exponential dynamic model of diphthongs and Its application in speech synthesis]. *Yuyan Yanjiu [Studies in Language and Linguistics]* 1986. 2: 1–8.