康复学报
Rehabilitation Medicine

·临床论著·

Acoustic Analysis of Mandarin Chinese Vowels Produced by Young Adults

WANG Zhenni[1], CHEN Yang[2], NG Manwa L.[3], YAO Liqun[4*], ZHANG Weiming[5*]

[1] *Shanghai Ruijin Rehabilitation Hospital, Shanghai 200125, China;*
[2] *Duquesne University, Pittsburgh, Pennsylvania PA15282, USA;*
[3] *Faculty of Education, University of Hong Kong, Hong Kong 999077, China;*
[4] *Nursing College, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian 350122, China;*
[5] *Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China*
*Correspondence: YAO Liqun, E-mail: yaoliqunpt@163.com; ZHANG Weiming, E-mail: zwm40397@rjh.com.cn

Received 2020-03-17; accepted 2020-04-28
Fundation: Supported by the Public Health Bureau of Shanghai Municipal Huangpu District (HKQ201808) and the University Research Funding, Fujian University of Traditional Chinese Medicine (X2017019)
DOI:10.3724/SP.J.1329.2020.03004
开放科学(资源服务)标识码(OSID):

ABSTRACT **Objectives:** Acoustic analysis is a kind of objective assessment of speech-sound which can offer a relatively simple and visual way to examine production of a vowel. Mandarin Chinese is a tone language in which the same phonetic segment carries a different meaning when produced at different lexical tones. Previous studies have examined American English vowels produced by native adult speakers. The aim of the present study was to establish the vowel formant space by examining the formant frequencies associated with Mandarin vowels produced by Chinese young adults aged 23 to 33 years old and identify the characteristics of vowels in Mandarin Chinese at 4 different lexical tones. **Methods:** Acoustic signals of the six Mandarin vowels (/a/, /ɔ/, /e/, /i/, /u/, /y/) produced by native young adult speakers of Mandarin Chinese (n=11) were recorded, and the participants were instructed to produce all the speech samples using a comfortable loudness level and speech rate. The speech samples were produced was randomly. The first two formants were analyzed by using a professional acoustic measurement system (Multi-Speech, KayPentax, USA). Multi-Speech provides a time-domain waveform and a frequency-domain wide-band spectrogram (filter bandwidth=300 Hz). **Results:** The results showed that, generally speaking, the vowel /a/ exhibited the highest mean F1 value whereas vowel /i/ had the lowest F1. The vowel /i/ showed the highest F2 while /u/ showed the lowest F2 value. The vowel of /ɔ/ and /e/ with different four tones in F1 and F2 had significant differences (P<0.05). However, no significant differences (P>0.05) in F1 and F2 were observed in the other vowels across tones, like /a/, /i/, /u/, /y/. **Conclusion:** An acoustic vowel space was established in which /i/, /u/, and /a/ were corner vowels, and /ɔ/ and /e/ were found centrally, contributing to the generally triangular vowel space associated with these six core vowels of Mandarin Chinese. The investigation of the first two formant frequencies of vowels across tones shows significant differences were found in the vowels /ɔ/ and /e/ with different tones in F1 and F2. When /ɔ/ was produced with level tone, the tongue was more advanced and depressed than with other three tones. For /e/ with rising tone the tongue was more retracted, compared to the tongue for /e/ with dipping tone.

KEY WORDS Chinese Mandarin vowel; formant frequency; vowel space; acoustic analysis; young adults.

1 Introduction

Acoustic analysis offers a relatively simple and objective way to visualize and examine production of a vowel (monophthong or diphthong), often by means of a waveform and a spectrogram. While the former depicts the changes of sound pressure in the time domain, the latter allows examination of the related formant (frequency) characteristics of the speech sound [1]. A vowel produced in a carrier phrase can be examined

引用格式:王璨旎,陈旸,吴民华,等. 成人普通话元音的声学分析[J]. 康复学报,2020,30(3):183-191.
WANG Z N,CHEN Y,NG M L,et al. Acoustic analysis of Mandarin Chinese vowels produced by young adults [J]. Rehabilitation Medicine,2020,30(3):183-191.
DOI:10.3724/SP.J.1329.2020.03004

in detail based on a wide-band spectrogram derived using the well-known Fourier transformation, as well as through the use of Linear Prediction Coding (LPC). A major advantage associated with formant or spectral analysis is it allows to better understand the frequency content and thus resonance characteristics of the vocal tract during production of a particular speech sound. As such, information obtained from spectrally analyzing a speech sound provides insight into the articulation of that sound. Articulation refers to the physical relationship among articulators during production of speech sound. Speech articulators include tip, blade and dorsum of the tongue, jaw, lips, soft palate, as well as pharyngeal cavity during speech production. All of them play a significant role in determining the resonance characteristics of the vocal tract. Simply put, knowledge of formant frequencies allows us to understand articulation during production of the sound and it allows us to know the details of vocal tract configuration, including positioning of the tongue, lips, jaw, etc. during sound production. This is achieved by merely acoustically analyzing the acoustic signal of the sound, which is considered a non-invasive and simple analytic method.

