Study on the airflow signal model of voiceless plosive in Xiahe Tibetan based on Glottal MS110

Jing Zhang[1a], Yonghong Li[2b]*

[1]Lanzhou University of Finance and Economics
Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University
[2]Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University

[a]zjwawm226@126.com ,
[b]lyhweiwei@126.com

Abstract. This paper takes Xiahe dialect as the research object, writes a Matlab program to extract the airflow signal parameters of the voiceless plosive initial, and establishes the airflow model of the plosive on the basis of more than 200 monosyllabic corpus, and describes the characteristics of different plosives from the perspective of aerodynamics of speech. From the airflow signal parameters, the duration, airflow volume, airflow velocity and peak value of the aspirated plosive are significantly greater than that of the non-aspirated plosive. The airflow volume, airflow velocity and peak value of the velar are significantly smaller than those of the bilabial sound and the alveolar. From the air flow model, the non-aspirated plosive can be fitted as a linear model, and the aspirated plosive is two-stage and can be fitted as a polynomial model. The most significant difference between different pronunciation points of plosives is the slope. The slope of the alveolar is the largest, followed by the bilabial, and the velar is the smallest.

Keywords: Tibetan, Xiahe dialect, plosive, airflow signal, model

1 Introduction

With the development of modern science and technology, more and more experimental instruments have been applied to the study of speech, and the scope of phonetics research has expanded from the conventional acoustic analysis to the explanation of the physical mechanism of speech production. The breath is the ' driving force ' of speech production. Without breath, there is no speech. The study of pressure and airflow changes during speech generation is called speech aerodynamics [1]. Through the analysis of the airflow and air pressure signal, it can explain some special speech phenomena that cannot be explained by traditional research methods, reveal the physiological characteristics of different speech and the principle of pronunciation, and play a positive role in supplementing and promoting the research of speech acoustics.

G. Guan et al. (eds.), *Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023)*, Atlantis Highlights in Computer Sciences 16,
https://doi.org/10.2991/978-94-6463-264-4_81

The application of airflow and pressure instruments to study the physiological phenomena of speech have dated from the 1950s. Many linguists have carried out in-depth studies and discussions on airflow and air-pressure. Isshiki (1964) studied the airflow during the production of selected consonants [2]. Rothenberg (1977) developed Time-based and Amplitude-based Measures and did prediction and differential calculations for the glottal airflow, respectively [3]. Shadle (1995) studied an articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences [4]. Zraick and Smith-Olinde (2012) collected speech acoustic signals from one hundred fifty-seven speakers, analyzed the differences for participants, and verified the effects of different methods in data processing [5].

In the 1980s, Professor Fant introduced the method of studying speech by air pressure into China. Wu (1987) first studied the distinctive features between aspirated and non-aspirated consonants in Mandarin [6]. He argued that the air-pressure of non-aspirated consonants is higher than that of aspirated ones, and there is no significant difference in airflow volumes. In recent years, Chinese dialects and minority languages have been studied in terms of aerodynamics. Cheung (2004) researched the rhythm of Hongkong dialect, and analyzed the correlation between pitch and oral airflow, and finally set basic aspiration units [7]. Li (2022) argued that acoustics and airflow characteristic of the compound consonants initial about Xiahe of Tibetan dialects[8]. Zhang jing (2023) studied the acoustic characteristics of fricative and the role of air flow in the process of fricative pronunciation of Xiahe [9].

Domestic Tibetan is divided into three dialects: Wei, Kang and Amdo. Xiahe dialect is a typical representative of Amdo Tibetan, which is between agricultural dialect and pastoral dialect [10]. Xiahe dialect has a special character different from other dialect areas and has always been a hot spot in the study of Tibetan phonetics. This article selects the plosive of the Tibetan Xiahe dialect as the research object, and compiles a Matlab program to analyze the parameters extracted from its airflow signals. On this basis, an airflow model of the points of articulation and pronunciation method of the Xiahe dialect plosives is established. The establishment of the model can describe the pronunciation characteristics of the plosives from an aerodynamic perspective, laying the foundation for the study of other consonants, and also providing reference for the airflow analysis of other minority languages.

