Acoustic Analysis of Plosives in Yushu Dialect Based on Experimental Phonetics

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

[1]Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Menzu University Lanzhou, China
[a]825760876@qq.com, [b*]lyhweiwei@126.com

ABSTRACT

This paper takes the plosives of Tibetan Yushu dialect as the research object, extracts and analyzes its acoustic parameters such as sound intensity and duration by means of experimental phonetics, and summarizes its acoustic characteristics. Firstly, the experiment uses the recording software Audition3.0 single channel adopts voice signal, in which the sampling accuracy is 16 bits and the sampling frequency is 22050Hz; Secondly, the voice cutting program of Matlab is used to generate a voice signal corresponding to a single name and voice; Praat speech analysis software is used for the extraction and analysis of all acoustic parameters in this experiment. If necessary, the self-written program of this laboratory shall be modified appropriately to meet the needs of this experiment. After analysis and induction, it is concluded that the acoustic parameters extracted by the acoustic experiment, such as VOT, sound intensity and duration, can well distinguish and describe the plosives of Yushu dialect.

Keywords: *YuShu dialect, plosives, experimental phonetics, acoustic analysis*

1. INTRODUCTION

Yushu is located in the south of Qinghai, with Tibetan as the main language and at the junction of the three dialect areas of Wei Zang, Kang and Ando. The overall phonetic appearance presents a transitional feature and is a living fossil of Tibetan language research[1]. Yushu dialect is traditionally classified as Kang dialect. Its phonetic features are: the initial consonant system is greatly simplified, and there are clear (aspirated / non aspirated) and voiced opposites among plosives, affricates and fricative initials; Rich vowels, with 2 to 5 complex vowels; Tones are initially produced to make up for the confusion caused by the weakening and disappearance of many phonemes.

The research on the pronunciation of Yushu dialect is less than that of Weizang and Anduo dialects, mainly including Huang Bufan's (1994) exploration of the phonetic characteristics and historical sound change law of Yushu dialect, which believes that the diversity of vowel evolution is more prominent [5]; Sangta and Dawa pengcuo (2010) analyzed the phonetic features of Yushu dialect [8]; Deng zhenwengmu (2015) studied and summarized the phoneme system of Yushu dialect, and mentioned that most dialects in Yushu dialect have voiced nasal crown initials, and there may be separate voiced initials[12]; Anseraga (2018) investigated and studied the transitional nature of the tone system of Yushu dialect, and believed that the complex consonant system and rhyme tail system of Yushu dialect have been greatly simplified, and tone, as a late rising phenomenon, reflects a transitional phonological feature [1].

Plosives are a class of consonants. When pronouncing, the pronunciation part is blocked, the soft palate rises, blocking the nasal passage, and the air flow breaks through the obstruction, burst out and burst into sound. There are 10 plosives in Yushu dialect, which can be divided into double lip, tip of tongue, root of tongue and laryngeal plosive; From the perspective of pronunciation method, it presents the opposition of clear aspirated sound, no aspirated sound and voiced sound According to our hearing, it is recorded as:

P、p^h、b；
t、t^h、d；
k、k^h、g；ʔ

In order to fully tap the linguistic value of Yushu Tibetan dialect and provide some reference for the description and research of more single point dialects. This paper makes a systematic analysis of four groups of plosives in Yushu dialect of Tibetan from the aspects of

A. El-Hashash et al. (Eds.): IEIT 2022, ACSR 100, pp. 721–728, 2023.
https://doi.org/10.2991/978-94-6463-058-9_113

VOT, sound intensity, duration and straight bar spectrum. As a laryngeal plosive, ʔ has special properties and will be listed separately.

2. EXPERIMENTAL METHOD

2.1. Experimental Materials

The pronunciation vocabulary used in this study is selected from the *Tibetan Dialect Questionnaire* [7]. The pronunciation partners select 557 monosyllabic words commonly used in the export language from monosyllabic words. Examples of pronunciation materials are shown in Table 1 below.