Formant measurement can be used to quantify production of speech sounds in connected speech, as an attempt to better understand articulation of speech sounds, especially for vowels. According to basic speech acoustics, formant frequencies are generally inversely proportional to vocal tract length, and can be derived from a signal spectral cross section through the steady-state portion of the acoustic signal [2-4]. The first two formants are frequently said to correspond to the high/low and front/back dimensions respectively, which have traditionally, though not entirely accurately, been associated with the contraction and position of the tongue [5]. Thus, the first formant (F1) is high for a low vowel (such as /a/) and low for a high vowel (such as /i/ or /u/). For the second formant (F2), it is high for a front vowel (such as /i/) and low for a back vowel (such as /u/) [6].

Previous studies have acoustically examined American English vowels produced by native adult speakers [7-9]. The most widely referred study relating to the American English vowels was reported by Peterson and Barney that dated back to 1952 [2]. Based on ten English vowels produced by 61 adult males and females and 15 children, they reported the average F0 and formants values. The study was later replicated and extended by Hillenbrand and colleagues [7], in which two more vowels and a fourth formant were included in the analyses of data obtained from 45 men, 48 women, and 46 children. Both studies revealed that articulation of vowels can easily be indexed acoustically by the first three formant frequencies (F1, F2, & F3), although some researchers suggested that F1 and F2 are sufficient for vowel classification. According to Raphael, Borden, and Harris [10], F1 and F2 are sufficient in distinguishing English vowels, and F3 can be used to reflect tongue tip configuration. Generally speaking, a lower F3 is associated with a retroflexed tongue, while a higher F3 reflects a flat tongue tip configuration. In other words, knowledge of F0 or higher formants probably adds no additional information to vowel articulation [11-14].

In a recent study of how body position may affect production of quadrilateral point vowels of English produced by 27 male and female native English speakers, Vorperian and colleagues [15] correlated fundamental frequency (F0), the first four formant frequencies (F1-F4) and a number of volumetric measures derived by using acoustic pharyngometry. They found that body posture did not seem to affect major vowel formants.

More recently, F0 and formant frequencies have been used in studies of speech production by children and adults across different languages, including Japanese [16], Russian [17], Swedish [18], Korean [19], Arabic [20], Mandarin [21], and Cantonese [22]. In an Arabic study [20] participants sustained the six steady-state Arabic vowels (/i:/, /e:/, /a:/, /o:/, and /u:/). In the Swedish study by White [18], the subjects were asked to sustain the vowels /e:/, /u:/, /i:/ and /ɔ/. As mentioned, most researchers [14-15, 21] focused on sustained vowels and syllables, instead of sentences. There is a paucity of information on formant frequencies of vowels produced in sentences. In addition, these studies failed to report significant differences in formant frequencies across different vowels. Ting and Zourmand [23] obtained F0 and the first two formant frequencies of vowels produced by 360 Malay children aged between 7 and 12 years. Their results showed a nonsystematic decrement in formant frequencies with age, implying a lengthening of resonating cavity, or the vocal tract. There was a significant difference in the first three formants between different races reported by Mayo and Grant [24].

Mandarin Chinese is a tone language in which, unlike a non-tonal language such as English, the same phonetic segment carries a different meaning when produced at different lexical tones. There are four lexical tones in Mandarin Chinese: level tone (Tone 1), rising tone (Tone 2), dipping tone (Tone 3), and falling tone (Tone 4) [25]. Tones of Mandarin Chinese have been studied by many researchers. For example, Duanmu [26] reported that Tone 1 was associated with a flat F0 contour, and Tone 2 with a rising F0 contour.

For example, F1 acoustic parameters have developmental and gender changes in vowel production in Mandarin-speaking children as found by Chen et al. [21]. Formant frequency analysis of Chinese vowels were compared between tongue carcinoma patients (before surgery and 3 months, 9 months, and 2 years after surgery) and a control group in Liang [27]. However, a

few studies have established vowel space and analyzed the position of articulators in Chinese Mandarin vowels across tones. Unfortunately, so far, there is no documented formant frequency of Chinese Mandarin vowels across tones for Chinese young adults.

The present study was a preliminary investigation of the six Mandarin Chinese monophthong vowels produced by native young adult speakers of Mandarin Chinese. The first two formant frequencies were obtained from the vowels, based on which by formant space associated with Mandarin vowels was established. The Chinese Mandarin vowels consist of six monophthongs: /a/, /ɔ/, /e/, /i/, /u/, and /y/. In addition, F1 and F2 values associated with all six Chinese vowels were compared across the four lexical tones of Mandarin.

2 Methods

2.1 Participants

A total of 11 native speakers of Mandarin Chinese (five males, six females) who were international students at Duquesne University participated in the study. They were aged between 23 and 34 years (M=25.64 years, SD=3.67 years). They were included only when all the inclusion criteria were met: (1) they were born in and brought up in the mainland China, and had lived in the mainland China until at least 20 years of age, and (2) they were holders of the Certificate of Putonghua Proficiency Test. The last criterion ensured that all participants spoke native standard Mandarin Chinese. All participants with a history of speech, language, or hearing problem were excluded. Informed consent was obtained from each participant prior to the experiment.

