An MDVP Analysis on Basic Vowel Voice of Mandarin Chinese

Mingxiao Gu[a], Yonghong Li[b], Dawa Pengcuo[*]

Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Gansu 730030, China

*Corresponding author: dwpc@163.com, [a]270523584@qq.com, [b]lyhweiwei@126.com

Keywords: Persistent vowels, Voice, Fundamental frequency, Frequency and amplitude jitter.

Abstract: Pronunciation research is an important part of speech acoustics research. In this study, the MDVP is used to extract and analyze ten basic vowels of Mandarin Chinese. The parameters mainly involve F_0, Jita, Jitt, ShdB, Shim, APQ and sAPQ. This study also compared the differences in voice parameters between different genders. The study found that the jitter of a [A] is very obvious and has a notable statistical significance, the frequency and amplitude jitter of ü↙ ↘ are the weakest. In addition, F_0, Jita, APQ, and sAPQ have significant differences in gender.

1. Introduction

Speech production can be divided into two parts: tuning and pronunciation. Pronunciation refers to the sound produced by different vibration modes of vocal cords under the action of airflow. [1] This study focuses on the voice characteristics of ten basic vowels in Mandarin Chinese and the differences between male and female voices. It intends to use MDVP and accurate statistical methods to analyze the voice acoustics of ten vowels. The research covers the voice data of ten basic vowels in Mandarin Chinese between different sexes, and finds out the distribution of voice parameters.

In the past, the study of tuning is more common, while the study of voice pronunciation and other aspects of acoustics is rare. The purpose of this study is to provide some references for exploring vowel properties, voice evaluation and speech synthesis in the future.

2. Materials and Methods

2.1. Selection of Research Objects and Data Acquisition

The pronunciation person is 5 males and 5 females, aged 22-25 years old. They are native speakers of Mandarin Chinese without obvious dialectal accent. The subjects have normal pronunciation, no professional vocalization or singing training, no throat diseases, no smoking and alcohol addiction, no history of voice disorders and respiratory diseases.

The recording software is Adobe Audition CC 2018, with a sampling rate of 44100HZ and a resolution of 16 bits (according to the condition that MDVP can handle speech). The voice collection is carried out in a quiet room with an ambient noise of 45 dB (A) or less. The speaker's mouth is about 15 cm from the microphone. They all take natural comfort, and then smoothly speak ten vowels a [A], ⊙↙⌐↘, ⊙↖↙∟↘, ↙↘, o↙∘↘, _↙_↘, ü↙ ↘, -↙ʅ↘, -↙ʅ↘, ⊙→↙‖→↘. The speakers should stretch the sound to make the vowel last for at least 4 seconds, and read each vowel three times. After that, the 3s stable section was taken and measured, then the average value was taken. [2]

2.2. Acoustic acoustic analysis

We use the Multi-Dimensional Voice Program (MDVP) of Kay Corporation of the United States to analyze the voice acoustics of 10 basic vowels in Mandarin Chinese. The MDVP can extract 33

  DOI: 10.25236/adlh.2019.034

acoustic parameters, which can be divided into six basic categories: pitch basic parameters, frequency jitter parameters, amplitude jitter parameters, voice index, voice clearing parameters, and basic parameters. [3]

This experiment selects several representative parameters of MDVP for analysis and comparison. By analyzing the Average Fundamental Frequency (F_0), Absolute Jitter (Jita), Jitter Percent (Jitt), Shimmer in dB (ShdB), Shimmer Percent (Shim), Amplitude Perturbation Quotient (APQ), Smoothed Ampl. Perturbation Quotient (sAPQ) between different vowels and different gender groups, we can explore the characteristics of basic vowels in Mandarin Chinese.

Brief description of the selected parameters [4]:

F_0 is a basic concept in speech acoustics and is defined as the reciprocal ($1/T_0$) of the vocal cord vibration period (T_0). The F_0 average fundamental frequency is the average of the fundamental frequency values of all extracted periods.

