A Voice Analysis on the Vowels of Tibetan Lhasa Based on Mdvp

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

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

[a] Gmx_Deity@163.Com, [b] Dwpc@163.Com

*Corresponding Author

Keywords: Tibetan Lhasa, Mdvp, Vowels, Voice

Abstract: in This Paper, Multi-Dimensional Voice Program (Mdvp) is Used to Extract the Voice Parameters of Tibetan Lhasa Vowels. the Main Parameters of Pitch Basic Parameters, Frequency Jitter Parameters and Amplitude Jitter Parameters Are Selected for Data Analysis and Mapping. the Study Found That the Fundamental Parameters of the Vowels of Male and Female Are Different, But the Frequency and Amplitude Jitter Parameters Have Some Similarities. the Study of the Tibetan Lhasa vowel's Pronunciation Has Great Practical and Theoretical Value for the Tibetan Voice Acoustic Modeling and Voice Lesions Treatment.

1. Introduction

The Vowel of Tibetan Lhasa is Not Available in the Traditional Grammar of Tibetan. Nowadays, Scholars Directly Transliterate the Past and Call It ཡུན་སྲུཝ. in Hua Kan and Suonan Cairang, the Compilation of the སྐད་བརྡ་རིག་པའི་སྤྱི་བཤད is Interpreted as ཡུན་སྲུཝ་ཞེས་བྱ་བ་ནི་ཧྲེང་སྲུཝ་ཡི་གཞུག་མའི་སྒྲ་རྒྱུ་ ཡིན།[1]. There Are 13 Core Vowels of Tibetan Lhasa Vowels, Which Are /a/, /Ao/, /Acb/, /Acu/, /Ae/, /I/, /Iu/, /U/, /Ua/, /Ue/, /e/, /o/, /Oe/[2]. Most of the Tibetan Language is Studied by Traditional Linguistic Methods, While the Experimental Phonetics of Tibetan Language is Rare, and the Study of Tibetan Voices is Lesser.

This Paper Intends to Use the Multi-Dimensional Voice Program (Mdvp) to Study the Vowels of Tibetan Lhasa. the Previous Researches in This Aspect Are: Kong Jiangping[3] Used Mdvp to Record and Extract 800 Chinese Speakers, and Established a Set of Chinese Voice Acoustic Parameter Library, Wang Shuwen[4] Conducted a Multi-Dimensional Phonetic Voice Analysis of Monosyllabic Mandarin Chinese, Shi Jing[5] Studied Multi-Dimensional Voice Program (Mdvp), Chen Xiaoying[6] Analyzed the Tibetan Lhasa's Monosyllabic Vocal Parameters, and Kong Jiangping[7] Studied the Tone Perception of Tibetan Lhasa.

The Study of the Vocal Pathology and Linguistics in Chinese National Language is Relatively Weak. Therefore, the Study of Tibetan Voice in Lhasa is Very Necessary and Valuable. in Linguistics, Multi-Dimensional Voice Analysis Can Be Used to Quantify Different Types of Phonation. in Speech Engineering, It Can Be Used to Quantify the Voices of Different People and Establish a Voice Source Model[8]. in Medicine, Multi-Dimensional Voice Analysis is Very Important for Voice Lesions.

2. Materials and Methods

2.1 Pronunciation Word List

There are 13 core vowels in the Tibetan Lhasa, which are /a/, /ao/, /acb/, /acu/, /ae/, /i/, /iu/, /u/, /ua/, /ue/, /e/, /o/, /oe/[9], which are developed into 53 vowels[10], as shown in Table 1:

  DOI: 10.25236/ICHCAD.2019.166

Table 1 the Vowels Of Tibetan Lhasa

Number	Vowel	IPA
1	/a/	[a] [a:] [am] [an] [aŋ/ã] [ap] [ak/aˀ] [aɹ]
2	/ao/	[ao]
3	/acb/	[əp]
4	/acu/	[əu]
5	/ae/	[ɛ:] [ɛn/ɛ]
6	/i/	[i] [i:] [im] [in/ĩ] [iŋ/ĩ] [ip] [ik/iˀ] [iɹ]
7	/iu/	[iu]
8	/u/	[u] [u:] [um] [un] [uŋ/ũ] [up] [uk/uˀ] [uɹ]
9	/ua/	[ua]
10	/ue/	[y:] [yn/ỹ]
11	/e/	[e] [e:] [em] [en/ẽ] [eŋ/ẽ] [ep] [ek/eˀ] [eɹ]
12	/o/	[o] [o:] [om] [on] [oŋ/õ] [op] [ok/oˀ] [oɹ]
13	/oe/	[ø:] [øn/õ]

2.2 Data Collection

The text recorder is 20 college students who are native speakers of Lhasa, 10 men and 10 women, they all have no speech disorder.