2.2 Data collection

Following a similar research protocol as Hillenbrand et al. [7], the six Mandarin vowels /a/, /ɔ/, /e/, /i/, /u/, /y/ produced in a consonant-vowel (CV) syllable were examined. Each of the syllables was produced using four different tones (T1: level tone, T2: rising tone, T3: dipping tone and T4: falling tone), yielding 24 distinctive yet meaningful Chinese words. During the recording, each CV syllable was embedded in a carrier phrase/ðISIZ__/ ("这是__") (meaning "This is___."), in order to maintain naturalness of productions. To eliminate possible order effect, the order at which the stimuli were produced was randomized. The participants were instructed to produce all the speech samples using a comfortable loudness level and speech rate. Upon completion of recording, a total of 264 (11 participants × 6 phrases × 4 tones) speech samples were recorded.

Before the formal recording took place, each participant was provided with a brief practice period in order to familiarize themselves with the recording materials and recording environment, as well as to warm up their voices. If an error occurred during recording, the stimulus was produced and recorded again. This process was repeated until an accurate and precise pronunciation was achieved. The recording took place in a sound attenuated room located in the Speech Perception/Production & Innovative Technology Lab (SPPIT) of Duquesne University with the background noise controlled below 20 dB. Audio recording was done using a high-quality condenser microphone which was placed approximately 10 cm in front of participant's mouth [28] via a professional-grade external sound card (PreUSB, M-Audio, USA). Each recorded speech sample was digitized and stored on a Lenovo E450 laptop computer for later analyses.

2.3 Acoustic analysis

Measurement of F0, F1 and F2 was obtained by using a professional acoustic measurement system (Multi-Speech, KayPentax, USA). Multi-Speech provides a time-domain waveform and a frequency-domain wideband spectrogram (filter bandwidth=300 Hz). F0 was determined by counting the number of single vertical lines per time unit (i.e., cycles per second). Also, the frequencies and relative distinctions of the first two formants (F1 and F2) were darker, rather than dim concentrations of energy. The cursor indicating F1 and F2 was cross-checked on the LPC algorithm, which found two peak points on the waveform that were equivalent to the first two formats.

2.4 Statistical analysis

Microsoft Excel was used to collect all data. All statistical analyses were performed using SPSS 16.0, by Descriptive Statistics, to get the mean and standard deviation of F1 and F2 of Mandarin vowels. The inter-rater and intra-rater reliability measurements from the speech samples were examined by a *Pearson* correlation test. A preset value of P=0.05 was used to determine for statistically significance.

2.5 Reliability

As human judgment is involved in the extraction of formant frequencies, human bias may exist. To ensure minimum human bias and measurement validity, speech samples of F1 and F2 values were re-analyzed by another investigator to obtain inter-rater reliability. The first and second authors calculated the two measurements using spectrograms and the LPC algorithm. The average absolute error for F1 and F2 inter-rater measurements were 12.59 Hz and 15.68 Hz, respectively. Results of the *Pearson* correlation tests for F1 and F2 were 0.99 (P<0.01) and 0.99 (P<0.01), respectively, which indicated significantly high inter-judge reliability of the two measurements by the two authors.

In addition, all speech samples of F1 and F2 values were obtained a second time by the first author in or-

der to obtain intra-rater reliability. The first author calculated the first and second measurements, which were then used to calculate intra-rater reliability. The average absolute error for F1 and F2 intra-rater measurements were 7.83 Hz and 29.11 Hz, respectively. Results of *Pearson* product moment correlation tests for F1 and F2 were 0.88 (P<0.01) and 0.90 (P<0.01), respectively, indicating significantly high inter-judge reliability of the two measurements by the first author. Both inter- and intra-rater reliability indicated that the measurements obtained were reliable and consistent, thus high reliability.

3 Results

3.1 F1 and F2 of the six vowels across all subjects

Average F1 and F2 values associated with different vowels produced by all participants are shown in Figure 1. Each data point represents the F1 and F2 values of each vowel across different tones. It can be observed from Figure 1 that the vowel space of the Mandarin Chinese resembles a triangle, which is distinctive from the quadrilateral shape of English vowels' formant frequencies. Figure 1 indicates that /i/, /u/, and /a/ were corner vowels.

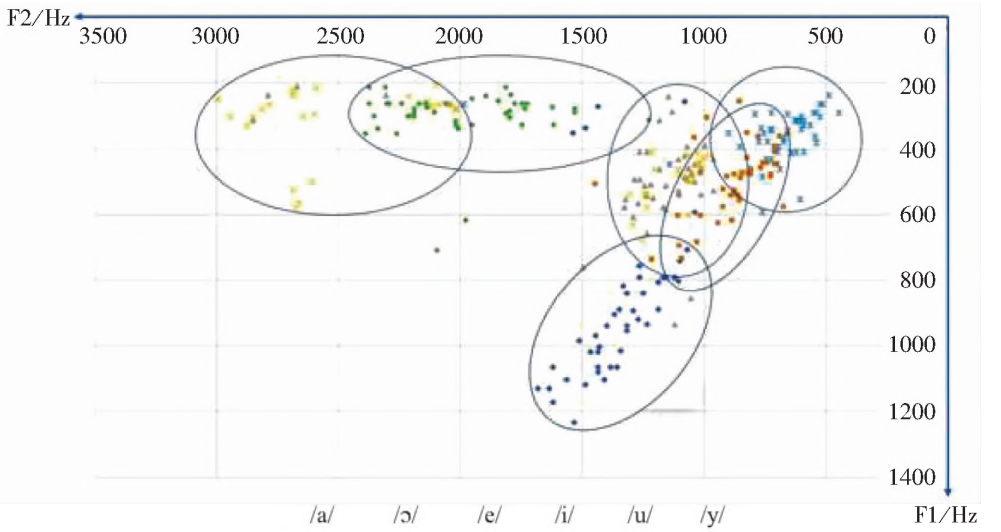

Figure 1 The F2–F1 vowel space containing the vowels across tones produced by all participants was plotted based on the value of F1 and F2 frequencies