2 Experimental method

2.1 Experimental equipment

This experiment uses the Aeroview aerodynamic system(Glottal MS110) produced by Glottal Corporation, as shown in Fig.1. The Aeroview system consists of both hardware and software, which can accurately record oral airflow, nasal airflow, sub-glottal pressure, and microphone data during pronunciation, the aeroview software is shown in Fig.2. Using these recorded signals, Aeroview software can automatically calculate syllable rate, glottal resistance, average sub-glottal pressure, oral and nasal airflow (volume and rate), pitch, vital capacity, and so on.

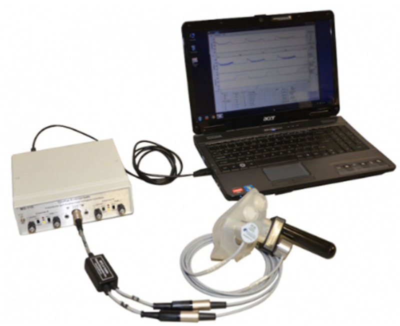

Fig. 1. Glottal MS110

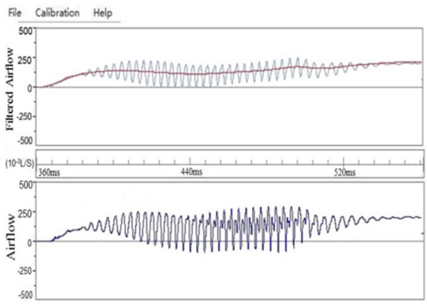

Fig. 2. Aeroview software

2.2 signal acquisition

Data collection in this experiment was carried out in a recording studio. For the convenience of comparative study, three kinds of signals including speech, voice and airflow were collected at the same time. The speech signal sample is stored in ' *.wav ' format, the sampling rate is 16 kHz, and the sampling accuracy is 16 bit. The calibration airflow volume of the airflow signal is 1.4 L and the flow rate is 0.5 L/ s. The experiment selected Xiahe County dialect, the representative of agricultural dialect, as the research object. The speaker is an undergraduate student of Northwest Minzu University, who has lived in Xiahe County since childhood, unaffected by other dialects, has no speech or voice disorders, and has no history of respiratory disease. In this study, we selected the spoken monosyllabic words containing 6 plosive initials in Xiahe Dialect, all of which were selected from the Survey of Tibetan Dialects [11].

2.3 main parameter

In the process of production, due to the different intensity of lung airflow output through the glottis and the change of organ obstruction, the changing airflow is produced, which constitutes different sound output effects. Therefore, the main parameters extracted and studied in this paper include:

- Airflow duration (AD; ms): The length of time in which air flows from the lungs through the glottis until exhaled.
- Airflow volume (EV; ml): the total amount of exhaled airflow during pronunciation.
- Airflow mean value (MEF; ml/s): Also known as average airflow velocity, represents the average exhalation velocity during articulation.
- Peak airflow (PEA; ml/s): also known as the peak airflow rate, refers to the maximum rate of airflow out of the glottis per unit time during pronunciation.

2.4 Airflow signal modeling

- Signal segmentation, using Audition software to segment the collected corpus, and extract the airflow signal of the stop sound. The starting position of the airflow signal is the end of the air pressure signal, and the ending position is the airflow signal tends to be gentle.
- Signal smoothing: Smoothing the airflow signal using a five-point smoothing algorithm.
- Normalization of signal duration. The duration of each consonant pronunciation is different. In order to facilitate comparison and statistics, it is necessary to normalize signal duration. The normalization method is to take 20 points equidistant from each airflow signal.
- Signal stacking: Using the Matlab platform to write a program to stack the segmented signals of the same type of plosives, we can observe the change mode of air velocity and range of air velocity.
- Signal model: Average the data from 20 points of the same type of consonants to obtain the average airflow curve of 20 data points, and then simulate it to establish the airflow model of the plosive in Xiahe dialect.