All the signal acquisition work in this experiment was completed in the professional voice recording studio of Northwest Minzu University. The recording software is Adobe Audition 3.0, with a sampling rate of 44100HZ and a resolution of 16 bits (according to the condition that MDVP can handle). The voice collection is carried out in a quiet room with an ambient noise of less than 45dB(A). The recorder wears a lavalier microphone and the recorder’s mouth is about 15 cm from the microphone. They all take natural comfort, and then smoothly speak all the vowels in Table 1.

The recorders should stretch the sound to make the vowel last for at least 3 seconds, and read each vowel three times. Finally, we stored all the sounds as voice files in the “*.wav” format.

2.3 Acoustic Analysis

We use the Multi-Dimensional Voice Program (MDVP)[11] of Kay Corporation of the United States to analyze the voice acoustics of the vowels in Tibetan Lhasa. The MDVP can extract 33 parameters, which can be divided into six categories: pitch parameters, frequency jitter parameters, amplitude jitter parameters, voice index, voice clearing parameters, and basic parameters.

This experiment selects several representative parameters of MDVP for analysis and comparison. By analyzing the Average Fundamental Frequency (F0), Highest Fundamental Frequency (Fhi), Lowest Fundamental Frequency (Flo), Standard Deviation of F0 (STD), Phonatory F0-Range in semi-tones (PFR), Absolute Jitter (Jita), Jitter Percent (Jitt), Pitch Perturbation Quotient (PPQ), Smoothed Pitch Perturbation Quotient (sPPQ), Shimmer in dB (ShdB), Shimmer Percent (Shim), Amplitude Perturbation Quotient (APQ), and Smoothed Ampl. Perturbation Quotient (sAPQ) between different vowels and genders in Tibetan Lhasa, we can explore the characteristics of Tibetan Lhasa vowels.

Brief description of the selected parameters: [12]

F0 is a basic concept in speech acoustics and is defined as the reciprocal (1/T0) of the vocal cord vibration period (T0). The F0 average fundamental frequency is the average of the fundamental frequency values of all extracted periods.

Fhi is the maximum value of the fundamental frequency values of all extracted periods.

Flo is the minimum value of the fundamental frequency values of all extracted periods.

Standard Deviation of F0 (STD) can be calculated as follows:

$$\mathrm{STD} = \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_0 - F_0^{(i)})^2} \tag{1}$$

In (1), i = 1, 2, 3, ... N is the extracted periodic parameter, N is equal to the number of extracted periods, and F0 is the fundamental frequency.

PFR is the range of the fundamental frequency between Fhi and Flo, and it expressed in chromatic scales. [13]

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:

$$\text{Jita} = \frac{1}{N-1}\sum_{i=1}^{N-1} | T_0^{(i)} - T_0^{(i+1)} | \qquad (2)$$

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:

$$\text{Jitt} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | T_0^{(i)} - T_0^{(i+1)} |}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \qquad (3)$$

PPQ is a relative evaluation of the fundamental period variation of the voice, which can well describe the short-range tonal perturbation of the voice. Calculated as follows:

$$\text{PPQ} = \frac{\frac{1}{N-4}\sum_{i=1}^{N-4} | \frac{1}{5}\sum_{r=0}^{4} T_0^{(i+r)} - T_0^{(i+2)} |}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \qquad (4)$$

sPPQ (%) is the relative assessment of short-term or long-term fundamental frequency variation. This measurement is widely used in the study of voice perturbations, which is very sensitive to fundamental frequency variations during continuous fundamental frequency periods. It is typical of hoarse voices. Calculated as follows:

$$\text{sPPQ} = \frac{\frac{1}{N-sf+1}\sum_{i=1}^{N-sf+1} | \frac{1}{sf}\sum_{r=0}^{sf-1} T_0^{(i+r)} - T_0^{(i+m)} |}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \qquad (5)$$

In (2), (3), (4), and (5), i = 1, 2, 3, ... N of T0(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:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1} | 20\log(A^{(i+1)} / A^{(i)}) | \qquad (6)$$

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:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | A^{(i)} - A^{(i+1)} |}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \qquad (7)$$

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:

$$APQ = \frac{\frac{1}{N-10}\sum_{i=1}^{N-10} | \frac{1}{11}\sum_{r=0}^{10} A^{(i+r)} - A^{(i+5)} |}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \qquad (8)$$

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:

$$sAPQ = \frac{\frac{1}{N-sf+1}\sum_{i=1}^{N-sf+1}\left|\frac{1}{sf}\sum_{r=0}^{sf-1}A^{(i+r)} - A^{(i+m)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \tag{9}$$

In (6), (7), (8), and (9), i=1, 2, 3...N of A(i) is the extracted peak amplitude parameter, and N is equal to the number of extracted periods.

