Acoustic Analysis of Monophthongs in Tibetan of Yushu Dialect

Lingzhen Li [1], Yonghong Li [1*]

[1] *Northwest Minzu University, China National Information Technology Research Institute, Lanzhou, China*

Abstract

Based on experimental phonetics, this paper further reveals the acoustic characteristics of monophthongs of Tibetan in Yushu dialect with the help of Adobe Audition 3.0, Praat and other speech analysis software. Firstly, the spectrogram is analyzed and the vowel acoustic parameters are extracted; Secondly, the formant pattern diagram and acoustic vowel diagram of Yushu dialect are drawn by using the values of F1, F2 and F3, which clearly reflect the acoustic characteristics and spatial distribution position of Yushu dialect monophthongs and the relationship between F1, F2, F3 and vowel acoustic characteristics. It is concluded that the lower the tongue position is, the larger value of F1 will be, and vice versa, the smaller value of F1 will be; The more anterior the tongue is, the larger value of F2 is; on the contrary, the smaller value of F2 is; The round lip effect can reduce the F2 value.

Keywords

Yushu dialect, monophthong, experiment phonetics, acoustic analysis

1. Introduction

Located in the southwest of Qinghai province, Yushu region has jurisdiction over six counties: Yushu, Chengduo, Baoqian, Zaduo, Zhiduo and Qumalai. With Tibetan as the main language, it is located at the junction of Wei Zang, Kang and Anduo dialect areas. The overall phonetic appearance presents a transitional feature. Yushu dialect is traditionally classified as Kang dialect [1]. Huang Bufan [2] believes that Yushu dialect has the nature of intermediary dialect or dialect chain due to the influence of the three dialects. Yushu dialect can be regarded as a dialect juxtaposed with the three dialects. Its phonetic features are: the initials system is greatly simplified, and the plosives, affricates and fricative initials have the opposition of unvoiced (aspirated / non aspirated) and voiced [3]; Tones were initially produced to make up for the confusion caused by the disappearance of many phonemes [3]; Rich vowels, with 2 to 5 compound vowels [4].

A monophthong is a vowel with the same tongue position, lip shape and opening degree. It can exist alone in a syllable without other vowels. At present, the research on the vowels of Yushu dialect mainly includes: Huang Bufan's *The phonetic characteristics and historical evolution law of Yushu Tibetan language* [2] thinks that the diversity of the vowel evolution of Yushu dialect is more prominent; Dengzhen Wengmu's *study on the phonology of Tibetan Yushu dialect*[4] mentioned that Yushu dialect has more simple vowels than other Tibetan Languages, and has 2 to 5 compound vowels; Sangta's *phonological study of Tibetan Yushu dialect* [5]shows that most simple vowels in Yushu dialect are the result of the loss and weakening of ancient Tibetan finals.

In this experiment, Yushu is taken as the investigation point. According to the listening and discrimination results, eight monophthongs of Yushu dialect are determined, which are: a、i、e、o、u、ʊ、y、ə. In this paper, the eight monophthongs are described from the three-dimensional spectrogram; Secondly, extracted the acoustic parameters and drawn formant patterns and acoustic vowel diagrams. By exploring the linguistic value of Yushu Tibetan dialect, we hope to provide some reference for the study of more single dialect and lay a certain foundation for phonological description.

ISCIPT2022@7th International Conference on Computer and Information Processing Technology, August 5-7, 2022, Shenyang, China
EMAIL: 825760876@qq.com (Lingzhen Li); lyhweiwei@126.com (Yonghong Li)

Phonological description is an important work of language formal description, so this work is also the most basic work of speech synthesis and recognition.

2. Experimental method

2.1. Experimental materials

The pronunciation vocabulary used in this experiment was selected from the *Tibetan dialect questionnaire* [6], and the pronunciation partners selected 557 commonly used words in the oral language from monosyllabic words. See Table 1 below for examples of pronunciation materials.

Table 1
Yushu Tibetan pronunciation vocabulary

Tibetan	Chinese	IPA	Tonal Category	Tone Pitch
ཨ	Tibetan scrip	ʔa	HA	D51
དཔང	right	oŋ	HC	D51
ཐར	ignited	bar	LB	D131
ཐོམ	crude	bɔm	LC	D131
ཐོང	soak	boŋ	LC	D131
དབུགས	breath	ʊʔ	HE	D55

2.2. Pronunciation partner

The pronunciation partner is a female college student (22 years old) from Northwest University for Nationalities who has clear enunciation, and can speak authentic Yushu dialect without being affected by other dialects. In order to ensure the accuracy of the signal, the partner is required to be familiar with the materials and read each word twice while signal acquisition.

