ACOUSTIC AND BREATHING STUDY OF NATURE DOCUMENTARY NARRATION DUBBING

Jingyi Sun
（M.Lin., PKU | dual-degree-M.Chi., NUS）

A THESIS SUBMITTED

**FOR THE DEGREE OF MASTER OF
LINGUISTICS & APPLIED LINGUISTICS
DEPARTMENT OF CHINESE LANGUAGE AND LITERATURE
PEKING UNIVERSITY**

**FOR THE DUAL-DEGREE OF MASTER OF CHINESE STUDIES
FACULTY OF ARTS & SOCIAL SCIENCES
NATIONAL UNIVERSITY OF SINGAPORE**

2021

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

孙静怡

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor at Peking University, Associate Prof. Lin You-Jing, and the director of the Linguistics Lab, Prof. Jiangping Kong, for every academic conversation where I was convinced by their extensive accumulation of knowledge. I also thank my supervisor, Associate Prof. LEE Cher Leng in NUS, as well as Associate Prof. Yuzhi Shi and THAM Shiao Wei for their instructions during my study in Singapore. Without the first freshman party, it would be difficult for me to quickly familiarize myself with the local customs.

I would like to thank my group members in the lab, such as Wen Liu, Changwei Liang and Jie Yang , for their care in my major studies, especially in helping me to contact the subjects when I was stuck in the investigation.

Thanks to my families. My parents have been doing their best to provide me with the comfortable living and learning conditions all the time, so that I can understand the importance of knowledge and build up a deep fascination for wisdom.

Finally, thanks to myself, I hope to continue on the path of learning, and I hope this is far from the end of my life reflection.

CONTENTS

ABSTRACT

Speech makes social interactions possible, while breathing plays an important role in maintaining body functions. This paper examines the acoustic characteristics and respiratory rhythmic patterns of speech articulation through a comparison three speech styles: documentary narration, narrator reading aloud, and non-narrator reading aloud.

The thesis is consisted of four chapters. Chapter 1 introduces the characteristics of nature documentaries and focuses on the history of domestic and international research on the acoustics of speech art and oral cultures, as well as phonological studies using respiratory signals.

Chapter 2 lays out the experimental design, data analysis and parameter definitions.

Chapter 3 compares in detail the acoustic and thoraco-abdominal breathing characteristics of three different speech styles: nature documentary narration, narrator reading, and ordinary people non-narrator reading. Acoustic evidence shows that the duration and interval of rhyming phrases are flexible and variable in nature-documentary narration, and the rhythm is mainly achieved by lengthening or accentuation of syllables within the rhyming unit, rather than by just simply changing the duration of intervals; toward the end of the rhyming phrase, the phonation tends to switch from modal to creaky. When the narrators and non-narrators are only reading texts aloud, on the other hand, the number of rhyming phrases is greatly reduced, sometimes even to zero; with invariant rhythm and modal voice all throughout. The respiratory signals show the following facts: (1) the duration and amplitude of respiratory resets were significantly correlated in both documentary

narration and reading aloud; however, only in documentary narration the two parameters also maintain a high positive correlation with the slope of the inspiratory section; (2) the average amplitude, average duration and number of resets of abdominal respiration were higher than those of thoracic respiration in narration and reading, while the average duration and number of resets were approximate to those of thoracic respiration in non-narrator reading; the numbers of thoracic and abdominal respirations were not very different from those in narration; (3) abdominal respiration was more regular and stable than that of thoracic respiration in all three speech styles, and respiratory resets were always present after pauses, but the breathing signals are most stable in non-narrator reading, and pauses always co-occurs with breath resetting. (4) tones affect thoracic respiratory resetting in narration, not in reading aloud; (5) thoracic and abdominal respiratory signals in narration generally change toward the same direction, though the starting point of their changes are not the same; in narration and non-narration reading, on the other hand, the directions of changes are basically opposite to each other; (6) breathing patterns of different speech styles in narration have different respiratory resets, relative distinct resetting points, and respiratory groups. (6) the breathing patterns of different speech styles do not differ significantly in terms of the respiratory reset, the relative position of the reset points, and the way the breaths are grouped. In summary, narration and reading aloud show more obvious differences in six dimensions: the type of voice change at the end of rhyming sentences; the number of resets of thoracic and abdominal breathing; the correlation between the duration, amplitude of breathing resets, and the slope of the inspiratory segment; the relationship between breathing resets and pauses, the overall stability of the thoracic breathing signal curve; and the characteristics of the thoracic and abdominal breathing signal trend.