Table 1: Yushu Tibetan pronunciation vocabulary.

Tibetan	Chinese	IPA	Tonal Category	Tone Pitch
ཨ	Tibetan scrip	ˀa	HA	D51
དབང	right	oŋ	HC	D51
སྦར	ignited	bar	LB	D131
སྦོམ	crude	bɔm	LC	D131
སྦོང	soak	boŋ	LC	D131
དབུགས	breath	ʊʔ	HE	D55

2.2. Pronunciation Partner

The pronunciation partner is a female college student (22 years old). She has clear speech and no obvious pronunciation defects. She can speak authentic Yushu dialect and is not affected by other dialects. In order to ensure the accuracy of the signal, the speaker is required to be familiar with the text before signal acquisition, so that the speaker can pronounce clearly and sound roundly.

2.3. Speech Signal Acquisition

The recording was conducted in the special recording room of Northwest Minzu University, which has good sealing and sound insulation. Recording equipment includes laptop, 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. Single channel recording is adopted, 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

First, use Adobe audition3 0 pre-processes the original voice and divides it into a single voice file with MATLAB. Praat speech analysis software is used to extract and analyze all acoustic parameters in this study. The acoustic parameters extracted from the experiment include duration, sound intensity, spectrum and so on.

Duration (Ms) is the length of the initial segment when the speech sample is labeled.

The sound intensity (dB) is the intensity of the sound, and the measurement position of the plosive sound intensity is usually selected on the straight bar.

VOT (voice onset time) is the voice onset time, which refers to the time from the plosives burst to the vocal cord vibration onset.

Spectrum is to analyze the relationship between frequency and energy in the frequency domain.

3. ANALYSES OF EXPERIMENTAL RESULTS

3.1. Spectrogram Analysis

In the process of plosive blasting, there is a high energy from low frequency to high frequency, which is commonly referred to as "straight strip"[7]. The sequential relationship between blasting and the starting time of vocal cord vibration is an important acoustic parameter to distinguish different types of plosives. We use double lip plosives p、p^h and b as the representative of three pronunciation methods: clear and no aspirated plosives, clear and aspirated plosives and voiced plosives. When labeling, the first layer labels the syllable, and the second layer labels the initial segment and vowel segment of the syllable.

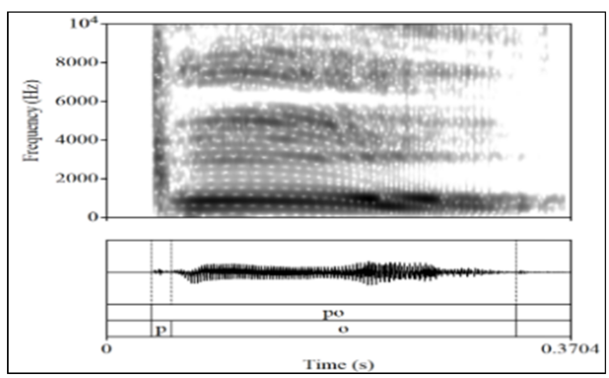

Figure 1: po ("person" suffix).

As can be seen from Figure 1, there is no obvious front voiced segment of the clear aspirated plosives p. the burst segment starts with a straight bar, followed by the formant of the vowel. Therefore, VOT is basically equal to zero or a little positive. The frequency of F1 and F2 of vowel o is low and the energy is strong, while the frequency of F3 is high but the energy is weak. Affected by the plosive p, the starting section of formant points to the low frequency, and then quickly transitions to the stable section.

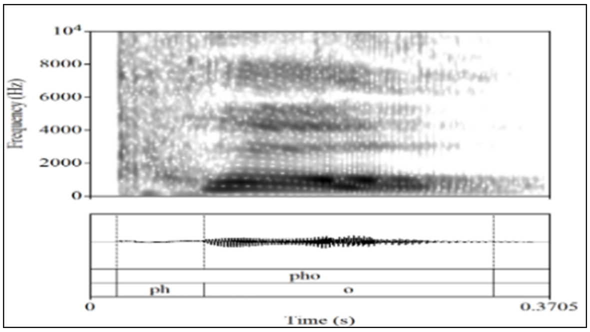

Figure 2: p^h o (splash)).

