Analysis of Chinese Vowel Pattern of Bangladeshi Students Based on Formants

Junying Niu

***Abstract*—With China's increasing international status, there has been a rise in the number of Bangladeshi students studying in China. Consequently, more students from Bangladesh are learning Chinese pronunciation to aid them in their work, studies, and daily life. Without proper instruction, they may encounter difficulties in learning Chinese pronunciation and make more errors. To enhance the Chinese language proficiency of Bangladeshi international students, this paper proposes a method for analysing the morphology of Chinese consonants based on their morphological features. The method analyses Chinese consonants from the perspective of their pronunciation, and the extracted consonant data are used to draw consonant coordinate maps and vowel space maps. Comparative analyses were conducted to identify differences in Chinese vowel pronunciation between Chinese and Bangladeshi students. The analysis provides a theoretical guide to correct pronunciation for Bangladeshi students.**

***Index Terms*—Bengali international students, Chinese vowels, Vowel pattern, formant feature.**

I. Introduction

As one of the countries that has established diplomatic relations with China, Bangladesh has increasingly frequent contacts with China. Therefore, Chinese has become a tool for Bangladeshis to conduct business and trade in China. With the increasing number of Bangladeshi students studying abroad, it has been observed that they have non-standard pronunciation while learning Chinese. However, research on the acquisition of Chinese by Bangladeshi students at this stage is still lacking. Studying the acquisition of Chinese vowels by Bangladeshi students will help us understand their overall Chinese language learning, standardize their pronunciation, and provide valuable teaching references. Therefore, understanding the pronunciation of Chinese vowels by Bangladeshi students is of great significance.

From an acoustics perspective, the pronunciation of vowels is mainly determined by their formant mode [1]. Formants are crucial in distinguishing vowel quality, with resonant peaks labelled in order of appearance: the first formant is F1, the second is F2, the third is F3, and so on.

Research has shown a correlation between tongue position and formant. Specifically, the first formant (F1) decreases as tongue position height increases, while F1 increases as tongue position height decreases. Additionally, the second formant (F2) increases as tongue position moves forward and decreases as it moves backward. Finally, the third formant (F3) decreases as the degree of lip rounding increases and increases as the degree of lip rounding decreases [2]. Based on the function of human organs, vowels can be identified by analyzing 3-4 characteristics, such as tongue area, tongue height, and lip roundness [3]. These characteristics can be plotted on a Cartesian plane, known as the formant space. The vowel triangle is then described based on the distribution of vowels in the formant space, which represents the movement of the tongue in the vocal cavity during vowel production.

This paper proposes a method for analysing the vowel patterns of Bangladeshi foreign students in Chinese based on formant characteristics. The formant parameters were extracted using Praat software to draw a vowel pattern map. The differences in Chinese vowel pronunciation between Chinese and Bangladeshi students were compared and analysed..

II. Related Work

In recent years, research on the acoustic characteristics of Chinese learners has increased. Munro [4] studied the dynamic change characteristics of vowel pronunciation rules of learners with different native language characteristics over time. The results indicate that the speed of target language acquisition for new language learners is fast at first but gradually stabilizes with increased learning content. The research indicates that the process of second language vowel learning is complex and heavily influenced by the learner's mother tongue. Ugarte [5] proposed a formant extraction method based on discrete wavelet transform to analyze the vowel characteristics of Spanish speakers from Antioquia, Colombia. Tripathi [6] proposed a method based on continuous wavelet transform coefficients and phoneme boundaries to detect vowel regions from different modes of speech signals. Yadav [7] proposed a method to determine the vowel transition region based on the rate of change of formant frequency using zero time window and group delay function molecules. Aloqayli [8] analysed the vowels of classical Arabic and modern standard Arabic based on formants. Ramya [9] conducted speaker verification using vowel and non-vowel regions. Shahriar [10] analysed the linear predictive coding (LPC) coefficients of vowel-consonant-vowel (VCV) sequences to determine the time transfer function of the vocal tract.

