Acoustics Analysis of Plosive in Zhongdian Tibetan

Jing Zhang[a, *], Yonghong Li[b]

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

[a]274159100@qq.com, [b]lyhweiwei@126.com

*Corresponding author

Keywords: Zhongdian Tibetan, plosive, acoustic analysis

Abstract: In this paper, the plosive system of Zhongdian Tibetan of Kang dialect is taken as the research object, and the acoustic experimental method is used to extract, analyze and summarize the distribution of its sound intensity and duration, etc, and describe its acoustic characteristics, which provides reference for future research in related aspects.

1. Introduction

Zhongdian (Shangri-La) is located in the grand triangle region of Yunnan, Sichuan and Tibet in the northwest of Yunnan Province, and is located in the heart of Shangri-La, Diqing. Zhongdian is a multi-ethnic community, consisting of more than 20 ethnic groups, of which the main ethnic group is Tibetan and the main language is Tibetan. As one of the most representative dialects of Tibetan in Yunnan, Zhongdian Tibetan belongs to Nanlu dialect of Kang-Tibetan. Its voice is characterized by the presence of voiced plosive, fricative, affricate and tone. In addition, compared with the Wei and Amdo Tibetan, the Kang is in the simplification stage of the phonetic system, so the Zhongdian Tibetan language also presents the instability and complexity of the system.

In recent years, the research on the pronunciation of Kang-Tibetan dialect is less than that of Wei and Amdo, mainly focusing on Dege dialect, for example, Professor Gesang Juming (2001) combed the phonological system of Dege dialect, and Qingcuo (2015) described the tone category and tone value of Dege dialect. In terms of the phonetics of Zhongdian Tibetan, Zhao Jincan summarized the characteristics of Zhongdian Tibetan from ten aspects, such as whether there is a voiced initial, whether there is a devoiced initial, whether there is a compound consonant initial, and the phonetic variation in his article "A Comparison of the Lhasa Tibetan and Shangri-La Tibetan Phonetics", and summarized the phonetic system of Zhongdian Tibetan. In his subsequent paper "Tone experiment in Jiantang Tibetan", he determined the tone contour, tone value and tone type of Zhongdian Tibetan language through computer experiment analysis, and at the same time made a brief discussion on its historical origin. Related researches also include Suzuki Hiroyuki's description and analysis of the origin of the nasal consonant system in his thesis "The nasal initials of Yalang dialect in the Shangri-La Tibetan".

This paper uses the method of experimental phonetics to do some basic research on the plosive system of Zhongdian Tibetan, hoping to provide basic reference for the future Tibetan speech recognition, speech synthesis and Tibetan language teaching.

2. Plosive system of Zhongdian Tibetan

The coda consonants in Zhongdian Tibetan have almost fallen off, so the plosive system mentioned here is the pre-plosive (plosive initial) system. There are 16 single plosive initials in Zhongdian Tibetan, from the point of place of articulation, they are divided into bilabial consonants, alveolar consonants, palatal consonants and velar consonants. From the point of manner of articulation, there is a three-point pattern of voiceless unaspirated sound, voiceless aspirated sound, and voiced sound. In addition, there are four compound consonants that composed of plosive and nose crown sound in the same place of articulation. The list of plosive system is as follows:

p	p^h	b	mb
t	t^h	d	nd
c	c^h	ɟ	ɲɟ
k	k^h	g	ŋg

This is different from the plosive system summarized by Lu Shaozun and Zhao Jincan. Both of them did not list the palatal consonant /c/ group, but from the situation of listening to the notes, there are many palatal plosive example words, 22 in total, which are not in opposition to the conditions of the velar consonant /k/ group, so the /c/ group is an independent phoneme.