The consonant part of Figure 2 is the clear aspirated plosive p^h。 From the diagram, the blasting section of the plosive p^h starts with a straight strip followed by an aperiodic waveform, which is a disordered pattern. It is the air supply stage after blasting. The disordered pattern energy in both low-frequency and high-frequency air supply sections is very uniform. The straight bar appears before the vocal cord vibration, and there is no obvious voiced segment, so the VOT is positive and the value is large.

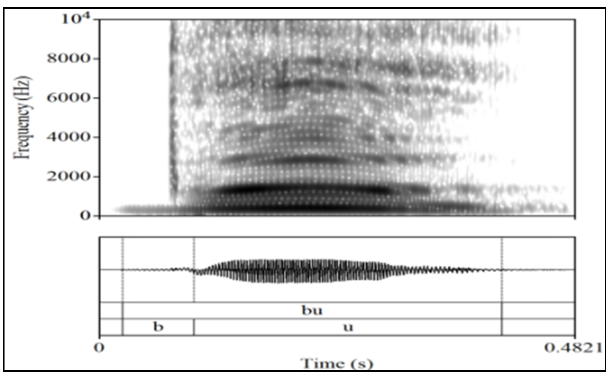

Figure 3: bu (swelling).

As can be seen from the "swelling" language diagram in Figure 3, the voiced plosives b has a low-frequency energy distribution before blasting, which is generated by vocal cord vibration. The acoustic characteristics are voiced bars, indicating that the vocal cords vibrate first when voiced plosives are issued. However, due to the small frequency range of the front voiced section, it will not exceed 500Hz, so the energy is mainly distributed in the low-frequency region. Then the air flow breaks through the obstruction and bursts into sound. Therefore, it can be seen that the voiced plosives VOT is negative.

In the process of observing the language map, it is found that there are often double blasting or multiple blasting phenomena, which are shown as two or more straight bars on the language map, and mainly occur in the root plosivepers of tongue. This is due to the large blocking area of the root of tongue and poor flexibility. Sometimes, the air flow rushes out from two or more points.

3.2. *VOT Analyses*

First of all, from the above diagram, we can intuitively see that the energy of plosive sound blasting is very strong, and there is no sharp decline with the increase of frequency, which is why we can see the whole straight bar in the diagram; Secondly, from the above diagram analysis, we can see that there is a close relationship between the plosive VOT of Yushu dialect and the pronunciation method. In order to more clearly express this law, we present the obtained data in the form of charts (taking the clear plosive as an example), as shown in Table 2 and figure 4 below.

Table 2: Absolute value of plosive VOT duration.

VOT(Ms) / Plosives	Max	Mix	Mean	SD
[p]	23.79	12.65	16.49	4.01
[t]	23.27	14.31	19.04	2.96
[k]	25.00	17.74	20.68	2.56
[pʰ]	73.48	29.94	54.13	15.76
[tʰ]	67.96	36.60	56.24	10.27
[kʰ]	85.06	52.97	69.37	9.36

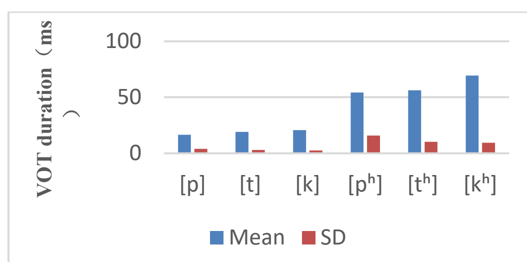

Figure 4: Comparison diagram of VOT duration of clear plosives.