Chapter 4 summarizes the main findings of this paper, and discusses the shortcomings of this study and the issues that still need to be addressed in the future.

Based on these studies, we should be able to understand more about the phonological and respiratory rhythms of narrations and reading aloud, which further enriches the physiological study of speech production.

Chapter 1 Introduction

Structural linguistics understands speech as the concrete output of words or phrases in actual contexts, rather than as an abstract, stable language system that can be transmitted from one generation to another①. In the actual speech process, in addition to the specific semantic and grammatical messages that can be conveyed, phonology often gives more subtle and precise meanings to speech in the form of specific accent assignments, rhythms, and phonation types. While semantics and syntax often vary from language to language in the transmission of such information, the matching of phonological expressions with specific paralinguistic levels of meaning can exist across cultures and languages, a phenomenon that is particularly evident at the level of speech art②. For example, when dubbing a documentary with a sense of history at home and abroad, we generally choose a thick and low voice, with a slow but powerful speech speed, so that even people who do not know the language will clearly feel that the slow but unhurried flow of speech seems to be broadcasting; when we call a foreigner in a foreign country who doesn't understand Chinese at all, and anxiously ask for advice on how to go to the doctor, they may not be able to figure out the content of our meaning, but they can basically tell that we are in an emergency situation by the speed, rhythm, delay and intonation of our speech. These are the cross-linguistic stylistic and informational aspects of speech when it comes to specific speech forms. The study of the acoustics of speech forms with specific articulation patterns can help us to grasp the

---

① 费尔迪南·德·索绪尔, 高名凯. 普通语言学教程[M]. 商务印书馆, 1980: 28-40.

② 这一部分以及随后举例的那部分，也是改写并举例扩充了孔老师《实验语音学基础教程》第 12.10 章节的内容，和脚注③引用的是同一部分，因此没有另外再做脚注说明。

characteristics of these specific articulation types in a concrete way on the one hand, and "provide an acoustic basis for modeling the speech synthesis of virtual announcers" (Kong, 2015)[①].

As a kind of documentary art, there are many different kinds of documentaries and different classification standards, but it is generally recognized at home and abroad that nature documentaries are unique and distinctive in style and subject matter (Zhu, 2002)[②]. After a series of documentary production processes, such as topic selection, planning, interviewing, recording, shooting, and editing, the key step of final synthesis still requires the clever integration of narration and even background music to become a complete work with scientific and aesthetic values, which is the result of a combination of factors such as the multiplicity of picture contents, the collocation of picture articulation, the abstraction of time and space transformation, and the complexity of cause-and-effect relationships (Wang; Yan, 2014)[③]. The narration characteristics of different types of documentaries also differ to a greater or lesser extent. At the same time, the dubbing of the narration needs to closely match the content of the video and not to overwhelm the main character, so it shows different characteristics from recitation and news broadcast. This study takes only nature documentaries as the object, and examines what characteristics on this particular genre of documentary has in terms of acoustics and breath resetting, and compares the formal dubbing by a voiceover artist with the voiceover’s reading, and the reading by an untrained general articulator from the above perspectives, to progressively analyze the

[①] 孔江平. 实验语音学基础教程[M]. 北京大学出版社, 2015: 189-190.
[②] 朱景和. 纪录片创作[M]. 中国人民大学出版社, 2002: 4-8.
[③] 王明军, 阎亮. 影视配音实用教程[M]. 中国传媒大学出版社, 2014: 1-2.

multidimensional acoustic characteristics of this particular dubbing.

This chapter will first give a general introduction to the research object, then sort out the research objects and specific research methods selected by previous researchers at the macro level of speech phonation using acoustic analysis tools and perspectives, then categorize and elaborate on the research results of breath combined with acoustic analysis, and finally state the specific research content and significance of this topic.