2.3. Voice signal acquisition

The recording was conducted in the professional recording room of Northwest University for Nationalities, with good sealing and sound insulation. Recording equipment includes notebook computer, microphone ecm-44b Lavalier microphone, eurorack ub1204fx-pro mixer, blaster X-Fi surround5.1pro external sound card, etc.; The recording software is Adobe Audition3.0, which adopts single channel recording, with sampling accuracy of 16 bits and sampling frequency of 22050hz. It can complete the recording work with high efficiency and quality, control the recording process, monitor the changes of technical indicators such as speech speed, energy and signal-to-noise ratio, and observe the voice state of the speaker. The recording samples are stored in (*.wav) format.

2.4. Experimental data processing and analysis

After the original speech is preprocessed with Adobe Audition3.0, Matlab is used to cut it into speech files corresponding to a single speech and a name, and Praat speech analysis software is used to mark the voice. When marking, syllables are marked on the first layer, and initials and finals are marked on the second layer, as shown in Fig.1, Praat speech analysis software was used to extract and analyze all acoustic parameters in this study.

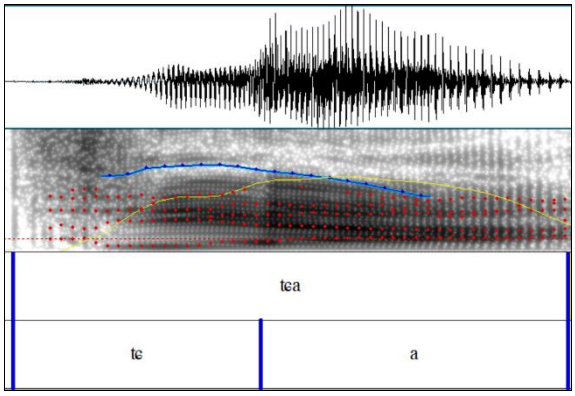

Figure 1: Voice annotation sample

Extraction of acoustic parameters: the vowel tongue position is mainly defined by the frequency of the first formant and the second formant. Therefore, the frequency of these two formants is the main parameter for the study of vowels [7].

3. Analysis of experimental results

3.1. Spectrogram analysis

Vowel is the most important component of voice, which is mainly reflected as formant in acoustics. The formant is the resonant frequency of the sound cavity, which is generally expressed in F, and the corresponding number is used to represent the number of formants. For vowels, F1 and F2 are closely related to the height of the vowel tongue position, the front and back of the tongue position, and the round spread of the lip shape. Therefore, the values of F1 and F2 will be taken as an important basis for describing the acoustic characteristics of vowels in phonetics. Next, select the representative sounds of eight vowels, draw a three-dimensional spectrogram, and show the acoustic characteristics of each category of vowels by analyzing the spectrogram.

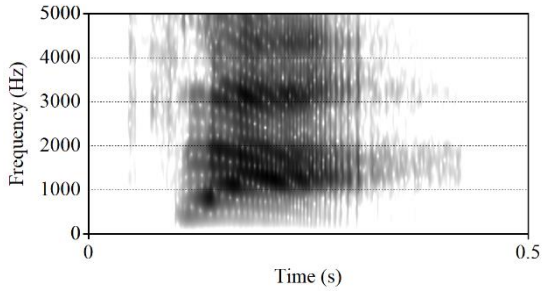

Figure 2: “ʦa” spectrogram

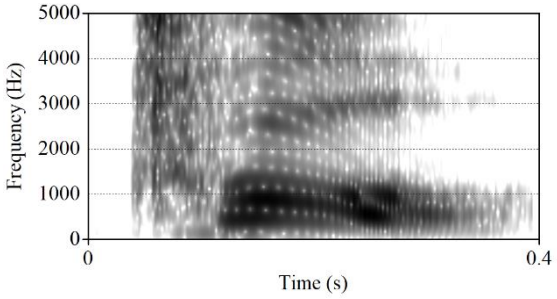

Figure 3: “ʈʂʰo” spectrogram

From spectrogram in Fig.2, the lasting segment energy of the vowel /a/ is strong, and the value of F1 is large. Affected by the initials, the initial segment of F1 is low, and then it rises rapidly and remains stable. Since /a/ channel is divided into front and rear cavities, F1 and F2 are relatively close at 900-1800hz. F3 and F4 are close to each other, and their energy is very strong, about 3000-4000hz. From the spectrogram in Fig.3, the initial parts F1 and F2 of the vowel are around 1000hz, and F3 has a high frequency and weak energy. Affected by the initials, F1 and F2 decrease during the pronunciation

process, while F3 shows an upward trend, and then reaches a stable trend. Comparing the two diagrams, the high frequency energy of /a/ is still very strong, and F1, F2 and F3 are relatively higher. Comparing the two spectrograms, F1 and F2 of the former are higher. Combined with the difference of the size opening of mouth, it is verified that F1 is related to the size opening of mouth (tongue position). The larger the opening, the larger F1.