Jita is the analysis of the irregularity of the fundamental frequency period in the short period voice. It is defined the change between a voiced pitch period as:

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right| \quad (1)$$

Jitt is the relative value of the fundamental frequency change (very short period) of the period of the voice. It is defined the relative change between a voiced pitch period as:

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}T_0^{(i)}} \quad (2)$$

In (1) and (2), i = 1, 2, 3, ... N of $T_0(i)$ is the extracted pitch period parameter, and N is equal to the number of extracted pitch periods.

ShdB is the evaluation of the inter-period (very short period) variation of the amplitude between peaks and peaks. It uses decibels to represent amplitude perturbations. Calculated as follows:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log\left(A^{(i+1)}/A^{(i)}\right)\right| \quad (3)$$

Shim is the relative assessment of the inter-period variation of amplitude between peaks and peaks. It represents the relative periodic variation between peaks and peaks. Calculated as follows:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i)} - A^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \quad (4)$$

APQ is the relative evaluation of peak-to-peak amplitude and period-to-cycle variation. In the analyzed voice zone, after smoothing 11 cycles of processing, it can describe the variation of short-term amplitude. Calculated as follows:

$$APQ = \frac{\frac{1}{N-10}\sum_{i=1}^{N-10}\left|\frac{1}{11}\sum_{r=0}^{10}A^{(i+r)} - A^{(i+5)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \quad (5)$$

sAPQ is the relative evaluation of the short-term or long-term amplitude variation between peaks and peaks. It is widely used in the research for measurement of voice perturbations, which is very sensitive to amplitude variations between successive pitch periods. Calculated as follows:

$$sAPQ = \frac{\frac{1}{N-sf+1}\sum_{i=1}^{N-sf+1}\left|\frac{1}{sf}\sum_{r=0}^{sf-1}A^{(i+r)} - A^{(i+m)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \quad (6)$$

In (3), (4), (5), and (6), i=1, 2, 3...N of A(i) is the extracted peak amplitude parameter, and N is equal to the number of extracted periods.

2.3. Statistical methods and mapping

Use SPSS 24.0 version for Independent Sample T test. The average number of each parameter was compared by gender, and the significance level was compared by the value of P value. At the same time, the mean and standard deviation of various parameters were compared. In addition, a one-way ANOVA test (analysis of variance) was performed. The LSD method was used to compare ten vowels, and P value was used to compare the significance levels among the vowels.

Use Microsoft Excel 2016 for data integration and drawing.

3. Data Result Analysis

3.1. Overall situation of Voice Acoustic Analysis

Table.1. Acoustic analysis results of male and female voices (x±s)

Gender	Numbers	F_0 (HZ)	Jita (μs)	Jitt (%)	ShdB (dB)	Shim (%)	APQ (%)	sAPQ (%)
Female	50	232.90±19.20	51.84±45.01	1.17±0.98	0.29±0.18	3.33±2.03	2.25±1.21	3.67±1.73
Male	50	122.52±17.57	87.94±59.88	1.05±0.65	0.34±0.13	3.89±1.42	3.18±1.32	5.67±2.41

Table.2. Acoustic analysis results of male and female voices (P value)

F_0(HZ)	Jita(μs)	Jitt(%)	ShdB(dB)	Shim(%)	APQ(%)	sAPQ(%)
0.000	0.001	0.482	0.122	0.112	0.000	0.000

Tables 1 and 2 show the results of the acoustic analysis of the voice between different genders by using independent sample T test. The data is the average of the ten vowels' voice data. It can be seen that female's F_0 is higher than male, the difference is statistically significant ($P<0.01$); Jita value is higher in male than female, the difference is significant ($P<0.01$); APQ, sAPQ male is significantly higher than female ($P< 0.01$). The remaining differences were not statistically significant.