III. Mathod

A. Analysis of vowel pattern Review

The systematic nature of each language's voice is reflected in its unique voice pattern. The vowel pattern includes internal variants, vowel positioning characteristics, and overall distribution relationships [11]. This paper will

Manuscript received February 28, 2024

Junying Niu, School of computer science and technology, Tiangong University, Tianjin, China.

compare and analyze the vowel spaces of Bangladeshi and Chinese students based on this. The first level vowel pattern serves as the foundation for all vowel patterns and is therefore considered typical. This paper presents the vowel formants of Chinese and Bangladeshi students' pronunciation obtained through phonetic experiments. A vowel pattern map is drawn and the pronunciation differences between the two groups are compared.

The formant data of the original signal is extracted using Praat software and converted into Bark values using formula (1):

$$Bark = 7ln\{(\frac{f}{650}) + [(\frac{f}{650})^2 + 1]^{1/2}\} \quad , \quad (1)$$

where f is the formants[6]. Input the obtained Bark value data into the normalization formula(2) to obtain its V value.

$$V = \frac{B_x - B_{min}}{B_{max} - B_{min}} * 100 \quad . \quad (2)$$

The Bark value is a unit of measurement used to make the distribution of vowels in the acoustic chart more closely aligned with human auditory perception. Both values are important in accurately representing vowel sounds. It is derived from the relativization of formant Hz data. Similarly, the V value is obtained through the normalization of formant Hz data. It is used as a scale on a three-dimensional graph to filter out individual pronunciation differences among speakers, retain the common characteristics of the relative relationship between vowels, and facilitate the comparison of voices between different languages, dialects, and speakers. Each sound unit can be located on the spatial map based on its Bark and V values, which reflect the distribution and relative acoustic and physiological positions of the sound units.

B. *Formants extraction*

The vocal tract can be seen as a tube with a non-uniform cross-section, acting as a resonator during pronunciation [12]. The excitation signal of the glottis entering the vocal tract as a quasi periodic pulse will cause resonance characteristics, resulting in a set of resonance frequencies, called formant frequencies or formants [13]. The formant carries a lot of useful information and is a frequency domain parameter with high stability in speech signal processing.

1) *Cepstrum method*

Cepstrum analysis assumes that speech signal is the result of convolution between excitation signal and linear time invariant system.

1.When the signal sequence is x(n), it is preprocessed to get $x_i(n)$.

2. The Fourier transform of $x_i(n)$ is performed to obtain

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n) e^{-j2\pi kn/N}. \quad (3)$$

3. Take logarithm of (3) to get:

$$\widehat{H}_i(k) = \log(|X_i(k)|). \quad (4)$$

4. The cepstrum sequence $\hat{x}_i(n)$ is obtained by inverse Fourier transform of (4) then:

$$\hat{x}_i(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}_i(k) e^{-j2\pi kn/N}. \quad (5)$$

5. Set a low-pass window function on the inverted frequency axis to obtain:

$$window(n) = \begin{cases} 1 & n \le n_0 - 1 和 n \ge N - n_0 + 1 \\ 0 & n_0 - 1 < n < N - n_0 + 1 \end{cases}, \quad (6)$$

where n_0 is the width of the window function.

6. After setting the window function, compare it with cepstrum sequence $\hat{x}_i(n)$ to obtain:

$$h_i(n) = \hat{x}_i(n) \times window(n). \quad (7)$$

7. After the Fourier transform of $h_i(n)$, we can get the envelope of $X_i(k)$,$H_i(k)$:

$$H_i(k) = \sum_{n=0}^{N-1} h_i(n) e^{-j2\pi kn/N} \quad . \quad (8)$$

8. The corresponding formant parameters can be obtained by finding the maximum value on the envelope line.

2) *Linear prediction algorithm*

The speech signal s(n) is generated by an all pole H(z) excited by the input sequence u(n), that is, the system transfer function is:

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-i}} \quad . \quad (9)$$

The relationship between s(n) and u(n) is expressed by the difference equation as:

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n) \quad . \quad (10)$$

The linear predictor is expressed as:

$$\hat{s}(n) = \sum_{i=1}^{p} a_i s(n-i) \quad , \quad (11)$$

where $\hat{s}(n)$It is obtained from the linear combination of the past p values, and a_i is the linear prediction coefficient.