3.2 Acoustic analysis of six vowels produced by male speakers

The mean, standard deviation, and range of F1 and F2 values for the six main vowels of Mandarin are showed in Table 1, which are produced by male speakers. It can be seen in Table 1 that the maximum value of F1 was /a/ [F1=(828.55±90.65) Hz], whereas the minimum was /i/ [F1=(248.30±28.15) Hz). Similarly, /i/ was the maximum F2, and /u/ was the minimum. As illustrated previously, these three special vowel sounds are found in the corners of the Mandarin triangle (Figure 2). Furthermore, the median values of both F1 and F2 belonged to the sounds /ɔ/ and /e/, which are the central part of the triangle (Figure 2).

Table 1 Mean and standard deviation values of F1 and F2 formant frequencies of Mandarin Chinese vowels produced by male speakers (n=5) Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
/a/	828.55	90.65	592–954	1 219.05	97.48	1 003–1 366
/ɔ/	487.60	87.03	355–737	830.10	144.48	658–1 213
/e/	456.15	47.45	392–601	1 072.60	76.59	976–1 267
/i/	248.30	28.15	179–280	2 196.75	200.71	1 941–2 172
/u/	330.00	44.10	237–411	729.35	315.17	448–1 988
/y/	279.15	33.23	214–329	1 785.85	144.09	1 527–2 057

3.3 Acoustic analysis of six vowels produced by female speakers

Table 2 demonstrated the mean, standard deviation, and range of F1 and F2 values for the six main vowels of Mandarin, which are produced by six female speakers. From Table 1, the maximum value of F1 was /a/

[F1=(1 034.75±96.89) Hz], whereas the minimum was /y/ [F1=(195.08±41.95) Hz]. Similarly, /i/ was the maximum F2, and /u/ was the minimum. As illustrated previously, these three special vowel sounds are found in the corners of the Mandarin triangle (Figure 2). Furthermore, the median values of both F1 and F2 belonged to the sounds /ɔ/ and /e/, which are the central part of the triangle (Figure 2).

Table 2 Mean and standard deviation values of F1 and F2 formant frequencies of Mandarin Chinese vowels produced by female speakers (n=6) Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
/a/	1 034.75	96.89	816–1 234	1 460.75	124.68	1 080–1 680
/ɔ/	544.38	94.05	362–735	921.42	166.27	698–1 448
/e/	517.25	86.51	313–675	1 194.58	76.23	1 185–1 253
/i/	282.30	41.88	214–354	2 769.58	143.12	2 593–3 184
/u/	360.58	65.26	263–486	666.50	137.78	513–1 142
/y/	195.08	41.95	214–382	2 167.88	233.01	1 223–2 389

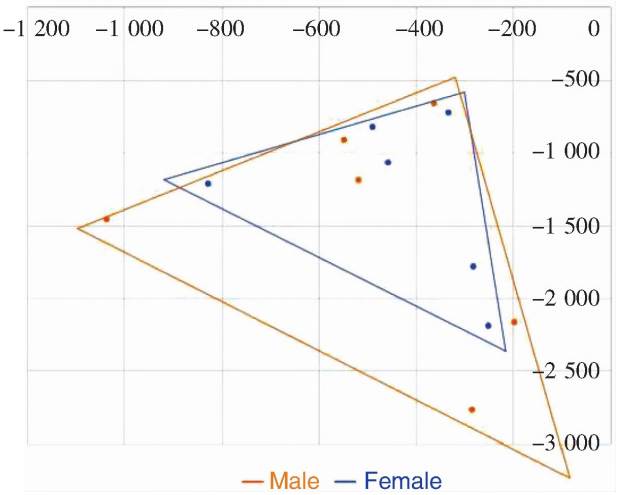

Figure 2 The F2–F1 vowel space was plotted based on the mean F1 and F2 formants frequencies of both male and female vowels

3.4 Acoustic analysis of the vowel /a/ across tones

The mean, standard deviation, and range of F1 and F2 formant frequencies for /a/ in different tones, produced by all participants, are showed in Table 3 below. Two two-way repeated-measures analyses of variance (*ANOVA*) were carried out, one for males and one for females, with both vowels and tones being the within measures. For both genders, results indicated a significant interaction effect for tones /a/ vowels (P<0.01). Subsequently, several one-way *ANOVA* were carried out. A one-way *ANOVA* test was used to identify the effect of different tones (Tone 1 *vs*. Tone 2 *vs*. Tone 3 *vs*. Tone 4) on the spectral features of F1 and F2 frequencies. The result of the one-way *ANOVA* test displayed no significant difference of the values of F1 and F2 for different tones of the vowel /a/ (F1:P=0.125, F2:P=0.989). However, results indicated in Table 2 show that the vowel /a/ for level tone (Tone 1) has the lowest mean value of F1 and F2, while T4 for falling tone (Tone 4) has the highest mean value of F1 and F2.