3.2. Fundamental Frequency Information

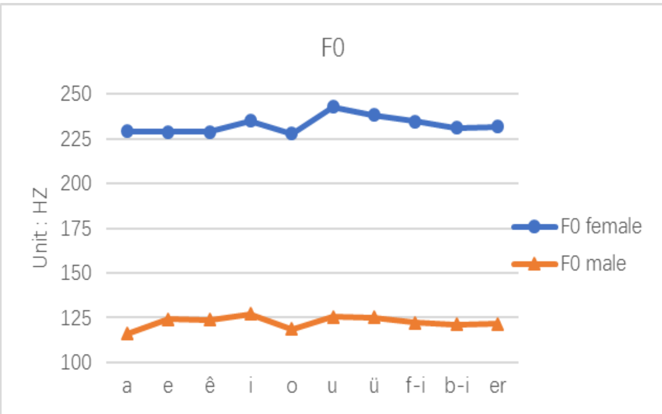

Fig. 1 Comparison of ten basic vowels' F0 between male and female

Fig. 1 shows the F_0 of ten basic vowels in Mandarin Chinese spoken by five males and five females. The curves are drawn by calculating the average vowels by gender (Note: "f-i" stands for "front i", "b-i" stands for "back i"). Fig. 1 shows that:

There are significant differences between the two curves, but the trend of the two curves is roughly the same. Firstly, there are obvious concave points at o↙∘↘, which indicates that the F_0 of o↙∘↘ is the lowest among the ten basic vowels. Secondly, there are convex points in ↻↙↻↘,

_↙_↘, ü↙ ↘, that is to say, ⟳↙⟳↘, _↙_↘, ü↙ ↘ have high F_0 values in ten basic frequencies.

The men's F_0 data is between 116 Hz and 127 Hz, and the F_0 span is about 11 HZ. The women's F_0 is between 228Hz and 242Hz, and the F_0 span is 14HZ. The vocal tract of men is long, wide and thick, but the vocal tract of women is short, narrow and thin [5], which results in that the F_0 of men is lower than women. However, the upper and lower limit span of male and female F_0 are not large.

3.3. Frequency Jitter Parameters

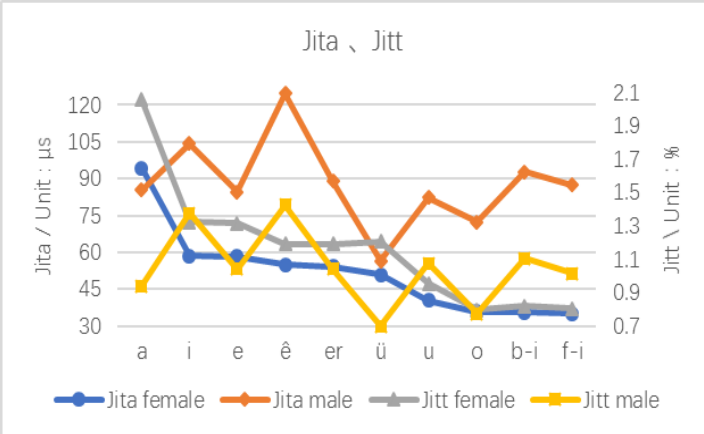

Fig. 2 Comparison of Jita and Jitt of ten basic vowels between male and female

Fig. 2 shows the Jita and Jitt of ten basic vowels in Mandarin Chinese, which are spoken by five men and five women. The curves are drawn according to the average value of each sex. Among them, the "Jita female" curve is plotted in order from big to small, which is convenient to observe the direction of other curves. Fig. 2 shows that:

There is a close relationship between Jita and Jitt. Under the same gender condition, the trend of Jita and Jitt curves is basically the same.

Firstly, the values of the four curves are relatively low at o↙∘↘, which indicates that the frequency jitter of o↙∘↘ is small. Secondly, the values of the four curves are relatively high at ⟳↙⟳↘, which indicates that the frequency jitter of ⟳↙⟳↘ is large. In addition, the Jita and Jitt values of a[A], ⊙↖↙∟↘, ü↙ ↘ differ greatly between men and women. For women, the frequency jitter of a[A] and ü↙ ↘ is large, while for men, the frequency jitter of ⊙↖↙∟↘ is large.