There are two methods for calculating the formant using the LPC method. One method involves estimating the formant value by solving the negative root of the linear prediction coefficient polynomial, known as the root method. The other method involves estimating the formant value by identifying the local maximum of the power spectrum response curve of the channel transfer function, known as the interpolation method. The formant of Chinese and Bangladeshi students was obtained using the LPC method through Praat software..

IV. Experimental Results

A total of 20 speakers participated in the experiment, including 10 Bangladeshi students and 10 Chinese students. Bangladeshi international students are healthy, have normal pronunciation organs. Their mother tongue is Bengali and second language is English. They also have a certain Chinese foundation. In addition, all the Chinese students who participated in the experiment obtained the Putonghua Proficiency Certificate of grade II grade A or above.

In modern Mandarin Chinese, vowels are classified into two categories: monosyllabic and compound vowels. There are 10 monosyllabic and 13 compound vowels. Monosyllabic vowels are further divided into glossal, tip, and rolling vowels. Compound vowels are classified into front, back, and middle vowels based on their pronunciation. When foreign students begin learning Chinese in China, they are first introduced to Chinese pinyin. When learning Chinese Pinyin, the initial step is to become familiar with the monosyllabic sounds. This paper focuses on testing students' knowledge of the monosyllabic sounds /a/, /o/, /e/, /i/, /u/, and /ü/. During the sampling phase of the experiment, six Chinese vowels were presented as slides, and the subjects were asked to become familiar with the Chinese vowel list. In a quiet classroom, the speaker takes turns to say each of the six vowels five times, with a 0.5-1s pause after finishing the first time. If any mispronunciations, stuttering or other issues occur during the recording process, it should be repeated. The recording software used is a professional mobile phone recording software that records mono, 44100HZ, 16bit audio and saves it as a wav format file. The formant data of the original signal is extracted using Praat software. The frequency values of F1, F2 and F3 for each stable vowel segment are extracted and averaged over five points. The average formant data for each vowel is then calculated, resulting in separate vowel formant average tables for 10 speech samples from Chinese students and 10 speech samples from Bangladeshi students.

A. Analysis of vowel pattern of Bangladeshi students

The V values were extracted from the formants and a discrete coordinate plot was generated using Excel software after normalization operations. The plot was then compared and analyzed with the lipogram of Mandarin tongue vowels. The results are presented in Figure1 and 2.