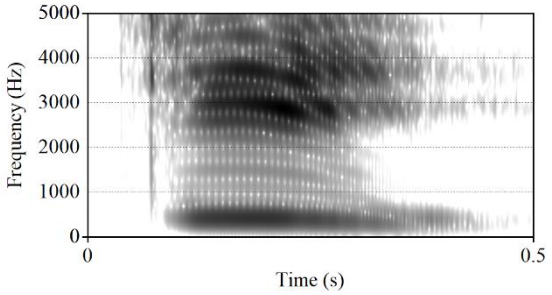

Figure 4: "ʈɕi" spectrogram

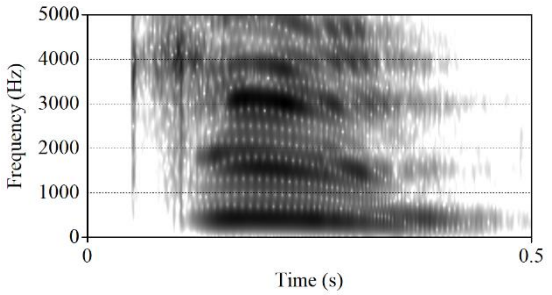

Figure 5: "tsu" spectrogram

In Fig.4, the formant appears at the time of vowel /i/ initial time. From the spectrogram, F1 is relatively small, less than 900hz. F2 is far away from F1 and close to F3. The energy of the three is very strong. Energy concentration at high frequency. Fig.5 shows that the vowel /u/ is relatively stable, and its formant features are: F1 is small, about 500hz, F2 is close to F1, about 1200hz, F3 is large, about 3000hz, and the energy from low frequency to high frequency is strong. Comparing the two diagrams, F1 is lower, but F2 and F3 of vowel /i/ are higher, which is more distant from F1. Combined with the difference between the front and back of the tongue position during pronunciation, it can be verified that F2 is related to the front and back of the tongue position. The more the tongue position is forward, the larger F2 is.

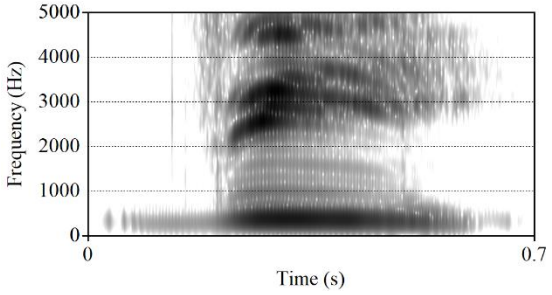

Figure 6: "ʤe" spectrogram

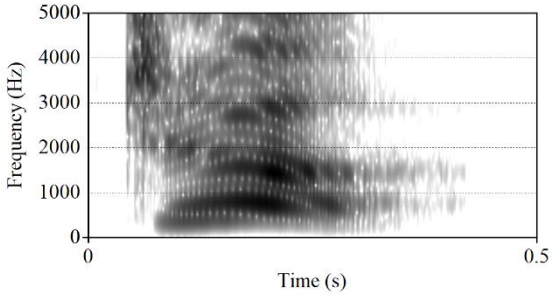

Figure 7: "tʂə" spectrogram

From the spectrogram in Fig.6, F1 of the vowel /e/ is about 400hz, F2 is far away from F1, about 2200hz, F2 is close to F3, the frequency energy is strong, and the distribution is relatively uniform. Compared with /i/, F2 and F3 are lower. Influenced by the front-end initials, F2 and F3 initially point to the low frequency, then rise rapidly and transition to the stable stage. The lowest end in the figure is the energy of fundamental frequency. Fig.7 is " knife " language spectrogram, F1 of vowel /ə/ is relatively high. Influenced by the previous initials, the initial value of F2 is large, and then it drops rapidly, which is very close to F3, about 1500hz. F4 and F5 have high values and relatively small energy.