1.1Introduction to Nature Documentaries

The definition of documentary film still varies from one school of thought to another, with social context, political objectives, cultural perceptions, artistic interpretations, and other factors leading to bias or emphasis in defining the concept. However, in general, truthfulness and narrative truthfulness are the common points① (but the fact that "truthfulness" is often compromised in the actual filming process due to the subjective position of the creator or the team in the creation of the script and the selection of the footage, while it does not prevent us from considering it as the most important pursuit in the creation of documentary multimedia products). From the viewpoint of filming purpose, documentaries can be classified as revealing or preserving, persuading or advocating, analyzing or questioning, and pure expression②; from the viewpoint of social benefits generated by documentaries, they can play different social roles, such as prophets of life's future, reporters, explorers of the earth's nature, painters, expositors of social problems, observers of human history, and dispassionate and objective chroniclers. The documentary can be a prophet of the future of

---

① 董春晓. 纪录片制作教程[M]. 浙江大学出版社, 2014: 2-4.
② Renov M . Theorizing Documentary[M]. 1993.

life, a reporter, an explorer of the earth and nature, a painter, an exposer of social problems, an observer of human history, a dispassionate and objective chronicler, etc[①]. From the point of view of documentary material, nature documentaries are one of the branches of human geography documentaries, and there are also human history and folklore documentaries in parallel[②]. Since the 1990s, nature documentaries have become a representative of documentary audiovisual products with high investment, high revenue and high reputation, with ocean, mountain, water, wildlife and environmental protection as their enduring themes[③].

Focusing solely on the dubbing of nature documentary narration, we find that the current literature on this type of documentary is mainly based on the "ear" and the "perception" of one's own dubbing experience to summarize its characteristics, and in China, students in broadcasting departments in colleges and universities often use abstract textual expressions in the classroom and teachers' guidance in practical courses to figure out the dubbing style that fits the content of the picture. Since different students' own vocal conditions can greatly restrict the selection of voiceover content, they have to learn voiceover methods and find their own voice types to better serve the presentation and meaning of the images. The only systematic discussion from the perspective of pedagogical practice rather than theoretical knowledge construction is found in Kristin's voice practice training textbook, which does not focus on training voice actors, but rather on mobilizing the various organs of the body involved in voice and learning to feel their presence, thus laying the foundation for the

① 巴尔诺. 世界纪录电影史[M]. 中国电影出版社, 1992.
② 李潇.人文地理类纪录片解说配音中的问题与策略研究[D].河南大学,2017: 8-20.
③ 朱景和. 纪录片创作[M]. 中国人民大学出版社, 2002: 7.

ultimate use of voice[①]. Although this book has almost become a required textbook for the media industry, and it does describe what "releasing the natural voice" means and how to improve vocal skills to achieve true freedom of expression in terms of body movements, facial expressions, mental states, brain imagination, oral movements, and breathing patterns[②]. However, there is still no concrete analysis and explanation of whether and what the differences are between voice acting and other linguistic contexts such as reading and broadcasting. Therefore, these omissions also provide room for further exploration and research in this paper.

Compared to other types of documentaries, nature documentaries are characterized by the following features based on the observation of auditory information:

1. thoracic resonance and complemented by oral resonance, flexible mobilization of the air sound in the lower register and the real sound in the middle register according to the content, the overall soothing and low, so as to achieve a better fit with the picture[③];

2. even and full amplitude, stable breath[④].

3. the overall length of voice is coherent and smooth, while the content of the picture will show a certain level of change[⑤], but still to ensure that the speed of speech is moderate, to find a balance between clear and too slow and dragging, so as to give the audience sufficient reaction time to guide them to think and organize information[⑥].

4. The rhythm will be based on the content of the picture to deal with the light and heavy

---

① [英] 克里斯廷·林克莱特 著/[荷] 安德烈·斯洛布 绘/彭莉佳 译，表演练声课[M]. 后浪|文化发展出版社, 2018.
② [英] 克里斯廷·林克莱特 著/[荷] 安德烈·斯洛布 绘/彭莉佳 译，表演练声课[M]. 后浪|文化发展出版社, 2018：15.
③ 张守鑫. 赵忠祥电视播音及主持艺术研究[D]. 四川师范大学，2018：23-25.
④ 张守鑫. 赵忠祥电视播音及主持艺术研究[D]. 四川师范大学，2018：26-28.
⑤ 高明敏.配音艺术——配音语言对角色的表现和塑造方式探析[D].上海师范大学,2020: 15-46.
⑥ 魏惠敏.新闻播音员做好各类配音工作需要掌握的技巧[J].科技传播,2019,11(23):55-56.

sounds, never flatly popping words or flatly pushing the flow of speech, but wave-shaped flow of speech, combined with the semantic and emotional needs, up and down with the trend, clear and natural when the speed of speech is fast, and when the speed of speech is slow, and can deal with the trend of speech and the end of speech, and finally use the exhale airflow to drag the voice[①].