3. Experimental Method

3.1 Pronunciation material

The pronunciation material of this paper is the "Zhongdian Tibetan Initials Pronunciation Table", which is designed according to the "Survey of Tibetan Dialect" and combined with the spoken Tibetan of Zhongdian. The table contains 570 Tibetan monosyllabic words, among which 153 monosyllabic words have the consonants of "plosive". Example of pronunciation material is shown in the following table:

Table 1. Zhongdian Tibetan Initials Pronunciation Text

Number	Tibetan	Latin transliteration	Chinese meaning	The phenotic symbol
1	ལྐུགས	lkugs	哑巴	ku
2	ལྐོག	lkog	悄悄地	ku
3	སྐམ	skam	肉牛;干	kaŋ
4	སྐད	skad	声;声音;	ce
5	སྐེ	ske	颈项;脖子;	ca
6	སྐུར	skur	寄发;带给	ku

3.2 Articulator

The articulator of this experiment is a middle-aged male (45 years old), who is in good health, without comprehension barriers and throat diseases. He can speak a native Zhongdian dialect, has not been influenced by other Tibetan dialects.

3.3 Sound Recording

The recording devices used in the experiment are notebook computer, clip-on microphone, external sound card, mixer, etc. The recording software is Adobe Audition3.0, which is used to collect the sound pressure wave and glottic impedance wave simultaneously through dual-channel. The sampling rate is 22050Hz, and the recording samples are stored in (*.wav) format. During the recording process, the articulator was required to read each word twice. The phoneme is recorded according to the articulator's listening and answering.

3.4 Analysis of Experimental Data

The annotation and analysis of speech samples were performed using praat speech analysis software. When labeling, the consonants and finals of the monosyllable are marked in a layer. The acoustic parameters extracted in the experiment mainly include duration, intensity, LTAS and so on.

Intensity is the intensity of sound, measured in decibels (dB). The measurement position of the stop sound intensity is usually selected on the release burst.

Duration is the length of the initial segment when the speech sample is marked, which represents the duration of the sound wave of the speech, and often plays a role of distinguishing features in consonants.

The FFT power spectrum is a long-time average of the short-time spectrum of each frame calculated by the fast Fourier transform method. The FFT power spectrum can intuitively reflect the pitch period and the energy, frequency of the harmonic, and can also reflect the channel formant corresponding to the peak with higher harmonic energy.

4. The Experimental Results

Except for the unique acoustic parameters of the first half of the nasal sound, the second half of the compound consonant is basically the same as the single consonant. Therefore, we only analyze the relevant acoustic parameters of the single consonant in this experiment.

4.1 Acoustics Description of Sonogram

The stop and vowel separate ends of the speech continuum. As a typical instantaneous sound, stop is shown as pulse waves with small amplitude on the waveform diagram, and as a release burst with a peak on the spectrogram, which is formed by the instantaneous breakthrough of air flow. Different manner of articulation will also be reflected on the sonogram, we choose the three stop sounds /p/, /pʰ/, and /b/ in Zhongdian Tibetan that are all bilabials to acoustically describe its sonogram.

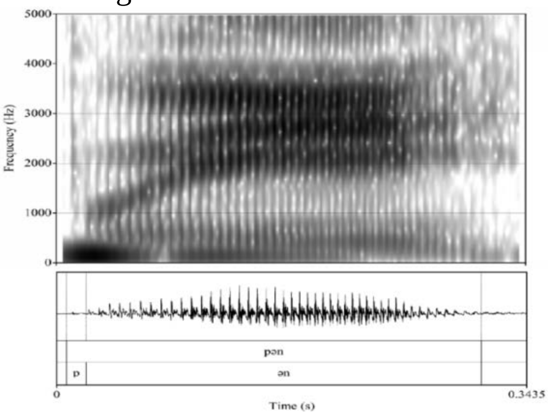

Figure 1. Sonogram of /p/

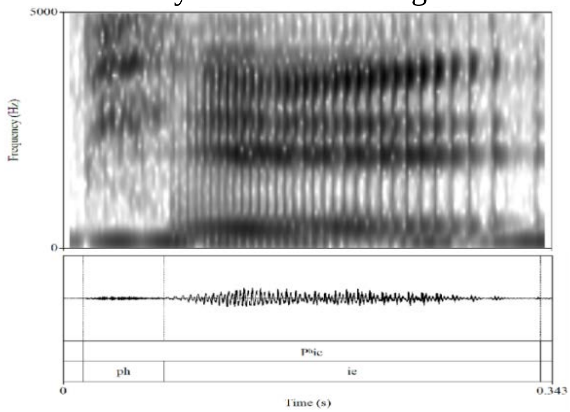

Figure 2. Sonogram of /p/

/p/ is an unaspirated voiceless stop sound with a short resistance removal period. The vowel followed immediately after the release of the stop closure and the VOT length is approximately zero.