It can be summarized as the following rules:

a) Figure 4 shows that the VOT of aspirated sound is longer than the corresponding non aspirated sound, which is shown as p^h> p、t^h> t、k^h> k。 The mean value of VOT can clearly distinguish the aspirated and non-aspirated plosives in Yushu dialect. It can also be seen from the table that no matter the maximum or minimum value, the value of VOT duration of non-aspirated sound is less than the corresponding value of VOT of aspirated sound.

b) According to the mean value of VOT duration of non-aspirated plosives, the sequence from length to segment is: k > t > p; The VOT duration of aspirated plosives is sorted according to the mean value, and the order from long to short is: k^h>t^h> p^h。 From the mean value of VOT duration of aspirated plosives and aspirated plosives, we find that the greater the mean value of VOT duration of the plosives with the back of tongue during pronunciation, that is, in Yushu dialect, the longer the VOT duration of the plosives with the back of the pronunciation position.

c) The difference of VOT duration between aspirated and non-aspirated plosivepers in Yushu dialect not only shows that the value of VOT duration is different, but also, we can see from the height of the standard deviation column in the figure that the value range of VOT duration is also very different. Generally, the value range of VOT duration of aspirated plosivepers is greater than that of non-aspirated plosivepers. And in each group of pronunciation methods, with the back of the pronunciation part, this range is also narrowing.

3.3. Duration Analysis

In order to avoid the influence of extreme values and obtain the experimental conclusions more clearly and intuitively, we take the average value after removing the maximum and minimum values of the duration of each type of plosives, and show the relationship between different pronunciation parts, pronunciation methods and duration by means of column diagram. Table 3 below:

Table 3: Absolute Statistical table of plosives duration.

Plosives category	Number of samples (units)	Average (duration: Ms)
p	16	14
pʰ	13	52
b	5	89
bilabial plosive	34	52
t	16	15
tʰ	14	58
d	21	117
Anterior stop of tongue tip	51	62
k	31	23
kʰ	16	67
g	14	130
root of the tongue plosive	61	73

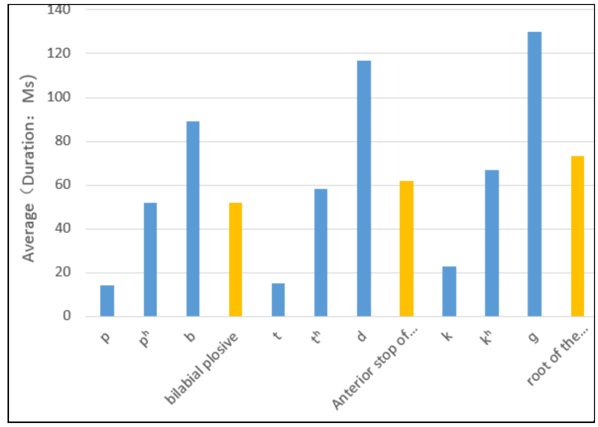

Figure 5: Plosives duration distribution.

In terms of pronunciation position, in the clear and non-aspirated plosives group, the length of the root of tongue plosives k is the longest, and the length of the double lip plosives p and the front of the tip of tongue plosives t are the same and shorter; In the clear aspirated plosives group, the root of tongue plosives k^h 's duration is the longest, followed by the anterior plosive t at the tip of tongue t^h， Double lip plosive p^h with the shortest duration; In the voiced plosives group, the duration of double lip plosives b is the shortest, the duration of anterior tongue tip plosives d is the second, and the duration of root tongue plosives g is the longest. From left to right, the orange bars represent the average duration of double lip plosive, anterior plosive and root plosive in turn. The figure shows double lip plosive < tip plosive < root plosive. To sum up, the farther back the pronunciation part is, the longer the plosives are pronounced, which reflects the influence of the flexibility of the starting organ and the contact area of the blocking part on the plosives. When the flexibility is poor and the contact area is large, it is more difficult to remove the obstruction, making the time of removing the obstruction longer.