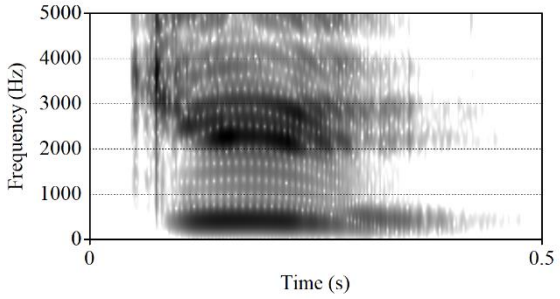

Figure 8: “ʈɕɣ” spectrogram

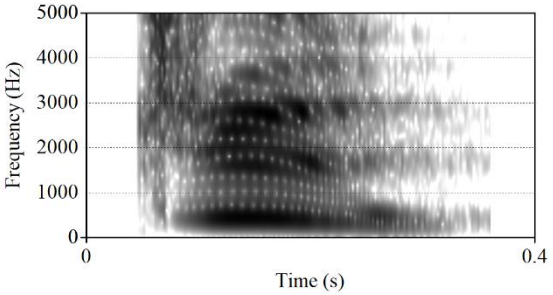

Figure 9: “tʂʊ” spectrogram

According to Fig.8, the distribution of F1, F2 and F3 is similar to that of /i/ of spectrogram. F1 is lower, about 300-400hz, which is related to the higher tongue position during pronunciation. However, since the tongue position is front, F2 is very high, but the F2 of /y/ is still smaller than /i/, about 2100hz. Fig.9 shows the "wash" spectrogram of Yushu dialect. The observation of spectrogram shows that F1 is about 400hz. F2 is close to F3, about 2000hz. In comparison with / ʊ/, the formant distribution is relatively uniform.

3.2. Vowel formant pattern of Yushu dialect

Drawing different vowel formants into a formant pattern diagram is conducive to observing the formant corresponding pattern between vowels, and can more vividly see the location and relationship of each vowel formant. After extracting the acoustic parameters of vowels in voice samples and averaging them, the frequencies of the first three formants F1, F2 and F3 of the eight monophthongs are obtained respectively, with vowel as the abscissa and the frequencies as the ordinate, and draw the formant pattern spectrogram of Yushu dialect, as shown in Fig.10:

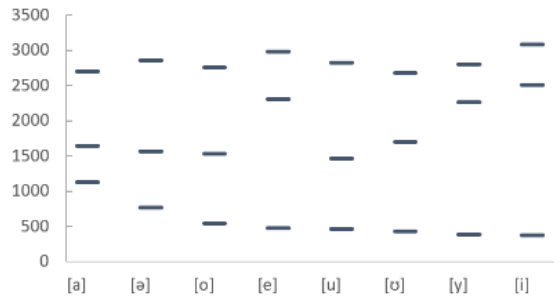

Figure 10: Vowel formant pattern

From the formant pattern of Yushu dialect, we can clearly see that each monophthong has its own formant distribution characteristics. It mainly shows that F1 and F2 are different in value and relative distance. According to the above Fig.10, F1 and F2 of /i/ are the largest, followed by /y/, the distance between /a/ is the smallest, followed by /ə/.

F1 values from small to large are:

/i/</y/</ ʊ/</ u/</e/</o/</ ə/</ a/；

F2 values in descending order are: /i/</e/</y/</a/</ ʊ/</ə/</ o/</u/.

It can be found that F1 and F2 values roughly form an inverse relationship. However, there are exceptions. For example, for the two monophthongs /e/ and /y/, the F1 value of /e/ is greater than /y/, but the F2 value is also greater than /y/, and they do not form a strict inverse proportional relationship. Considering that F2 is also related to the round spread of lip shape, that is, the round lip effect can reduce the F2 value, because the round lip effect and the back position of tongue can make the front resonant cavity larger when pronouncing.

3.3. Acoustic vowel diagram of Yushu dialect

The acoustic vowel diagram is different from the traditional vowel tongue bitmap. It is obtained according to the objective values of F1 and F2. At the same time, F1 is the vertical coordinate and F2 is the horizontal coordinate. The coordinate origin is set in the upper right corner, making its relative position roughly the same as that of the traditional vowel tongue bitmap. Jos (1948) [8] believes that although the formant frequencies of the same vowel uttered by different people are different, the relative positions of each vowel on the acoustic vowel map are stable. The position of each vowel in the Fig.11 is obtained by averaging the formant frequencies of all samples of each vowel.