3. Data Analysis

3.1 Fundamental Frequency Information

When studying the fundamental frequency information of Tibetan Lhasa vowels, we selected five parameters: Average Fundamental Frequency (F0), Highest Fundamental Frequency (Fhi), Lowest Fundamental Frequency (Flo), Standard Deviation of F0 (STD), and Phonatory F0-Range in semi-tones (PFR). We first take the average value of these parameters, and then make graphs.

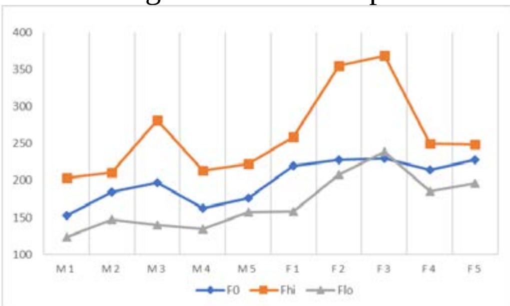

Fig.1 Curves of F0, Fhi and Flo.

Fig.2 Combination Diagram of Std and Pfr.

Figure 1 shows the parameter data of Fundamental Frequency (F0), Highest Fundamental Frequency (Fhi), and Lowest Fundamental Frequency (Flo) of the Tibetan Lhasa vowels. The ordinate is the value of the parameter data, and the abscissa is 10 recorders (M stands for male, F On behalf of women). The following figures are the same. From the figure we can see that F0, Fhi and Flo of the female voice are mostly higher than the male voice. In Highest Fundamental Frequency (Fhi), except for the Male 3's parameter data higher than the Female 1, Female 4 and Female 5, the value of other male voices is lower than the value of any female voices.

Figure 2 is a combination diagram of the standard deviation of Standard Deviation of F0 (STD), and Phonatory F0-Range in semi-tones (PFR) of vowels in Tibetan Lhasa. By the following figure, we can see that the variation of the male's STD is larger than the female. The STD of Male 1 and Male 3 increases with the rise of F0, while Male 2 and Male 3 and Male 4 decrease with the increase of F0. Only female 2 in female voice falls with F0 rise, the rest of the female voices are rising as F0 increases. The change of PFR is not as obvious as STD. Except for Male 2 and Female 4 and Female 5, the PFR of male and female voices mostly increase with F0 rise.

3.2 Frequency Jitter Parameters

When studying the frequency jitter parameters of Tibetan Lhasa vowels, we select four parameters: Absolute Jitter (Jita), Jitter Percent (Jitt), Pitch Perturbation Quotient (PPQ), and Smoothed Pitch Perturbation Quotient (sPPQ).

Fig.3 Combination Diagram of Jita and Jitt.

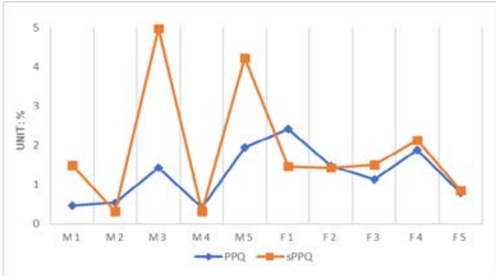

Fig.4 Curves of Ppq and Sppq.

Figure 3 is a combination diagram of Absolute Jitter (Jita) and Jitter Percent (Jitt) of the Tibetan Lhasa vowels. In the figure, we can see that the regular of Jita in the male and female vowels is not obvious. From the data of Jitt of male and female vowels in Tibetan Lhasa, the value of male voice fluctuated greatly, while the value of female voices fluctuated less and remained relatively stable.

Figure 4 is a curve of Pitch Perturbation Quotient (PPQ), and Smoothed Pitch Perturbation Quotient (sPPQ) in the Tibetan Lhasa vowels. From the figure, we can see that PPQ and sPPQ of the male and female vowels in Tibetan Lhasa are different. The trend is basically the same, but sPPQ is slightly more volatile than PPQ.

3.3 Amplitude Jitter Parameters

When studying the amplitude jitter parameters of Tibetan Lhasa vowels, we select four parameters: Shimmer in dB (ShdB), Shimmer Percent (Shim), Amplitude Perturbation Quotient (APQ), and Smoothed Ampl. Perturbation Quotient (sAPQ).

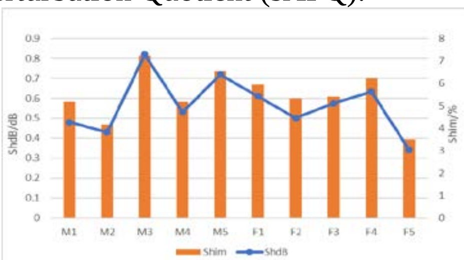

Fig.5 Combination Diagram of Shdb and Shim.