From the perspective of pronunciation method, the duration of voiced plosives b is the longest and that of voiceless plosives p is the shortest; Among the sounds at the tip of tongue, voiced plosive has the longest duration, followed by clear aspirated plosives t^h， the shortest is t; Among the root sounds of tongue, the voiced plosive has the longest duration is g, the shortest is the voiceless plosive K. To sum up, the relationship between different pronunciation methods and duration of plosives is voiced plosives > clear aspirated plosives > clear non aspirated plosives. The main reason for the longest duration of voiced plosives is that there is a long band before blasting.

3.4. Intensity Analyses

Intensity is an important acoustic parameter of consonants. The intensity of plosives is usually extracted on the straight bar. Through the statistical analysis of the average value, maximum value and minimum value of different plosives in Yushu Tibetan language, and according to the statistical table data, the comparison diagram of plosives is drawn according to the pronunciation parts, as shown in Figure 6.

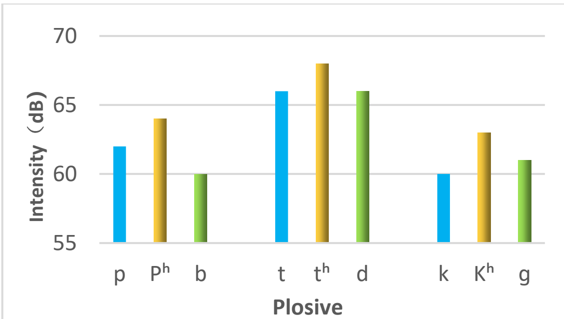

Figure6: Intensity comparison chart.

Figure 6 shows the intensity of the double lip plosive, the front plosive of the tip of tongue and the root plosive of tongue respectively, in which the column with the same colour represents the plosive with the same pronunciation method. The figure shows that among the three groups of plosives with different pronunciation parts, the intensity of clear aspirated plosives is always the strongest; Secondly, there is little difference in the intensity of clear and non-aspirated plosives and voiced plosives in plosives with the same pronunciation position. The clear and non-aspirated plosives in double lip plosives are stronger than voiced plosives, while the clear and non-aspirated plosives in tongue root plosives are weaker than voiced plosives, indicating that vocal cord vibration has no absolute influence on the intensity of plosives.

Overall, in Yushu dialect, the intensity of the plosive before the tip of tongue is the strongest and the intensity of the plosive at the root of tongue is the weakest. Firstly, the tip of tongue is more flexible than the root of tongue, and the instant time of resistance removal is shorter; Compared with the two lips, the resistance removal area is small and explosive.

3.5. Spectrum Analysis

As an important acoustic parameter of plosives, blasting pulse mainly involves the extraction of straight bar parameters [6]. We observe the spectral energy distribution patterns of different plosives by drawing the spectrum diagram of the central section of the straight bar.

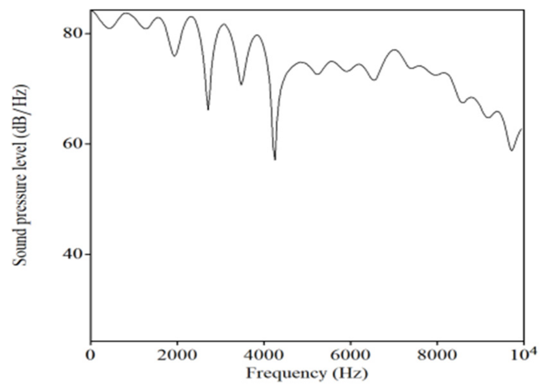

Figure 7: "p" straight bar spectrum.

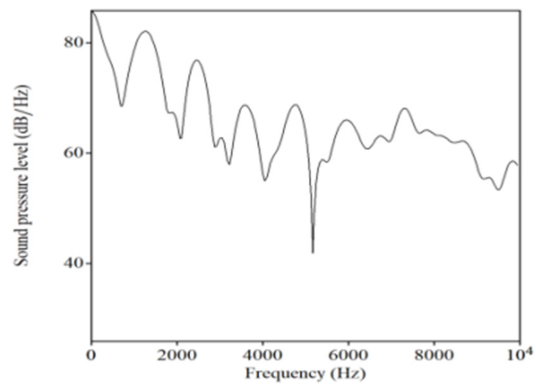

Figure 8: “p^h” straight bar spectrum.