/pʰ/ is aspirated voiceless stop sound. It can be clearly seen in the spectrogram that there is there is a much longer gap (about 54ms) between the release of the stop closure and the onset of voicing for the following vowel. The gap is aspirated segment.

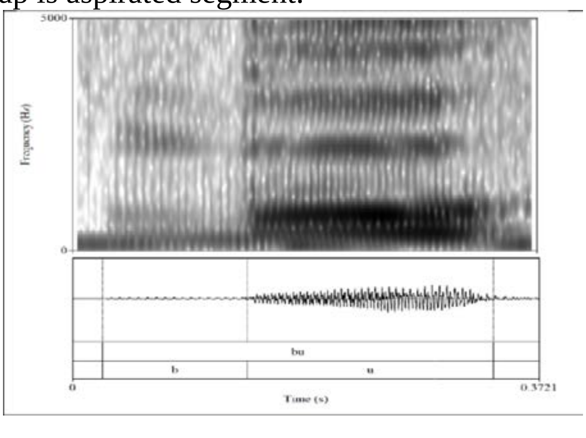

Figure 3. Sonogram of /b/

/b/ is a voiced stop, the waveform of the consonant segment has a periodic waveform with a very low frequency. In spectrogram, a low-frequency voice bar is seen, which indicates that the vocal cords have begun to vibrate before the release of the stop closure, and the VOT is negative.

4.2 Duration Analysis

In order to avoid errors caused by the influence of extreme values, the median is used to compare and analyze the length of the Zhongdian Tibetan plosive. Comparing the statistical data of all stops in Zhongdian Tibetan, the duration is arranged in order from long to short: /ɡ/ > /cʰ/ > /kʰ/ > /b///d/ > /pʰ//tʰ/ > /ɟ/ > /c/ > /k/ > /t/ > /p/.

In order to more intuitively observe the distribution of each stop sound, moreover, the relationship between the duration and the articulation position and method, we classify the 12 stops according to different articulation position and methods, and draw a bar chart, as follows:

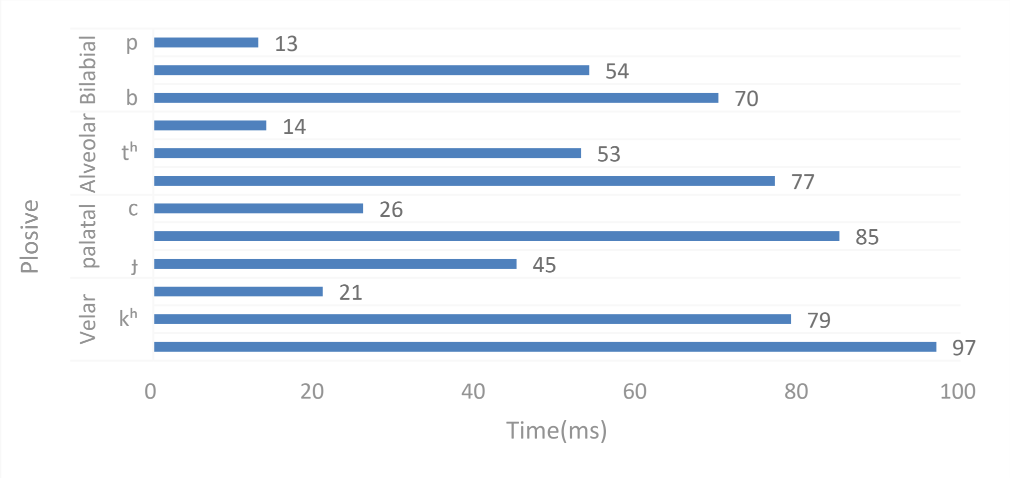

Figure 4. Duration Distribution Diagram of Plosive

From the perspective of the articulation position, in the unaspirated voiceless group, the palatal stop /c/has the longest duration, followed by the velar stop /k/, /p/ and /t/ have the similar duration. In the voiceless aspirated group, the longest is the palatal stop /cʰ/, followed by the velar stop /kʰ/, and the duration of /pʰ/and /tʰ/ is equivalent and shortest. In the voiced group, the length of the velar stop/g/is the longest, followed by /b/ and /d/, and the palatal stop /ɟ/ is the shortest. In general, for voiceless aspirated and voiceless aspirated plosives, the duration of palatal and velar consonant is much longer than that of alveolar and bilabial consonant. This shows that the duration is closely related to the position of articulation, and the more forward of the position, the shorter the duration is. In addition, it can be seen that the flexibility of the vocal organ and the contact area of blocked parts will also affect the duration. The duration of the /c/ group is longer because the position of the palatal sound is the hard palate and the lingual surface, and the active contact area with the passive vocal organ is larger. In the process of pronunciation, the tongue surface is less flexible than the lips, the apex of linguae, and even the root of tongue, it is more difficult to remove the obstruction, which results in a longer time.

From the perspective of articulation method, among the bilabial sounds, the voiced stop /b/ has the longest duration and the voiceless unaspirated stop /p/ has the shortest duration. In the alveolar, voiced stop /d/ has the longest length, followed by /tʰ/ and /t/. Among the palatal sounds, the longest duration for the voiceless aspirated/cʰ/, then the /ɟ/, voiceless unaspirated plosive [c] is the shortest. In the velar, the longest is the voiced topper /ɡ/, and the shortest is the voiceless aspirated stop /k/.According to the comprehensive analysis, except for a group of special ones in the palatal, the relationship between the plosive and duration of different articulation methods in Zhongdian Tibetan is voiced sound > voiceless aspirated sound> voiceless unaspirated sound. The voiced stop has a longer band before the outbreak, which is an all-voiced plosive.

4.3 Intensity Analysis

In this experiment, statistical analysis is performed on the average, maximum and minimum sound intensity of different stop sounds in Zhongdian Tibetan. The specific data are as follows:

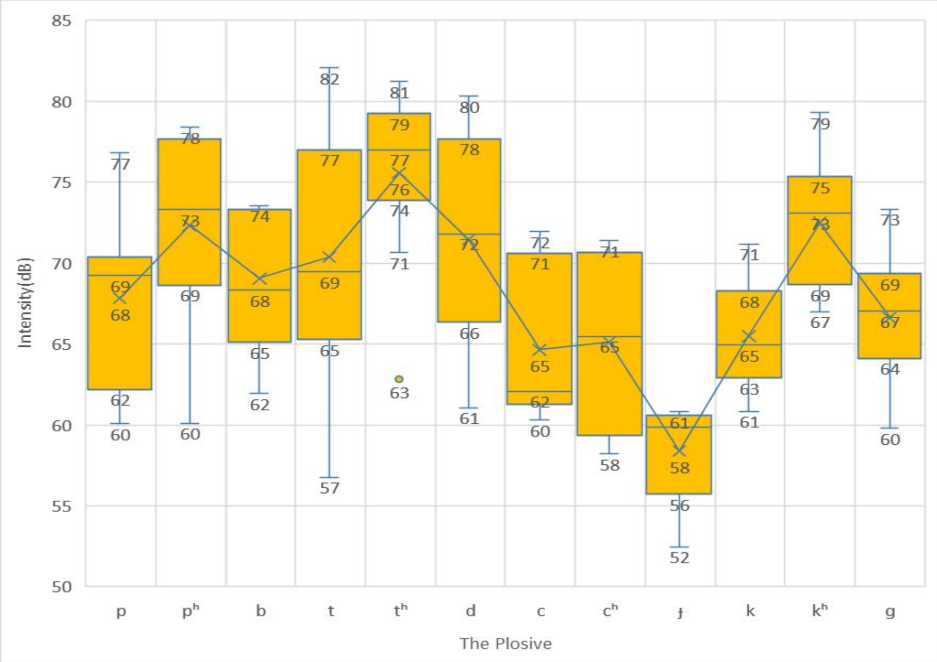

Figure 5. Contrast Chart of Plosive Intensity

As an auxiliary parameter of consonant acoustic characteristics, sound intensity is closely related to articulation method. Wu Zongji and Lin Maocan (1989) summarized the general trend of consonant sound intensity in their research: one is that the intensity of voiceless consonant is greater than the voiced consonant; the other is that the intensity of aspirated sound is greater than the aspirated sound. We can see from the above chart whether the Tibetan stop in Zhongdian accords with this trend.