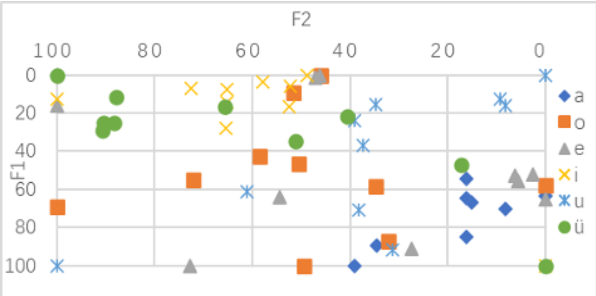

Figure1. Vowel Pattern of Bengali Students

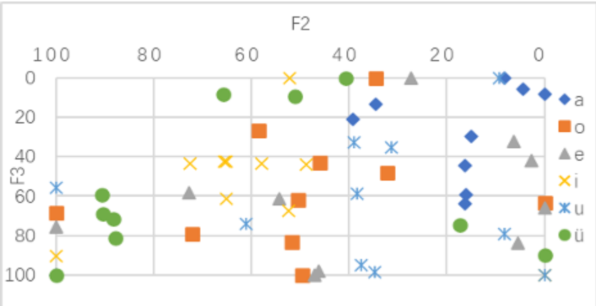

Figure2. Vowel Pattern of Bengali Students

Figure1 shows a coordinate diagram with the horizontal axis representing the normalized value of F2 and the vertical axis representing the normalized value of F1. As can be seen from Figure1, the pronunciation data set of vowel/a/ is relatively concentrated, which shows that the tongue position is low and the tongue position is backward, and its acoustic characteristics are similar /ɑ / The concentration of vowel/o/ pronunciation data is not obvious. Half of the samples' tongue position is backward, but the tongue position is significantly lower than that of Mandarin /o/. The data of vowel /e/ are relatively scattered, and most samples show that the tongue position is backward and low, which is lower than that of Mandarin /ɣ /, However, in the figure, there are also individual samples with high tongue position and relatively middle position in water square, whose acoustic characteristics are close to vowels / ə/. The pronunciation of vowel /i/ is also relatively concentrated, and each sample concentration shows high tongue position and front tongue position, and its acoustic characteristics are very similar to vowel /i/. The pronunciation samples of vowel /u/ are relatively discrete. Most of the pronunciation samples show high tongue position and backward, and their acoustic characteristics are close to Mandarin /u/. The pronunciation data of vowel /ü/ shows a certain concentration, but the range of values is relatively large. Its main feature is that the tongue is high and front, and its acoustic characteristics are close to Mandarin /y/.

To study the degree of roundness in vowel sounds, a coordinate diagram was created with the normalized value of F2 on the horizontal axis and the normalized value of F3 on the vertical axis. Figure2 illustrates the distribution characteristics of the round lip sounds /o/, /u/, and /ü/. The vowel /o/ is dispersed widely in the horizontal direction, but most of the samples are located in the middle of F3, reflecting the characteristics of semicircular lips. The values of the vowel /u/ are also discrete. Only three samples exhibit the characteristics of a backward tongue position and rounded lips, and their acoustic characteristics are similar to those of Mandarin /u/. The value of the vowel /ü/ also lacks concentration. Half of the samples exhibit rounded lip characteristics, while the rest exhibit low round lip and non-round lip characteristics.

B. Vowel space comparison between Bangladeshi students and Chinese students

In this part, the formants F1 and F2 of Chinese pronunciation of Chinese and Bangladeshi students are extracted by Praat software, and the obtained data are averaged to obtain Tables 1 and 2.

Table1. Average formants of Chinese vowel pronunciation of Bangladeshi students.

vowels	F1	F2
/a/	758	1258
/o/	410	819
/e/	462	1321
/i/	319	2425
/u/	341	743
/ü/	312	1989

Table2. Average formants of Chinese vowel pronunciation of Chinese students.

vowels	F1	F2
/a/	748	1239
/o/	478	856
/e/	501	1161
/i/	351	2840
/u/	354	763
/ü/	327	1961

According to the average data of the first common formant and the second formant in Table 1 and Table 2, the Chinese vowel space of Chinese students' and Bangladeshi students' respective pronunciation can be obtained, as shown in Figure3. Connect the basic vowel positions of the three limits /a/, /i/, /u/ in Figure1 with straight lines, and the triangular space formed by them is the vowel space of their respective Chinese pronunciation. As can be seen from Figure3, the vowel space emitted by Bangladeshi students and Chinese students is triangular, and the space is basically similar to the standard vowel system issued by the International Speech Association(Figure4). However, Bangladeshi students have a larger opening degree than Chinese students, but their tongue position is front and low.