3.4. Amplitude Jitter Parameters

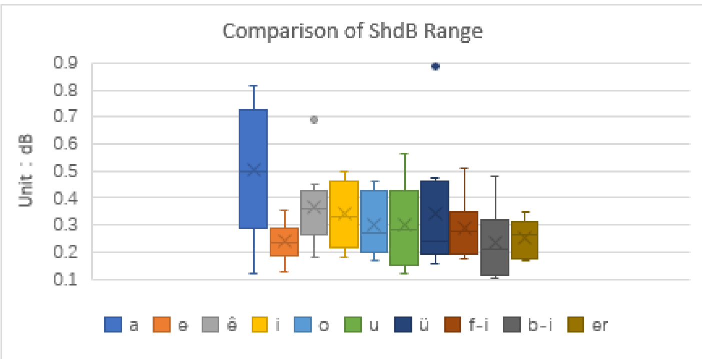

Fig. 3 Comparison of basic vowels ShdB range

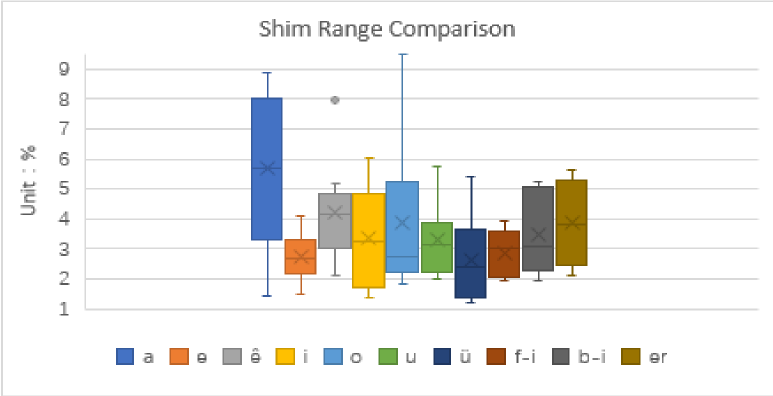

Fig. 4 Comparison of basic vowels Shim range

Fig. 3 and Fig. 4 are listed as box diagrams of ShdB and Shim of ten basic vowels in Mandarin Chinese spoken by five men and five women. Through these two figures, we can compare the range of ShdB and Shim of ten basic vowels. Fig. 3 and Fig. 4 shows that:

In terms of span, the span of ShdB and Shim of a[A] is the largest, while ⊙↙⌴↘ is the smallest. In addition, the span of ↙↘, o↙∘↘ is relatively large, while the span of ⊙↖↙∟↘, -↙⌴↘ is relatively small.

From the upper limit of span, the ShdB and Shim values of a[A] are the largest, indicating that the amplitude jitter of a[A] is the most obvious. From the lower limit of span, in ShdB data, the value of −↙⌴↘, ⎽↙⎽↘, ⊙↙⌴↘ three vowels is low. In Shim data, the values of ü↙ ↘, ↙↘, ⊙↙⌴↘ is low.

Table.3. ShdB,Shim (P<0.05)

ShdB	a									ê	
	e	ê	i	o	u	ü	Front i	Back i	er	ü	
	0.000	0.039	0.002	0.015	0.001	0.000	0.000	0.002	0.014	0.043	
Shim	a									ê	
	e	ê	i	o	u	ü	Front i	Back i	er	e	ü
	0.000	0.043	0.002	0.012	0.001	0.000	0.000	0.003	0.014	0.045	0.032

In addition, the ShdB and Shim values of ten vowels are tested by single factor ANOVA. Table 3 shows all samples with P values less than 0.05. The ShdB and Shim values of a[A] were significantly different from all vowels. The differences between a[A] and ⊙↙⌴↘, ↙↘, ⎽↙⎽↘, ü↙ ↘,-↙⌴↘, -↙⌴↘ is very obvious (P <0.01), indicating that the amplitude jitter of a[A] is very obvious. Besides, there are also significant differences between ⊙↖↙∟↘ and ⊙↙⌴↘, ü↙ ↘.

Because the APQ and sAPQ are more precise descriptions of the amplitude jitter. Fig. 5 is a plot of the APQ and the sAPQ of ten basic vowels of Mandarin Chinese spoken by five men and five women. Among them, "APQ female" curve is plotted in order from big to small. Fig. 5 shows that:

There is a close relationship between APQ and sAPQ. Under the same gender conditions, APQ curves and sAPQ curves are different in height, but tend to coincide roughly.