Table 3 Mean and standard deviation values of F1 and F2 formant frequencies of /a/ in different tones Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
Tone 1	860.36	142.90	592–1 066	1 257.27	151.89	1 037–1 514
Tone 2	950.09	125.19	708–1 105	1 332.64	153.46	1 070–1 563
Tone 3	955.82	116.08	790–1 132	1 404.73	149.27	1 168–1 634
Tone 4	997.82	151.61	740–1 234	1 408.91	178.15	1 103–1 680
F		2.03			2.24	
P		0.125			0.989	

3.5 Acoustic analysis of the vowel /ɔ/ across tones

Table 4 indicates the mean, standard deviation, and range of F1 and F2 formant frequencies for /ɔ/ in different tones produced by all participants. The effect of different tones (Tone 1 *vs*. Tone 2 *vs*. Tone 3 *vs*. Tone 4) on the F1 and F2 frequencies was assessed by a one-way *ANOVA* test, which demonstrated that the values of F1 and F2 of different tones for the vowel /ɔ/ were significantly different (F1:P=0.004<0.01, F2:P=0.007<0.01). The effect of any two different tone groups (Tone 1 *vs*. Tone 2, Tone 1 *vs*. Tone 3, Tone 1 *vs*. Tone 4, Tone 2 *vs*. Tone 3, Tone 2 *vs*. Tone 4, Tone 3 *vs*.

Tone 4) on the F1 and F2 frequencies was tested by the post hoc multiple comparisons (*LSD*) that showed that the mean value of F1 was significantly higher in the vowel /ɔ/ for level tone (Tone 1) rather than for rising tone (Tone 2) and dipping tone (Tone 3) (Tone 1 *vs*. Tone 2: P=0.001<0.01, Tone 1 *vs*. Tone 3: P=0.003<0.01). The mean value of F2 was significantly higher in level tone than in the other tones of the vowel /ɔ/ (Tone 1 *vs*. Tone 2: P<0.001, Tone 1 *vs*. Tone 3: P=0.001<0.01, Tone 1 *vs*. Tone 4: P=0.006<0.01).

Table 4 Mean and standard deviation values of F1 and F2 formant frequencies of /ɔ/ in different tones Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
Tone 1	594.45	98.10	474–737	1 035.82	191.93	806–1 448
Tone 2	466.64	90.23	355–619	787.64	88.17	658–919
Tone 3	483.82	69.72	408–625	825.91	116.40	698–1 015
Tone 4	529.36	69.79	474–696	870.27	116.95	672–1 103
F		5.22			7.315	
P		0.004*			0.007*	

Note: * P<0.01.

3.6 Acoustic analysis of the vowel /e/ across tones

The mean, standard deviation, and range of the F1 and F2 formant frequencies for /e/ in different tones produced by all participants are shown in Table 5 below. A one-way *ANOVA* test was used to determine the effect of different tones (Tone 1 *vs*. Tone 2 *vs*. Tone 3 *vs*. Tone 4) on the spectral features of F1 and F2 frequencies. The result of the one-way *ANOVA* test indicated that the value of F1 for different tones for the vowel /e/ were significantly different (F1: P=0.016<0.05). However, no significant difference of F2 among different tones of the vowel /e/ (F2: P=0.721) was noticed. The effect of any two different tone groups (Tone 1 *vs*. Tone 2, Tone 1 *vs*. Tone 3, Tone 1 *vs*. Tone 4, Tone 2 *vs*. Tone 3, Tone 2 *vs*. Tone 4, Tone 3 *vs*. Tone 4) on the F1 frequencies was tested by the post hoc multiple comparisons (*LSD*), which demonstrated the value of F1 was significantly higher in the vowel /e/ for rising tone rather than dipping tone (Tone 2 *vs*. Tone 3: P=0.005<0.01).

Table 5 Mean and standard deviation values of F1 and F2 formant frequencies of /e/ in different tones Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
Tone 1	519.18	71.01	448–656	1 166.73	96.37	1 027–1 292
Tone 2	526.09	79.43	395–675	1 123.36	88.75	1 013–1 253
Tone 3	436.27	76.26	313–576	1 125.45	107.12	976–1 273
Tone 4	476.36	51.48	411–551	1 141.00	103.87	1 004–1 306
F		3.87			0.45	
P		0.016*			0.721	

Note: * P<0.05.

3.7 Acoustic analysis of the vowel /i/ across tones

Table 5 shows the mean, standard deviation, and range of F1 and F2 formant frequencies for /i/ in different tones produced by all participants. A one-way *ANOVA* test was used to determine the effect of different tones (Tone 1 *vs*. Tone 2 *vs*. Tone 3 *vs*. Tone 4) on the F1 and F2 frequencies. The result of the one- way *ANOVA* test showed that there was no significant difference of the values of F1 and F2 for different tones of the vowel /i/ (F1: P=0.895, F2: P=0.936). However, the vowel /i/ with level tone had the lowest mean value of F1, while it also has the highest mean value of F2. The vowel /i/ with rising tone had the highest mean value for F1, while the lowest mean value for F2 (Table 6).