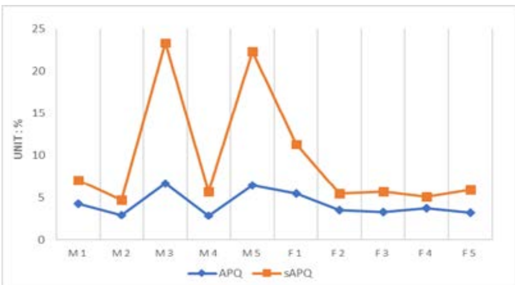

Fig.6 Curves of Apq and Sapq.

Figure 5 is a combination diagram of Shimmer in dB (ShdB) and Shimmer Percent (Shim) of the male and female vowels in Tibetan Lhasa. From the figure, we can see that the change of ShdB and

Shim are basically the same, and the difference between men and women is not obvious.

Figure 6 is the curve of Amplitude Perturbation Quotient (APQ) and Smoothed Ampl. Perturbation Quotient (sAPQ) of the male and female vowels in Tibetan Lhasa. As shown in the figure, APQ and sAPQ of the male and female vowels in Tibetan Lhasa are basically the same. However, there are only large fluctuations in Male 2 and Male 5. Overall, the amplitude jitter of the male voice is stronger than that of the female voice.

4. Summary

Through the above chart data analysis, we can roughly draw the following conclusions:

(1) In terms of fundamental frequency information, the F0, Fhi, Flo of female voice in the Tibetan Lhasa vowels is higher than the male voice, and PFR of the Tibetan Lhasa male and female voices increases with the rise of the fundamental frequency.

(2) In terms of frequency jitter parameters, the Jita of the male and female vowels in Tibetan Lhasa is not obvious. But Jitt of male voices fluctuates greatly. In addition, the PPQ and sPPQ are basically the same, but sPPQ is more fluctuating than PPQ.

(3) In terms of amplitude jitter parameters, the ShdB and Shim of the male and female vowels in Tibetan Lhasa are basically the same, and the difference between men and women is not obvious. In addition, the APQ and sAPQ have basically the same trend, but the amplitude jitter of the male voice is stronger than that of the female voice as a whole.

Acknowledgement

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

References

[1] Hua Kan, Hua Wenjing, Ai Guo et.al. (2002). Tibetan Anduo Dialect Vocabulary, Gansu Nationalities Publishing House.

[2] Hu Tan, Suonan Zhuoga, Luo Bingfen. (2001). Lhasa colloquial reader, National Publishing House.

[3] Kong Jiangping. (2009). The basic method of language vocal research, Commercial Press.

[4] Wang Shuwen. (2012). Phonetic analysis of multi-voice voices in monosyllabic Mandarin Chinese, Northwest Minzu University.

[5] Shi Jing, Zhou Ling, Yu Hongzhi. (2010). Research on Multi-dimensional Voice Analysis Software (MDVP). Journal of Changchun Institute of Technology(Natural Science Edition), Vol.11, No.02, pp.102-105.

[6] Chen Xiaoying, Chen Chen, Hu Yanru et.al. (2010). Research on the characteristics of vowel voices in Anduo Tibetan based on EGG signals. Journal of Northwest Minzu University (Natural Sciences), Vol. 31, No.01, pp.20- 23+78.

[7] Kong Jiangping. (1995). Tibetan (Lhasa) tone perception research. National Languages, No. 03, pp.56-64.

[8] ཟླ་བ་ཚེ་རིང་། སུམ་རྟགས་འགྲེལ་ཆུང་གོ་བདེའི་ལྡེ་མིག་བཞུགས་སོ། (2005).བོད་ལྗོངས་མི་རིགས་དཔེ་བསྐྲུན་ཁང་།

[9] Hu Tan. (2002). Theory of Tibetan Studies, China Tibetology Press.

[10] Gesang Jumian, Gesang Yangjing. (2002). Introduction to Tibetan Dialects, Ethnic Publishing House.

[11] Kay PENTAX. (2007). MDVP Instruction Manual.

[12] Mat Baki M, Wood G, Alston M, et al. (2015). Reliability of Opera VOX against

Multidimensional Voice Program (MDVP). Clinical Otolaryngology, Vol.40, No.1, pp.22-28.

[13] Hema N, Mahesh S, Pushpavathi M. (2009). Normative data for Multi-Dimensional Voice Program (MDVP) for adults-A computerized voice analysis system. Journal of All India Institute of Speech and Hearing, Vol.28, No.1, pp.1-7.