As can be seen from the figure, in the four different position of articulation, the sound intensity of the voiceless aspirated stop is greater than that of the same position of the voiceless unaspirated stop and voiced stop, in line with the general consonant sound intensity characteristics of "the intensity of aspirated sound is greater than the unaspirated sound". On the other hand, there is no obvious difference in the intensity of unvoiced and voiced sounds in the same position. In addition to the palatal, the voiced sounds in other position are even stronger than the unvoiced sound. This shows that in Zhongdian Tibetan, whether the vocal cord vibration affects the intensity of the plosive is not absolute. The voiced stop in Zhongdian Tibetan is a "hard plosive", the muscles are tense and the rupture is strong in pronunciation.

From the point of articulation position, in Zhongdian Tibetan, the intensity of stop sound that at the rear of vocal area is lower than the stop sound that at the front of the vocal area. In the case of equal air flow, the plosive that at the rear of vocal area has a short distance from the power organ to the blasting organ, so force of the articulator is small, which affects the sound intensity.

4.4 Spectrum Analysis

Each consonant has its characteristic part, which is shown as the energy concentrated frequency area on the sound spectrogram. Due to the different position of articulation, the concentrated frequency areas of each stop is also different. In this experiment, the energy distribution pattern of different plosive is observed by plotting the FFT power spectrum of the central segment of the release burst.

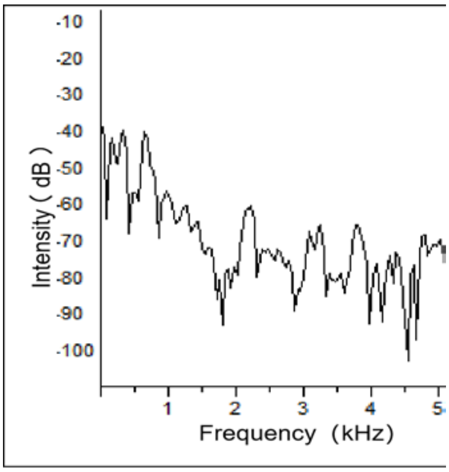

Figure 6. The LTAS of /

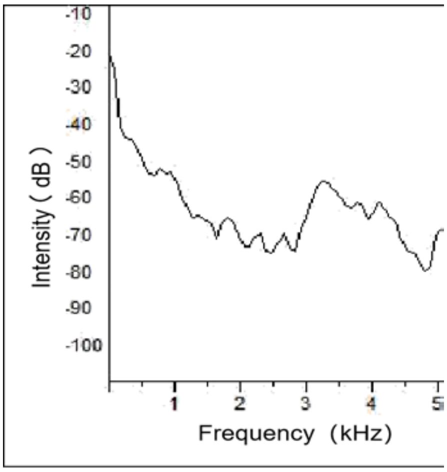

Figure 7. The LTAS of /t/

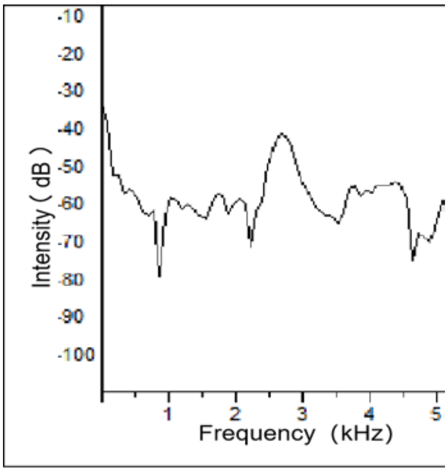

Figure 8. The LTAS of

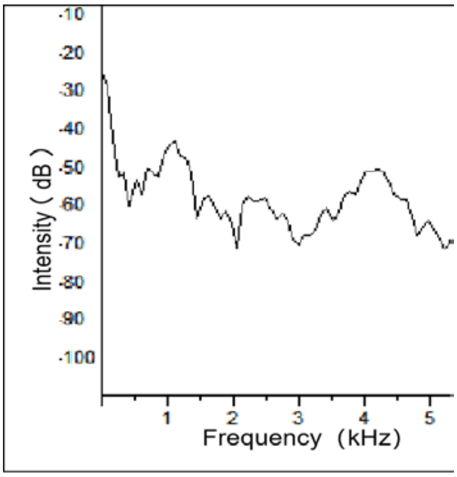

Figure 9. The LTAS of /k/

Fig.7-fig.10 respectively show the FFT power spectrum of four voiceless unaspirated stops /p/ /t/ /c/ /k/, and the power spectrum of other stoppers is not listed here. The frequency domain graph can reflect the distribution of the energy concentrated frequency of different stop sounds.