3 Model of plosive airflow

There are 9 plosive initials in Xiahe dialect. According to the different points of articulation, they are divided into three groups: bilabial [p] [ph] [b], alveolar [t] [th] [d] and velar [k] [kh] [g]. According to the different pronunciation methods, it is divided into non-aspirated voiceless plosive [p] [t] [k], aspirated voiceless plosive [ph] [th] [kh], non-aspirated voiced plosive sound [b] [d] [g]. The non-aspirated voiced plosives only appear with the pre-consonant, and are not used as initials alone. Therefore, this paper only analyzes the airflow of the six voiceless plosives in Xiahe dialect.

In this experiment, there are 211 speech samples of plosive monosyllables, including 94 non-aspirated speech samples, 117 aspirated speech samples, 16 samples for [p], 38 samples for [t], 40 samples for [k], 22 samples for [ph], 57 samples for [th], and 38 samples for [kh]. Fig.4 is the stack drawing of the non-aspirated airflow signal, Fig.5 is the stack drawing of the aspirated airflow signal, and Fig. 6 is the scatter diagram of plosive airflow signals (the horizontal axis is the duration and the vertical axis is the airflow).

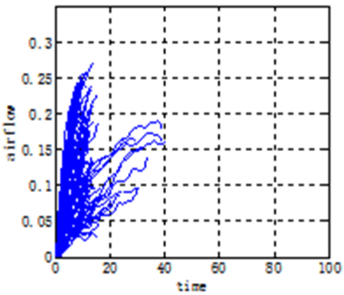

Fig. 3. Unaspirated plosive set

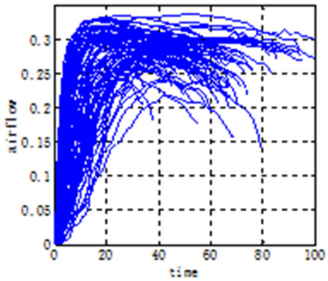

Fig. 4. Aspirated plosive set

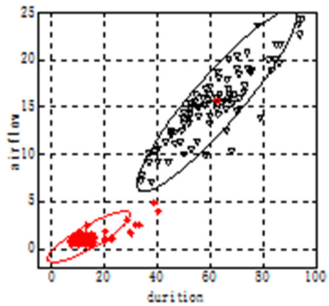

Fig. 5. Plosive distribution

From Fig.3 and Fig.4, the difference between the non-aspirated stop and the aspirated stop airflow signals is relatively large. After the explosion of the non-aspirated plosive, the glottis closes and the vocal cord begins to vibrate. It is reflected in the airflow signal that the consonant airflow is directly connected to the vowel airflow, and the airflow curve is basically shown as an inclined straight line.The airflow signal of the aspirated plosive is divided into two parts : the blasting section and the aspirated section. The airflow curve of the blasting section is a straight line inclined upward. The airflow curve of the aspirated section continues to rise next to the blasting section, and begins to decline slowly (or remains straight) after reaching the peak position. Due to the different parts of the pronunciation will also affect the airflow characteristics of the initial. The friction degree of the velar is large, which is relatively large in terms of duration and airflow.

As can be seen from Figure 5, the non-aspirated plosive is relatively stable in terms of airflow duration and airflow rate, the distribution of the aspirated plosive is relatively scattered, and the span of duration and airflow are very large. The average airflow rate of aspirated plosive is 15.7ml, and the average airflow duration is 62.5ms. The average airflow rate of non-aspirated plosive is 1.2ml, and the average airflow duration is 14ms. The peak airflow of aspirated plosive is about 300ml/s, and the peak airflow of non-aspirated plosive is 100-200ml/s.

In the following, each plosive airflow signal is analyzed and simulated, and the plosive airflow curve model is established. The curve model of airflow signal is shown in Fig.6 (the horizontal axis is the actual average duration (ms) of each plosive, the vertical axis is the average airflow (l/s), the solid line represents the original data, and the dotted line represents the simulated data).