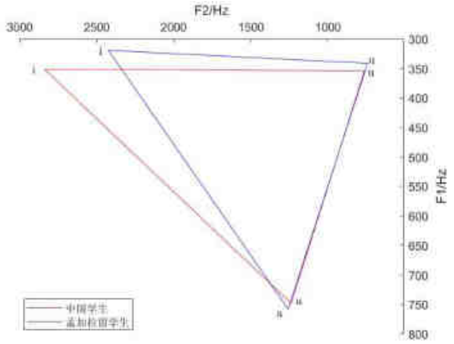

Figure3. Vowel Space Comparison between Bangladeshi and Chinese Students

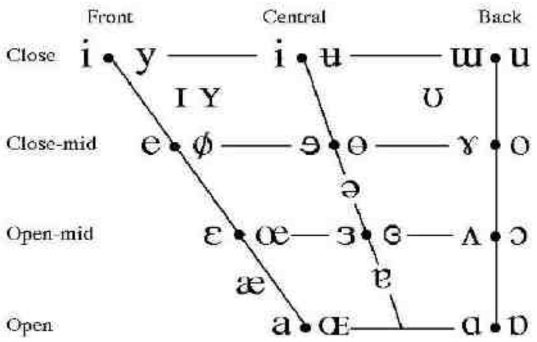

Figure4. International Standard Vowel System

V. Conclusion

The results of the experiment showed that Bangladeshi students were closest to Chinese students in the pronunciation of the vowels /a/ and /u/. However, the other consonants had different degrees of difficulties, including unstable pronunciation and uneven distribution. Therefore, Bangladeshi students may encounter difficulties in learning the pronunciation of Chinese consonants. The characteristics of the Chinese language, the existence of tones in Chinese, and the influence of the native language of Bangladeshi students are all factors contributing to this situation.

References

[1] Ladefoged.P.VowelsandConsonants[M].LosAngeles:BlackwellPublishers.2001.

[2] Shiliang L , Yasheng J , Ning M .Research on the vowel pattern of the Amdo Tibetan's Xiahe dialect[C]//2014 International Conference on Audio, Language and Image Processing (ICALIP).IEEE, 2015.DOI:10.1109/ICALIP.2014.7009819..

[3] Luo Anyuan. A Brief Analysis of Kazak Vowel Quality[J]. Journal of The Central University for Nationalities (Humane and Social Science Edition), 2007 (1):96-99..

[4] Munro M J,Derwing T M. Segmental acquisition in adult ESL learners: A longitudinal study of vowel production[J].Language learning,2008, 58(3): 479-502.

[5] Orellana, Simon,Ugarte, Juan P..Vowel characterization of Spanish speakers from Antioquia-Colombia using a specific-parameterized discrete wavelet transform analysis[J].Applied acoustics,2021,107635.1-107635.7.

[6] Tripathi, Kumud,Rao, K. Sreenivasa.Robust vowel region detection method for multimode speech[J].Multimedia tools and applications,2021,80(9):13615-13637.

[7] Yadav, Jainath.Detection of vowel transition regions from Hindi language[J].Computer speech & language,2021,7.

[8] F. M. Aloqayli and Y. A. Alotaibi, "Analyzing vowel triangles in spoken Arabic dialects," 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2018, 1-6.

[9] Ramya R , Shanmuga Priya P . SVM based speaker verification system using vowel like and non-vowel like regions[C].International Conference on Communications & Signal Processing. IEEE, 2014:1630-1634.

[10] S. Shahriar and M. N. Hoq, "Evaluation of LPC trajectory for Vowel-Consoant-Vowel sequence," 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, 235-239.

[11] Jianguo Wei, Jianwu Dang. Morphological Normalization of Vocal Tract Shape[C]. The 35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dalas, USA, 2010, 5: 1-10.

[12] Schroeder.M.R, Atal B S, Hall J.L.Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human EarlJl. The Journal ofthe Acoustical of Acoustical Society of America,1979 (6) :1647-1652.

[13] H. Aouani and Y. Ben Ayed, ''Emotion recognition in speech using MFCC with SVM, DSVM and auto-encoder,'' in Proc. 4th Int. Conf. Adv.Technol. Signal Image Process. (ATSIP), Mar. 2018, pp. 1–5.

[14] D. Jurafsky and J. H. Martin, Speech and language processing vol. 3: Pearson London, 2014.