/p/ /pʰ/ /b/, these three bilabial plosives' spectrum energy is a steep decline, energy mainly concentrated in the low frequency band of 0-1500Hz. Among them, /p/ and /b/ showed two rising fluctuations in the energy between 2000-3000Hz and 3000- 4000Hz, and /pʰ/ only occurred in the energy rising in 3000-4000Hz.

To alveolar sound /t/ and /tʰ/, the energy energy curve fell more gently, and their energy concentration area appears at high frequencies above 3000Hz. specifically, the energy concentration area of /t/ is between 2000-2200Hz, /tʰ/ is between 3500-4000Hz. There are two energy concentration areas of /d/, which are between 500-1000Hz and 3500-4000Hz. The energy of the alveolar plosive is higher than that of the bilabial plosive.

The first part of the spectrum energy of the palatal /c/ /cʰ/ /ɟ/ is relatively stable, with an obvious peak value at 2000-3000Hz. The peak value of /c/ in appears in the 2000-3000Hz band, the peak value of /cʰ/ appears in the 3000-4000Hz band, the peak value of /ɟ/ appears in 2500-3000Hz band, and the energy of /c/ and /cʰ/ decreases significantly at about 1000Hz.

The velar /k/ /kʰ/ /g/ does not show a significant downward trend in the energy curve, there are two obvious energy concentration areas in the low frequency and high frequency areas. The first energy concentration area appears in /k/ at 700-1200Hz, there is a small increase in energy at 2000-3000Hz, and the second energy concentration area appears at 3800-4200Hz; at 1000-1500Hz, /kʰ/ appears an energy concentration area appears, and there is a small increase in energy at 3500-4000Hz. The energy concentration area of /g/ appears at 1400-1500Hz, and the energy increase zone is not obvious, there is no peak in the high frequency zone.

5. Conclusion

Through the collation of the Zhongdian Tibetan phonetic system and analysis of relevant acoustic parameters, we can draw the following conclusions:

There are 20 stop sounds in Zhongdian Tibetan, among which 16 are single consonants and 4 are compound consonants. The position of articulation involves bilabial, alveolar, palatal and velar, the method of articulation involves voiceless unaspirated, voiceless aspirated, and voiced. The compound consonant system has been greatly simplified, mainly the plosive plus the nasal crown sound of the same part.

The duration of the stop sound is closely related to the place of articulation, the more forward of the position, the shorter the duration is. From the perspective of articulation method, the voiced consonant in the same position have the longest duration, followed by the voiceless aspirated ones, the shortest is the voiceless aspirated consonant. The voiced plosive is an all-voiced plosive.

The sound intensity of the plosive is more closely related to the articulation method, which is mainly manifested as that the intensity of the voiceless aspirated stop is greater than that of the voiceless unaspirated stop and the voiced stop in the same position. There is also a relationship between sound intensity and the place of articulation, that is, the sound intensity of the stop at the back of the place of articulation is generally less than that of the stop at the front of the place of articulation.

Acknowledgments

This work was financially supported by NSFC grant fund (No.11964034).

References

[1] Kong Jiangping, R. A. Lupton, Basic course of experimental phonetics, Beijing, 2015.

[2] Wu Zongji, Lin Maocan, Summary of experimental phonetics, Beijing, 1989.

[3] Gesang Juming, Introduction to Tibetan dialect, Beijing, 2002.

[4] [Zhao Jincan, A comparison between the Lhasa Tibetan and the Shangri-La Tibetan, Journal of Sichuan University for Nationalities, 2010.

[5] Wang Wenquan, Acoustic research of consonants of Amdo Tibetan in Tianzhu, Northwest Minzu University, 2016.