3.8 Acoustic analysis of the vowel /u/ across tones

The mean, standard deviation, and range of the F1 and F2 formant frequencies for /u/ in different tones produced by all participants are shown in Table 7. The effect of different tones (Tone 1 *vs*. Tone 2 *vs*. Tone 3 *vs*. Tone 4) on the values of F1 and F2 frequencies was analyzed by a one-way *ANOVA* test. The result of the *ANOVA* test demonstrated that there was no significant difference between the values of F1 and F2 for different tones of the vowel /u/ (F1: P=0.921, F2: P=0.9081). However, results in Table 7 indicated that the vowel /u/, for dripping tone, has the highest mean value for F1, while it has the lowest mean value for F2.

Table 6 Mean and standard deviation values of F1 and F2 formant frequencies of /i/ in different tones Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
Tone 1	261.55	32.22	211–315	2 542.55	334.99	2 014–2 940
Tone 2	272.00	44.37	208–354	2 466.82	345.44	2 008–2 993
Tone 3	262.82	55.44	179–354	2 541.64	364.39	2 008–3 184
Tone 4	271.45	25.15	237–329	2 485.82	335.18	1 941–2 861
F		0.20			0.14	
P		0.895			0.936	

Table 7 Mean and standard deviation values of F1 and F2 formant frequencies of /u/ in different tones Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
Tone 1	340.73	72.21	263–486	642.55	83.88	513–761
Tone 2	345.00	39.26	290–394	676.00	126.16	448–853
Tone 3	357.73	66.52	237–448	614.09	78.00	487–748
Tone 4	343.27	56.05	263–446	847.64	414.03	527–1 988
F		0.18			2.41	
P		0.921			0.081	

3.9 Acoustic analysis of the vowel /y/ across tones

The mean, standard deviation, and range of F1 and F2 formant frequencies for /y/ in different tones produced by all participants are shown in Table 8 below. A one-way *ANOVA* test was used to assess the effect of different tones (Tone 1 *vs*. Tone 2 *vs*. Tone 3 *vs*. Tone 4) on the spectral features of F1 and F2 frequencies. The result of a one-way *ANOVA* test indicated that there was no significant difference with the values of F1 and F2 for different tones of the vowel /y/ (F1: P=0.114, F2: P=0.754).

Table 8 Mean and standard deviation values of F1 and F2 formant frequencies of /y/ in different tones Hz

Vowel	F1			F2		
	M	SD	Range	M	SD	Range
Tone 1	287.00	28.30	257–341	1 994.45	216.63	1 725–2 370
Tone 2	277.18	31.84	214–329	1 980.82	293.91	1 530–2 304
Tone 3	276.00	46.93	214–341	2 066.18	232.91	1 777–2 369
Tone 4	311.18	39.07	263–382	1 935.45	355.36	1 223–2 389
F		2.12			0.41	
P		0.114			0.754	

4 Discussion

One aim of the study was to identify the characteristics of vowels in Mandarin Chinese. Based on the average format frequencies of F1 and F2 from the 11 subjects, it can be hypothesized that the six main vowels of mandarin form a trilateral on the acoustic vowel diagrams intuitively. The three vowels (/i, u, a/) with extreme values can be located in Figure 2, and are found at the three angles of the triangle, which are distinctive from the quadrilateral formation of English vowels.

The various points of each sound in the diagrams represent the different length of the subjects' vocal tracts, which was determined by tongue contraction and jaw height. According to the findings of other studies, we can primarily conclude that the larger the value of F1, the lower the jaw. Similarly, the larger the value of F2, the less contractive the muscles of the tongue, which means the tongue position is relatively in the front of the oral cavity. If the height of the diagram is divided into partitions corresponding to F1 and F2, the high vowels are /i, u, y/, the mid vowels are /o, e/, and the low vowel is /a/. In the same way, the front vowel is /i/, the central vowels are /y, a, e/, but the /y/ is quite close to the front, whereas /e/ is near the back, and the back vowels are /o, u/.

The results of the one-way *ANOVA* tests identified that F1 and F2 did not show a significant difference in the vowels /a/, /i/, /u/ or /y/ across tones (Tables 3, 6,

7, & 8). These results are in agreement with the study of Shaw [25], who found that some aspects of lingual articulation were stable across tones. In this study, both /a/ and /u/ were produced at T4 (/a/ and /y/), and they showed the highest mean for F1 and lowest mean for F2, in contrast to T1, T2, and T3. This indicated when producing /a/ with falling tone and /y/ with falling tone, the anterior oral cavity became longer, and the posterior oropharyngeal space was much shorter than other tones, as the tongue retracted and the tongue root depressed. However, /i/ produced with level tone has the lowest mean for F1 and the highest mean for F2, when compared to the other tones. This indicated that when producing /i/ with level tone, the anterior oral cavity was shorter, and the posterior oropharyngeal space was much longer than other tones, as the tongue advances and the tongue root elevated [10, 29-30]. When the vowel /u/ was articulated for dripping tone with the tongue was more elevated and retracted than other tones. This indicated the oral cavity was lengthened, and the posterior oropharyngeal cavity shortened [10, 29-30]. These results were in agreement with the study of Erickson [31] who reported that when producing the Tone 3 compared to the Tone 1, the tongue and jaw are more retracted. This can be explained by a general rule in acoustic-articulatory relationship, that the F1 frequency varies inversely with tongue height, and F2 varies inversely with tongue advancement [29].