Based on the above auditory observations, we will explore whether the relationships between fundamental frequency, phonation, breath resetting and rhythm show different characteristics in dubbing and reading aloud; to what extent the switching or lengthening of picture content during dubbing affects the regulation of thoracic and abdominal breathing during dubbing and the cooperation between the two; whether the type of semantic description of the narration affects the rhythm of the dubbers; whether ordinary pronouncers without broadcasting training will adjust their reading consciously by paying attention to the different semantic content of the narration; the characteristics of the duration of the syllables before and after the breathing boundary, the slope of the exhalation and inhalation segments at the breathing reset point that differ from the reading, and how it relates to the base frequency of the syllables before and after each breathing reset boundary? Clear answers to these questions will not only help us to understand the unique phonation, articulation and breathing characteristics of natural documentary voice-over, but also provide concrete and objective reference and guidance for the study of broadcasting courses, thus further contributing to the professional development of the field of dubbing.

[①] 吴峰.电视专题纪录片配音技巧的把控[J].记者摇篮,2020(04):112-113.

1.2 Current Status of Phonological Research on Speech Art

In terms of macro aspects related to the art of dubbing, the current domestic case studies using phonological methods to analyze speech art or oral culture fall into two main categories.

The first category is the experimental sampling investigation using purely acoustic analysis, and these studies include: Qiang Huang and Ping Huang (2001) investigated and preliminarily analyzed the acoustic characteristics of the singing pronunciation of a special performance form of the northern operatic art of "Han Dang Da Gu". "Compared with normal pronunciation, the resonance peaks of the "candleholder-bearing people" when pronouncing vowels with bamboo chopsticks in their mouths may change due to changes in the opening degree and lip area, but the overall pattern of formant does not change, and clear vowels can still be produced[①]. Then he specifically analyzed the differences between the two by comparing the formant, LTF, LTAS, F0, open quotient, speed quotient and other multidimensional acoustic parameters of unitary tones, phrases and repertoire of the same articulator in candleholder-bearing and normal conditions; Jing Wu (2014) compared folksinging and bel canto through acoustic analysis, and investigated that the convergence of the formant in width, height, and frequency domain of the two singing styles led to the similarity in terms of timbre; Qichao Han (2014) used acoustic analysis to sample and summarize in detail the tone, duration, pitch, and the relationship between duration and rhythm, pitch and melody of Kunqu[②]; Xie Genggaocaidan (2015) used Praat to extract and

---

① 黄强，黄平. "含灯大鼓"唱声声学特征的初步分析[C]// 新世纪的现代语音学——第五届全国现代语音学学术会议论文集. 2001.

② 韩启超.南昆念白声学实验分析(一)[J].中国音乐学,2014(04):102-114.

analyze the F0, power, and other acoustic parameters of the Tibetan folk song Maqu playing in Zhanianqin[①]; Yu Chen (2016) studied the F0, range, and harmonic power of the special artistic falsetto form of "Tiao Housang" of Henan Huaiqing Bangzi from the perspective of acoustical analysis[②]; Yang Cao (2016) conducted an experimental phonetic study on the speech signals during singing, recitation and reading of lyric songs in two lines of Liu Zi opera, Laosheng and Qing Yi, and focused on the differences in three aspects: pitch, LTF, and LTAS; Fang Wang (2016) also used acoustic analysis to compare and analyze the rhythmic characteristics, duration, F0, and power of five Tianzhu Tibetan folk songs; Ruiqi Ma (2019) from the perspective of comparing the rhythmic performance of Shandong Clapper and the common rhythmic reading, discussed in detail the differences exhibited in two dimensions: the F0 distribution and the LTAS.