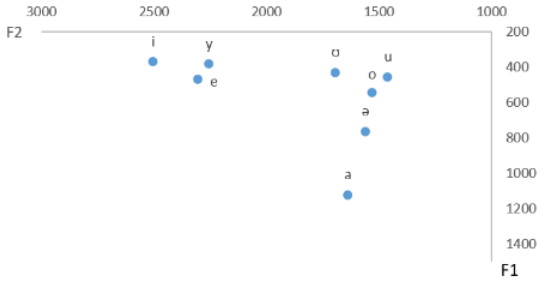

Figure 11: Acoustic vowel diagram of Yushu dialect

As shown in the Fig.11, the whole distribution identified by three vertex vowels /i, u, ɑ/ of the acoustic vowel map of Yushu dialect can be determined. /i/ is the front high unrounded lip vowel, /a/ is the central rear low unrounded lip vowel, /u/ is the rear high rounded lip vowel, thus forming a triangular region. By observing the above figure, the first formant frequency of Yushu vowel is between 300-1200hz, and the second formant frequency is between 1300-2600hz. In the vertical direction, the lowest vowel is /a/, /ə、o、e/ are in the middle position, /i, y, u, ʊ/ are in the highest area; In the horizontal direction, /i, y, e/ are at the front, / a、ʊ/ are at the back of the center, /u, o ə / are located at the rear.

The above figure shows more clearly that F1 is related to the height of tongue position: the vowel tongue position of /a/ is the lowest, F1 is the highest, the tongue position of /i, u, y ʊ/ vowels is higher, but the F1 value is smaller; F2 is related to the front and back of the tongue position: the vowel tongue position of /i/ is the most front, and the F2 value is the largest. The later the tongue position is, the smaller the F2 value is. It can be seen from the figure that when the vowels of Yushu dialect are pronounced, the eight monophthongs are concentrated and scattered in the three vertex areas, and the monophthongs in the central area of the tongue are less distributed.

4. Summary

The eight monophthongs of Yushu dialect are: a、i、e、o、u、ʊ、y、ə. /a/ is the central back low unrounded vowel, /i/ is the front high unrounded vowel, /u/ is the rear high rounded vowel, /y/ is the front high rounded vowel, /o/ is the rear medium high rounded vowel, /e/ is the rear medium high unrounded vowel, /ʊ/ is the middle high rounded lip vowel behind the center. /ə/ is the second half of the high unrounded lip vowel. The distribution of formants was consistent: the higher the tongue position was, the smaller the F1 value was; the lower the tongue position was, the larger the F1 value was; The more anterior the tongue is, the greater the F2 value is. The more posterior the tongue is, the smaller the F2 value is; In the same case, the round lip effect can reduce the value of F2.

The tone of Yushu dialect is very special. Through the analysis and research of its pronunciation, it can supplement the blank of the other three major Tibetan dialects and the world tone language. At the same time, the acoustic analysis of Yushu dialect using experimental phonetics can promote the development of phonetic information and visualization. It is hoped that with the development of science and technology, computer technology and digital signal analysis technology can be more and more applied in phonetics, and promote the further development of phonetics to fill the shortcomings of traditional phonetics.

5. Acknowledgements

This work was financially supported by NSFC grant fund (No.11964034) and Research and innovation Projects (No.2021CXZX-674).

6. References

[1] Peng Jin, Tibetan Jianzhi, 2nd. ed., Ethnic Publishing House, 1983.
[2] Bufan Huang, Suonan jiangcai, Minghui Zhang, Phonetic characteristics and historical evolution of Yushu Tibetan language, Chinese Tibetology, (1994) (2) 24.
[3] Anseraga, A survey of Tibetan Yushu dialect (Labu) phonology, Tibet studies, (2018) (1) 7.
[4] Dengzhen Wengmu, Phonological study of Yushu dialect in Tibetan, Henan science and technology, (2015) (22) 1.
[5] Sangta, Phonological study of Yushu dialect in Tibetan, Master's thesis, Northwest University for Nationalities, 2012.
[6] Jiangping Kong, Tibetan dialect questionnaire, 2nd. ed., Commercial Press, 2011.
[7] Yasheng Jin, Ruishan Zhang, A study on the unit sound acoustics of Dongxiang language , Northwest ethnic studies, (2010)(4)10.
[8] Joos,M. Acoustic Phonetics,Language, 2nd. ed., No.24, (suppl.2).
[9] Gesang Jumian, Gesang Yangjing, Introduction to Tibetan dialect, 2nd. ed., Ethnic Publishing House, 2002.
[10] Jiangping Kong, Basic course of experimental phonetics, 2nd. ed., Peking University Press, 2015.