Statistical analyses showed that there were significant differences between the vowels of /ɔ/ and /e/ across tones for F1 and F2 (Table 4 and 5). These findings suggest that production of /ɔ/ and /e/ across tones was much different in the anterior oral cavity and posterior oropharyngeal placement. For /ɔ/ produced with level tone, the anterior oral cavity became shorter, and the posterior oropharyngeal space was much smaller than with other tones, as the tongue blade was advanced and the tongue root was depressed. For /e/, the only significant effect of F1 was found between rising tone and dripping tone. It demonstrated that when producing /e/ with rising tone, the posterior oropharyngeal space was much smaller than in other tones, as the tongue retracted and the tongue root depressed more than when producing /e/ with dripping tone.

5 Conclusion

This study established a vowel formant space and investigated the first two formant frequencies of vowels across tones in 11 healthy young adult Chinese between 23 and 33 years old. An acoustic vowel space diagram was created showing that /i/, /u/, and /a/ are corner vowels, /ɔ/ and /e/ are in the central part. The six main vowels of Mandarin Chinese are shaped as a trilateral. Statistical analyses showed that there were significant differences in vowels of /ɔ/ and /e/ across tones in F1 and F2. When /o/ with level tone is produced, the tongue could both advance and depress more than with other tones. For /e/ with rising tone, the tongue was more retracted than /e/ with dripping tone.

The present study is a preliminary acoustic analysis in Mandarin-speaking young adults. Further studies should investigate a larger sample of Mandarin-speaking young adults. The clinicians may obtain references for acoustic assessment of Mandarin-speakers from the data provided in this study.

Acknowledgements

Sincerely gratitude is extended to our eleven Mandarin participants from Duquesne University of Pittsburgh, USA. We are also very grateful to Sarah Leech for her valuable inputs to this manuscript.

References

[1] FANT G. A note on vocal tract size factors and non-uniform F-pattern scaling [J]. Speech Transmission Laboratory Quarterly Progress and Status Report, 1966, 1: 22-30.
[2] PETERSON G E, BARNEY H L. Control methods used in a study of the vowels [J]. J Acoust Soc Am, 1952, 24(2): 175-184.
[3] PETERSON G E. Parameters of vowel quality [J]. J Speech Hear Res, 1961, 4(1): 10-29.
[4] STRANGE W. Evolving theories of vowel perception [J]. J Acoust Soc Am, 1989, 85(5): 2081-2087.
[5] LADEFOGED, PETER. A Course in Phonetics (Fifth Edition) [M]. Boston, MA: Thomson Wadsworth, 2006: 188.
[6] LADEFOGED, PETER. Vowels and Consonants: An Introduction to the Sounds of Language [M]. Maldern, MA: Blackwell, 2001: 40.
[7] HILLENBRAND J, GETTY L A, CLARK M J, et al. Acoustic characteristics of American English vowels [J]. J Acoust Soc Am, 1995, 97(5 Pt 1): 3099-3111.
[8] SYRDAL A K, GOPAL H S. A perceptual model of vowel recognition based on the auditory representation of American English vowels [J]. J Acoust Soc Am, 1986, 79(4): 1086-1100.
[9] FOX R. A, JACEWICZ E. Dialect and generational differences in vowel space areas [C]. Athens: Third ISCA Workshop on Experimental Linguistics, 2010.
[10] RAPHAEL L J, BORDEN G J, HARRIS K S. Speech Science Primer: Physiology, Acoustics, and Perception of Speech [M]. 5th ed. Baltimore: Williams & Wilkins Co, 2007: 105-130.
[11] ANDRIANOPOULOS M V, DARROW K, CHEN J. Multimodal standardization of voice among four multicultural populations' formant structures [J]. J Voice, 2001, 15(1): 61-77.
[12] CHEN Y Y. The acoustic realization of vowels of Shanghai Chinese [J]. J Phon, 2008, 36: 629-648.
[13] MAJEWSKI W, HOLLIEN H. Formant frequency regions of Polish vowels [J]. J Acoust Soc Am, 1967, 42(5): 1031-1037.
[14] SAKAYORI S, KITAMA T, CHIMOTO S, et al. Critical spectral regions for vowel identification [J]. Neurosci Res, 2002, 43(2): 155-162.
[15] VORPERIAN H K, KURTZWEIL S L, FOURAKIS M, et al. Effect of body position on vocal tract acoustics Acoustic pharyngometry and vowel formants [J]. J Acoust Soc Am, 2015, 138 (2): 833-845.
[16] HOMMA Y. An acoustic study of Japanese vowels: their quality, pitch, amplitude, and duration [J]. Stud Sounds, 1973, 16: 347-368.
[17] PURCELL E T. Formant frequency patterns in Russian VCV utterances [J]. J Acoust Soc Am, 1979, 66(6): 1691-1702.