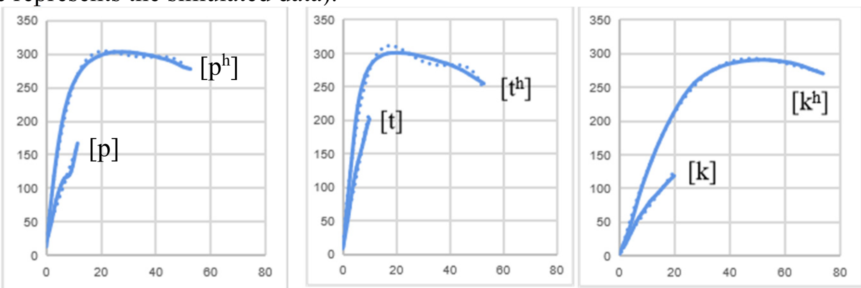

Fig. 6. Plosive airflow model of Xiahe dialect

The difference of the plosive model between different articulation methods: the non-aspirated plosive can be simulated with a linear model; the airflow curve of the aspirated plosive is in two stages, with the first half rising rapidly to the apex, accounting for about 1/3 of the total length of time, and the second half dropping slightly. It can be fitted by linear model in sections or polynomial fitting as a whole. In this paper, polynomial fitting is selected.

The difference of the plosive model between different pronunciation points: The main difference between the non-aspirated sound is the slope. The slope of the airflow curve of the velar is the smallest, followed by the bilabial, and the alveolar is the largest, that is, the airflow rate of [t] is very fast; in aspirated sound, the airflow velocity of bilabial and alveolar increases rapidly with obvious inflection point, while the airflow velocity of velar is slower and closer to parabola.

4 Conclusion

This paper takes the Tibetan Xiahe dialect as the research object, extracts the airflow parameters of the plosives in Xiahe dialect and establishes the model, discusses the characteristics of different plosives from the perspective of speech aerodynamics. From the airflow signal parameters, the duration, airflow volume, airflow velocity and peak value of the aspirated plosive are significantly greater than that of the non-aspirated plosive. The airflow volume, airflow velocity and peak value of the velar are significantly smaller than those of the bilabial sound and the alveolar. From the perspective of the airflow model, the non-aspirated plosive is linearly distributed and fitted as a linear model, while the aspirated plosive is a two-stage model, rising first and then falling, and fitted as a polynomial model. The study of this paper provides reference for the phonetic acoustic analysis of Tibetan dialects, and also provides data for phonetic teaching, phonetic engineering and phonetic technology application.

Acknowledgment

This work was financially supported by the Fundamental Research Funds for the Central Universities (No.31920220019) and NSFC grant fund (No.11964034).

References

1. B. Huaiqiao,L.Maocan.(2014) Summary of Experimental Phonetics. Peking University Press. Beijing.
2. Isshiki,N. (1964)Regulatory mechanisms of vocal intensity variation. Journal of Speech and Hearing Research. 7:17-29.
3. Rothenberg. (1977) Measurement of Airflow in Speech. Journal of Speech and Hearing Research. 20: 155-176.
4. Shadle, C. (1995) The aerodynamics of speech. In Hardcastle ,W. J. and Laver,J. (eds.) The Handbook of Phonetic Sciences. Oxford: Blackwell
5. Zraick, R.I., Smith-Olinde, L.S., & Shotts, L. (2012) Adult Normative Data for the KayPentax Phonatory Aerodynamic System Model 6600. Journal of Voice.164-176.
6. W. Zongji,L.Maocan.(2004) The Linguistic proceedings of Wu Zongji. The Commercial Press. Beijing.
7. Cheung Yuk-Man. (2004) An Aerodynamic Analysis of Intonation in Hong Kong Cantonese. Speech Prosody
8. Li yonghong, Zhang jing Wangjing.(2022) An acoustics and airflow analysis of the compound consonants initial about Xiahe of Tibetan dialects, Experimental linguistics, 11:49-53.
9. Z. Jing, L. Yonghong.(2023) Study on fricative initials of Tibetan Xiahe Dialect based on speech and airflow, In: 2nd International Conference on Educational Innovation and Multimedia Technology. Nanchang. pp.276-282. https://www.atlantis-press.com/proceed.
10. W. Shuangcheng.(2012) The phonetic research of Tibetan Anduo dialect. Zhongxi Book Company, Shanghai.
11. K. Jiangping. (2015)Basic Course of Experimental Phonetics. Peking University Press. Beijing.