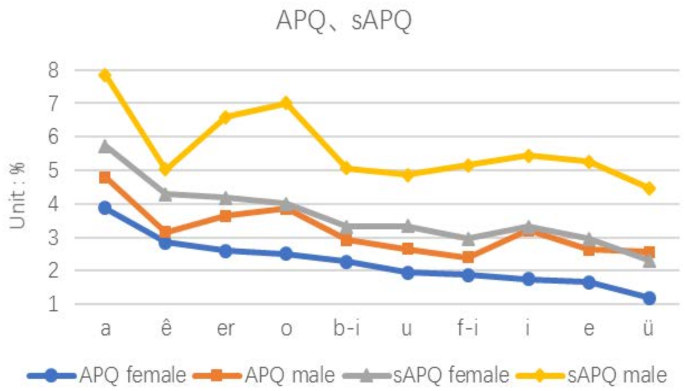

Fig. 5 Comparison of APQ and sAPQ of ten basic vowels between male and female

The amplitude jitter of a[A], o↙◦↘, ⊙₋↙‖₋↘ is obvious, and the amplitude jitter of ⊙↙⎵↘ and ü↙ ↘ are small. In addition, for females, the amplitude jitter of ⊙↖↙⌞↘ is large, while for males, the amplitude jitter of i↙◌↘ and -◌↙⎵˥↘ ◌₋ large.

Table.4. APQ,SAPQ (P<0.05)

APQ	a									ü		
	e	ê	i	o	u	ü	Front i	Back i	er	ê	o	er
	0.000	0.015	0.001	0.036	0.000	0.000	0.000	0.002	0.026	0.042	0.018	0.025
sAPQ	a									ü		
	e	ê	i	u	ü	Front i	Back i			o	er	
	0.008	0.035	0.018	0.008	0.001	0.007	0.011			0.037	0.048	

In addition, the APQ and sAPQ values of ten vowels were tested by single factor ANOVA. Table 4 shows samples whose P values were less than 0.05. Among them, the APQ value of a[A] has very significant difference between⊙↙⎵↘, ◌↙◌↘, ₋↙₋↘, ü↙ ↘,-◌↙⎵˥↘, -◌↙⎵˥↘ (P<0.01). The sAPQ value of a[A] is different from ⊙↙⎵↘, ₋↙₋↘, ü↙ ↘,-◌↙˥↘. Besides, ü↙ ↘'s APQ is different from ⊙↖↙⌞↘, o↙◦↘, ⊙₋↙‖₋↘, ü↙ ↘'s sAPQ is different from o↙◦↘, ⊙₋↙‖₋↘.

3.5. Frequency jitter parameters and amplitude jitter parameters

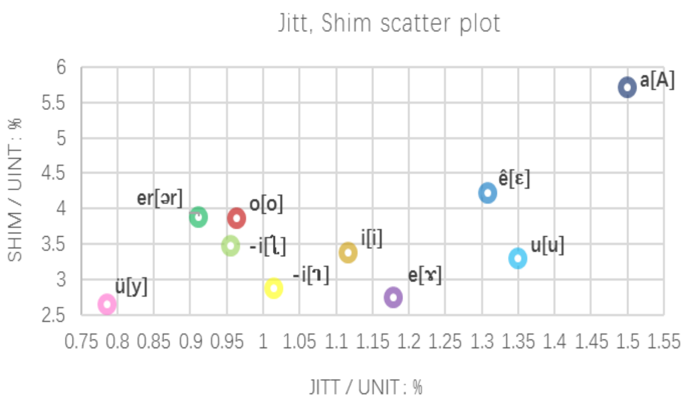

Fig. 6 Ten basic vowels Jitt and Shim scatter plot

Fig. 6 shows Jitt and Shim scatter plot of the ten basic vowels spoken by five men and five women. Among them, Jitt and Shim respectively averaged. Jitt, which is displayed on the horizontal axis, and Shim, which is displayed on the vertical axis. Fig. 6 shows that:

As a whole, a[A] is located at the upper right position of the figure, indicating that the Jitt and Shim values of a[A] are large; ü↙ ↘ is located at the lower left position of the figure, indicating that Jitt and Shim values are both small.