Another type of multidimensional speech studies is voice and/or combined with speech and breathing signals, mainly: Chunyan Qu and Yongxiang Liu (2000) investigated the voice characteristics of different roles in Beijing opera with the help of computerized voice spectrum analysis system and SPSS, including F0, jitter, shimmer, NNE and vocal range span[③]; Shanying Yu (2010) compared the vowel formant of bel canto with those of Chinese folk singing, Kunqu, and Yueju through spectral analysis and CT experiments, and pointed out that singers' voices of Chinese folk and opera singing can also obtain singers' formant, but the high frequencies are rich and the low frequencies are insufficient, and the laryngeal state changes are also different from bel canto[④]; Ji-deok Yin (2010) analyzed the speech

---

① 谢更高才旦．玛曲弹唱口传文化的声学特征分析[D].西北民族大学, 2015:1-36.
② 陈彧.怀梆“挑后嗓”的音乐声学分析[J].天津音乐学院学报,2016(02):23-29.
③ 曲春燕,刘永祥.京剧嗓音的声学分析及音域特征[J].中华耳鼻咽喉科杂志,2000(01):58-60.
④ 于善英.不同歌唱类型歌手共振峰特征及音色形成的机理研究[J].音乐研究,2010(02):74-80+58.

rhyme of poetry readings in the linguistics laboratory of Peking University from the perspective of the phonation and articulation, and used electroglottography to analyze the basic voice characteristics, phonation, breathing rhythms and voice characteristics of Chinese utterance stress in poetry reading[①]; Yifan Qian (2009) used voice characteristics as the basic observation object to compare and analyze the different phonation methods of ordinary speech vocalization and those of pronouncers who had received bel canto and folksinging training, and analyzed them from the perspectives of power, F0, open quotient, speed quotient, etc[②]. Feng Yang (2012) compared and analyzed the rhyme, melody, phonation, and respiratory characteristics of different traditional Chinese chants using multimodal speech studies of video, speech, voice, and thoracoabdominal breathing signals[③]; Ikuyo Yoshinaga (2012) studied the rhythmic, melodic, vocal, and respiratory characteristics of Japanese Noh drama, Chinese Buddhist chanting, and Tibetan chanting by extracting and analyzing the parameters of speech signals and electroglottography signals[④]. Hu, A. Xu, and Gegenthana (2013) also studied the phonation parameters of Mongolian Long-tune by using the EEG signal and the difference between male and female[⑤]. Li Dong and Jiangping Kong(2017) studied the fluctuating characteristics of the vibrato and the changing pattern of phonation of the Kunqu Guimendan by analyzing the EEG signal[⑥]. It can be seen that whether it is purely acoustic or combined with voice and respiratory analysis, domestic research on speech art or oral culture has become more in-depth and detailed, trying to

---

[①] 尹基德.汉语韵律的嗓音发声研究[D]. 北京大学, 2010.
[②] 钱一凡. 言语发生与艺术发声嗓音特性比较研究[D].北京大学, 2009.
[③] 杨锋.中国传统吟诵研究[D]. 北京大学, 2012.
[④] 吉永郁代. 能剧嗓音发声研究[D]. 北京大学, 2012.
[⑤] 胡阿旭,格根塔娜.蒙古长调的嗓音特征声学分析[J].科技信息,2013(36):8.
[⑥] 董理,孔江平.昆曲闺门旦颤音的嗓音特征[J].清华大学学报(自然科学版),2017,57(06):625-630.

establish a more comprehensive understanding of the research object from multiple aspects, and also objectively laying the foundation for voice sample modeling and synthesis.

From the microscopic aspect closely related to the art of dubbing, Yudong Chen (2004) analyzed and summarized the basic types and variations of the focus of small sentences, and then proved the reality and regulation of the accent level in the segmental structure, and applied the laws and means of segmental regulation to the analysis of the rhythm of media speech, providing a two-way reference for the teaching and communication practice of speech language in theory and application[①]. Xiaohua Li (2007) explored the causes of the rhythmic characteristics of news broadcasts from the perspective of duration patterns and pitch structure within and across syllables[②]; Jingjing Tan (2008) focused on the analysis of the breathing signals of common speakers when they read aloud, and investigated the breathing resets, the distribution of high and low points, the breathing boundaries and the rhythm of news in different types of speech materials ("modern style" poetry, words, novels, prose, etc.)[③]. The rhythmic characteristics of breathing in Mandarin Chinese are explained in detail by taking various parameters such as the correlation between the boundary and syllable duration. It can be seen that in the process of empirical phonological research on a particular artistic form of pronunciation, we tend to compare the target pronunciation with the common speech pronunciation, which is a very obvious trend in artistic vocalization research related to voiceover in a narrow sense. In our project, we likewise chose three types of speech samples for collection: formal dubbing by professional dubbers, normal reading by