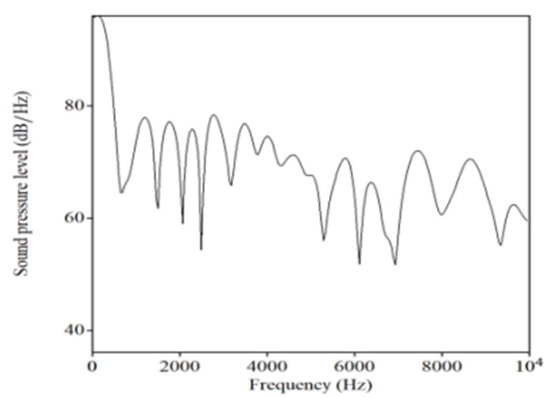

Figure 9: "b" straight bar spectrogram.

The above is the spectrum of straight bar of double lip plosive.

Fig. 7 is the spectrum diagram of double lip plosives p blasting. It is found that although the energy decreases with the increase of frequency, the energy of the whole frequency band is still very strong. Figure 8 shows p^h ’s the spectrum at the straight bar, the energy of the whole frequency band is still very strong, but the downward trend is more obvious than that of the plosive p. In Figure 9, compared with p and p^h during blasting, the overall spectral energy of b is lower, and the energy distribution from low frequency to high frequency is more uniform. With the increase of frequency, the overall energy decreases slightly. To sum up, the energy distribution of straight bar of double lip plosive is mainly between 60-

80dB, and the spectrum feature is that the higher the frequency, the smaller the energy.

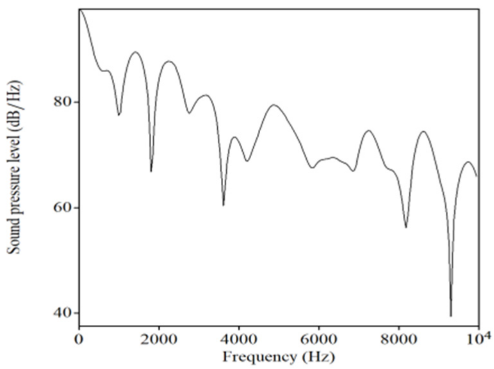

Figure 10: "t" straight bar spectrum.

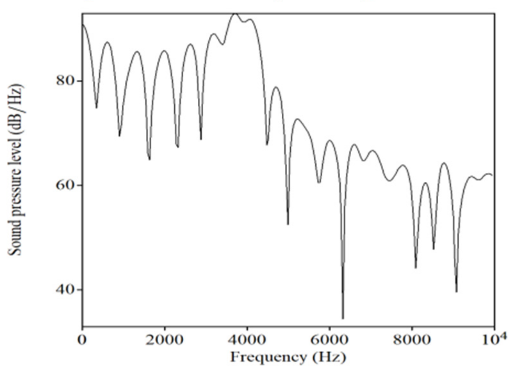

Figure 11: “t^h” straight bar spectrum.

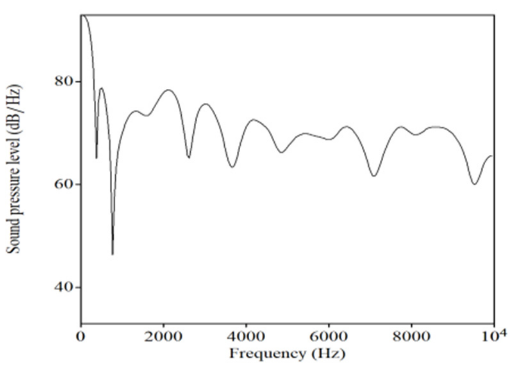

Figure 12: "d" straight bar spectrum.