From the horizontal axis, the points at the front are ü↙ ↘, ⊙→↙‖→↘, -ๆ↙︶ʅ↘, o↙∘↘, -ๆ↙ʅ↘, ๆ↙ɕ↘, indicating that its Jitt value is small. The points behind the position are a[A], _↙_↘ ⊙↖↙∟↘ ⊙↙⅃↘, indicating that the Jitt value is large.

From the vertical axis, the points below the position are ü↙ ↘ ⊙↙⅃↘ −ๆ↙︶ʅ↘ _↙_↘ ๆ↙ɕ↘ -ๆ↙︶ʅ↘, o↙∘↘ ⊙→↙‖→↘ indicating that the Shim value is small. The point above the position are a[A] and ⊙↖↙∟↘, indicating that the Shim value is relatively large.

4. Conclusion and discussion

Through comparison and analysis above, we can draw the following main conclusions:

The F_0 of male and female voices is significantly different, and women's F_0 is significantly higher than males. For the ten basic vowels, the F_0 of o↙∘↘ is the lowest, while the F_0 of ๆ↙ɕ↘ and _↙_↘ is relatively high, but it is not statistically significant.

In the frequency jitter parameters, the Jitt value does not differ significantly between men and women. However, the female Jita value is higher than male as a whole, so it can be seen that the female frequency jitter is stronger than the male. For ten basic vowels, the Jita and Jitt values of o↙∘↘ are small, but ๆ↙ɕ↘ are relatively large.

In the amplitude jitter parameters, the APQ and sAPQ values of males are higher than females. For ten vowels, the value of ShdB and Shim of a[A] is the largest, and ⊙↙⅃↘ is small. In addition, the APQ and sAPQ values of a[A], o↙∘↘, ⊙→↙‖→↘ are large. While, the the APQ and sAPQ values of ⊙↙⅃↘ and ü↙ ↘ are small. In all, the amplitude jitter is most obvious in a[A], weakest in ⊙↙⅃↘.

Wu Xiyu [6] mentioned that “Amplitude jitter (ShdB and Shim) and Jita decrease with the increase of pitch.” For this conclusion, we put Fig. 1 compare with Fig. 3, and Fig. 5 to observe the trend and points of the curve. It is found that the F_0 and ShdB, Shim, APQ, sAPQ are negatively correlated. It corresponds to the conclusion "Amplitude jitter decreases as the pitch increases".

The above conclusions show that there are differences in the voice parameters between different genders and different vowels, and some parameters have certain correlations. This study is a reference and guidance for future research. However, the conclusions drawn from this study are still mostly descriptive nature. Although statistical methods are used, the data is small and not statistically significant. These shortcomings are expected to improve in future research.

Acknowledgments

This work was financially supported by NSFC grant fund (No. 61562076).

References

[1] Kong Jiangping, 2001. On Language Pronunciation [M]. *Beijing: Central University for Nationalities Press.*

[2] Li Xiaohui, Shan Chunguang, Hao Fang, 2012. Analysis of Acoustic Characteristics of Voices in Variable Period [J]. *Journal of Audiology and Speech Pathology*.

[3] Kong Jiangping, 2015. Basic Course of Experimental Phonetics [M]. *Beijing: Peking University* Press.

[4] Shi Jing, Zhou Ling, Yu Hongzhi, 2010. Research on Multidimensional Voice Analysis Software (MDVP) [J]. *Journal of Changchun Institute of Engineering (Natural Science Edition).*

[5] Shen Mixia, Kong Jiangping, 1998. MDVP (multi-dimensional voice processing) study on sustained vowels of Mandarin through EGG signal[C]. *Proceedings of Conference on Phonetics of the Languages in China*.

[6] Wu Xiyu, 2016. Voice Acoustics Research in the Period of Change of Voice [J]. *Linguistic Theories Series.*