[18] WHITE P. Formant frequency analysis of children's spoken and sung vowels using sweeping fundamental frequency production [J]. J Voice, 1999, 13(4): 570–582.
[19] LEE S, IVERSON G.K. The development of monophthongal vowels in Korean: age and sex differences [J]. Clin Linguist Phon, 2008, 22(7): 523–536.
[20] NATOUR Y S, MARIE B S, SALEEM M A, TADROS Y K. Formant frequency characteristics in normal Arabic–speaking Jordanians [J]. J Voice, 2011, 25(2): e75–e84.
[21] CHEN C C, LIN K C, WU C Y, et al. Acoustic study in Mandarin speaking children: developmental changes in vowel production [J]. Chang Gung Med J, 2008, 31(5): 503–509.
[22] NG M L, CHEN Y. Proficiency in English sentence stress production by Cantonese speakers who speak English as a second language (ESL) [J]. Int J Speech Lang Pathol, 2011, 13(6): 526–535.
[23] TING H N, ZOURMAND A, CHIA S Y, et al. Formant frequencies of malay vowels produced by Malay children aged between 7 and 12 Years [J]. J Voice, 2012, 26(5): 664.
[24] MAYO R, GRANT W C. Fundamental frequency, perturbation, and vocal tract resonance characteristics of African–American and white American males [J]. J Natl Black Assoc Speech Lang Hear, 1995, 17: 32–38.
[25] SHAW J A, CHEN W R, PROCTOR M I, et al. Influences of tone vowel articulation in Mandarin Chinese [J]. J Speech Lang Hear Res, 2016, 59(6): S1566–S1574.
[26] DUANMU S. The phonology of standard Chinese [M]. Oxford, UK: Oxford University Press, 2000: 1–15.
[27] LIANG Y, NUMAN F A, LI K, et al. Spectrum analysis of Chinese vowels formant in patients with tongue carcinoma underwent hemiglossectomy [J]. Int J Clin Exp Med, 2015, 8(2): 2867–2873.
[28] CHEN Y, NG M L, LI T S. English vowels produced by Cantonese–English bilingual speakers [J]. Inte J Speech Lang Pathol, 2012, 14(6): 557–568.
[29] KENT R D, WEISMER G, KENT J F, et al. Acoustic studies of dysarthric speech: methods, progress, and potential [J]. J Commun Disord, 1999, 32(3): 141–180, 183–186.
[30] FANT G. Acoustic theory of speech production [M]. The Hague: Mouton, 1970: 107–135.
[31] ERICKSON D, IWATA R, ENDO L, et al. Effect of tone height on jaw and tone articulation in Mandarin Chinese [C]. Beijing: Proceedings of the International Symposium on Tonal Aspects of Languages, 2004: 53–56.

成人普通话元音的声学分析

王臻旎[1], 陈　旸[2], 吴民华[3], 姚立群[4*], 张伟明[5*]

1 上海市瑞金康复医院,上海 200125;
2 杜肯大学,匹兹堡,宾夕法尼亚州 PA15282,美国;
3 香港大学教育学院,香港 999077;
4 福建中医药大学护理学院,福建 福州 350122;
5 上海交通大学医学院附属瑞金医院,上海 200025
* 通信作者:姚立群,E-mail:yaoliqunpt@163.com;张伟明,E-mail:zwm40397@rjh.com.cn

摘要　**目的:**声学分析是对语音的一种客观评估方法,能够相对简单、直观地检测元音的发音情况。普通话是一种声调语言,同一音段在不同的声调下产生不同的意义。先前的研究已经检测分析了母语为英语的美国成年人的元音发音特征。通过检测 23～33 岁中国青壮年人群普通话元音产生的共振峰频率来确定普通话元音 4 个声调的共振峰特性。**方法:**在安静的环境下,记录 11 名以普通话为母语的青壮年 6 个元音(/a/、/ɔ/、/e/、/i/、/u/、/y/)的声学信号,并指示参与者使用舒适的声音响度和语速阅读所有的语音样本。其中语音样本是随机产生。同时,使用专业的声学分析系统(Multi-Speech, KayPentax,美国)对第一、第二共振峰(F1、F2)进行分析,声学检测系统使用时域波形和频域宽带声谱图(滤波器带宽=300 Hz)。**结果:**汉语元音/a/的 F1 均值最高,而汉语元音/i/的 F1 均值最低;汉语元音/i/的 F2 均值最高,而汉语元音/u/的 F2 值最低。汉语元音/ɔ/和/e/在不同声调下,其 F1 和 F2 差异有统计学意义($P<0.05$)。而/a/、/i/、/u/、/y/4 个汉语元音在不同声调下的 F1 和 F2 差异无统计学意义($P>0.05$)。**结论:**建立普通话的元音三角形模型,其中以/i/、/u/、/a/为角元音,/ɔ/和/e/为中间元音,形成了普通话 6 个核心元音的三角形模型。不同声调下,元音/ɔ/和/e 的/F1 和 F2 有显著差异。当元音/ɔ/发第一声调时,舌头在口腔的位置与其他 3 个声调相比更加前伸和下沉。当/e/发第二声调时,舌头在口腔的位置与第三声调相比更加向后收缩。

关键词　普通话元音;共振峰频率;元音间距;声学分析;青壮年
DOI:10.3724/SP.J.1329.2020.03004