The front sound of the tip of tongue is a consonant produced by the obstruction of the tip of tongue against the back of the upper incisor. The front plosive of the tip of tongue is a sound produced by suddenly removing the obstruction after the front occlusion of the tip of tongue.

Figure 10 is the spectrum diagram of t blasting. It can be seen from the figure that the energy of the whole frequency band is very strong during blasting, but the energy from low frequency to high frequency decreases more than that of double lip plosive. Figure 11 shows the spectrum diagram of t^h during blasting, the energy of the whole spectrum is very high, and the energy decreases greatly from low frequency to high frequency. Fig.12 is the spectrum diagram of d straight bar. Except that there is obvious energy weakening below 1000Hz, the overall energy is relatively uniform.

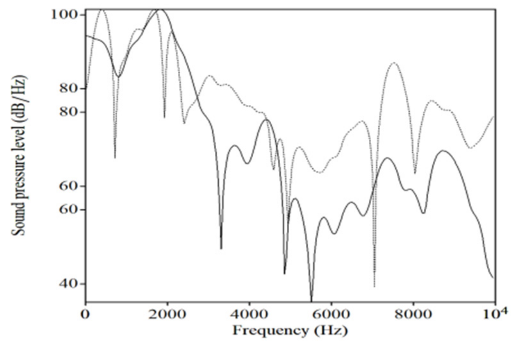

Figure 13: "k" straight bar spectrum.

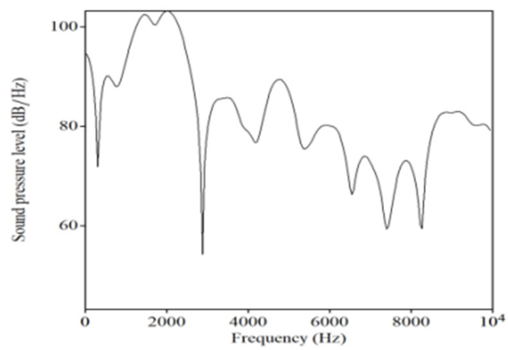

Figure 14: “k^h” straight bar spectrum.

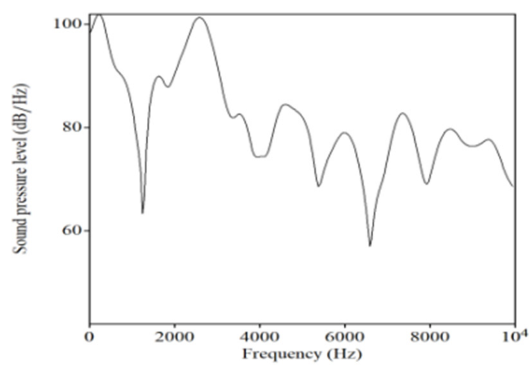

Figure 15: "g" straight bar spectrum.

Figure 13-15 are the spectrograms of straight bar of tongue root plosives. Compared with the first two groups of spectrum diagrams, the energy at low frequency is stronger, between 80-100dB, and the energy fluctuation range is larger.

The double straight bars phenomenon of Fig.13 is common in the root of the tongue. The solid line represents the front straight bar and the dotted line represents the rear straight bar. The energy of the whole frequency band is higher, and the high-frequency energy is weaker, even lower than 60dB. Figure 14 and Figure 15 are k^h and g respectively straight bar spectrum, the overall energy of the frequency band is strong, and the energy decreases from high frequency to low frequency.

4. GLOTTAL PLOSIVE ʔ

Also known as glottic plosive, it is a kind of consonant. It is a plosive formed by the instantaneous interruption of air flow caused by glottic closure. Because of its unique pronunciation position, glottic plosives only have clear consonants.

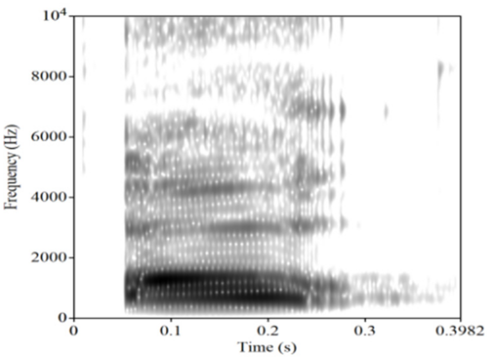

Figure 16: "breath" ʔo spectrogram.

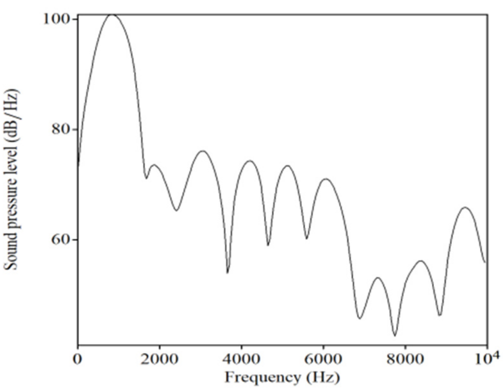

Figure 17: "ʔ" Straight bar spectrum.

It can be seen from the spectrogram of ʔo that the acoustic characteristics of laryngeal plosivepers are first shown as straight bars, which are directly followed by the formant of vowel o. therefore, VOT is equal to zero and the connecting segment is very small, which can be omitted. Figure 17 shows different from the above plosives blasting, the energy of laryngeal plosives blasting decreases sharply from high frequency to low frequency, and the energy at high frequency decreases significantly.

5. CONCLUSION

The mean value of VOT can clearly distinguish the aspirated and non-aspirated plosives of Yushu dialect; From the mean value of VOT duration of aspirated and non-aspirated plosives, the later the pronunciation part is, the longer its VOT duration is; The value range of VOT duration is also very different, generally speaking, the value range of VOT duration of aspirated plosives is greater than that of non-aspirated plosives, and in each group of pronunciation methods, this range is also narrowing with the back of pronunciation position; The energy from low frequency to high frequency is very strong during plosive blasting, and the energy at low frequency is also increasing with the back of the pronunciation part.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Anseraga. A survey of Tibetan Yushu dialect (Labu) phonology [J]. Tibet research, 2018 (1): 7.

[2] Chen Xiaoying, Hua Kan, Yu Hongzhi. Acoustic parameter analysis of vowel monosyllabic voice in Tibetan Lhasa [J]. Chinese Journal, 2010.

[3] Denzhenwengmu. Phonological study of Tibetan Yushu dialect [J]. Henan science and technology, 2015 (22): 1.

[4] Gesangjumian and GesangYangjing. Introduction to Tibetan dialect [M]. Ethnic Publishing House, 2002.

[5] Gesangjumian. Phonetic analysis of Tibetan Batang dialect [J]. National language, 1985 (2): 12.

[6] Huang Bufan, Suonan jiangcai, Zhang Minghui. Phonetic characteristics and historical evolution of Yushu Tibetan language [J]. Chinese Tibetology, 1994 (2): 24.

[7] Kong Jiangping. Basic course of experimental phonetics [M]. Peking University Press, 2015.

[8] Kong Jiangping. Tibetan dialect questionnaire [M]. Commercial Press, 2011.

[9] Santa, Dawa pengcuo. Study on phonetic characteristics of Yushu dialect in Kangba dialect [J]. Tibet research, 2010 (2): 89-93.

[10] Songbao Danzhou. On the Phonetic Evolution in Yushu dialect [D]. Qinghai Normal University, 2013.

[11] Suonan Lengzhi. On the contribution of experimental phonetics to the study of Tibetan phonetics [J]. Tibet research, 2016 (1): 7.

[12] Zhou Yixin. Acoustic analysis of syllables in Tibetan anduoxiahe dialect [D]. Northwest University for nationalities, 2011.edition.