---

① 陈玉东.传媒有声语言语段的构造和调节 [D]. 北京大学, 2004.
② 李晓华.新闻播音节律特征研究[D].北京大学, 2007.
③ 谭晶晶.汉语普通话朗读时的呼吸节奏研究[D].北京大学, 2008.

professional dubbers, and reading by ordinary pronouncers without any dubbing training, and then distinguished the acoustic and respiratory characteristics of formal voicing in a graded manner; at the same time, we compared them with previous related studies to check and identify the differences.

Linguistic research using respiratory signals has seen a spurt of quite a wealth of topics explored and found around 2010 at home and abroad. In addition to the above master's thesis by Jingjing Tan (2008), which used reading from a different corpus as the object of study, many universities and research institutes have used a variety of respiratory signal acquisition tools (mainly the MLT1132 Piezo Respiratory Belt Transducer produced by the Australian AD Instrument[①]) to conduct multidimensional analysis of breathing, speech, voice, and rhythm. In China, the *Phonetics Lab* of Peking University, the *Key Laboratory of Information Technology of Chinese Ethnic Languages and Writings* of Northwest University for Nationalities, Nankai University, and Jinan University are the main representatives, among which Peking University is the pioneer in using MLT1132 Piezo Respiratory Belt to study speech rhythm. In abroad, the research on this part is mainly published in the *Journal of Speech, Language,* and *Hearing Research*, and the tools used for breath signal collection and analysis are also different. The analysis of respiratory signals is an important bridge between the analysis of rhythm and voice parameters, so the following section will explain in more detail the study of respiratory signals.

Respiration plays an essential role in maintaining the basic vital signs of the body and in speech-mediated communication, providing the power source for speech production. The

① 吴君如. 呼吸带说明.

complete respiratory system consists of: nose, pharynx, larynx, trachea, bronchi and lungs. Externally the lungs are in contact with the ribs, diaphragm and mediastinum and are greatly influenced by the motion of these parts; internally its parietal, posterior and middle lobes are more mobile, while the lower, anterior and lateral lobes are more mobile. The thoracic cavity expands anteriorly, laterally, and vertically due to the movement of the ribs, sternum, spine, and diaphragm. The structure of the human lung is shown in Figure 1.1.

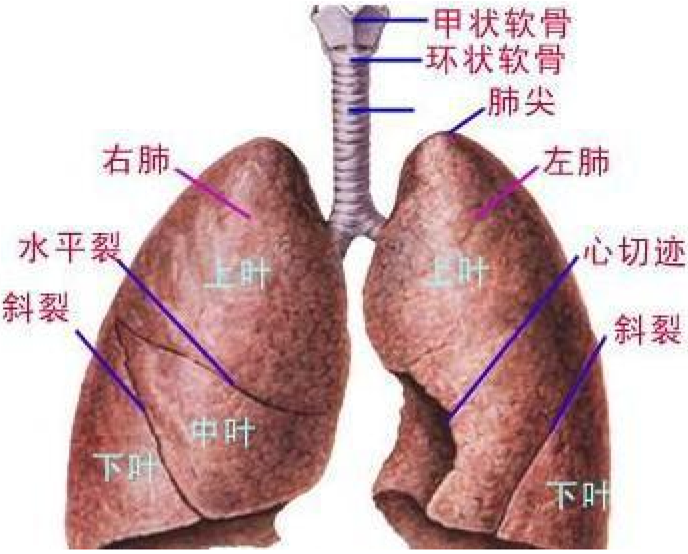


Figure 1.1 Diagram of human lung structure[①]

Andoh H (1978) summarized the displacement and changes in the corresponding respiratory organs resulting from the aforementioned lung movements into four categories.

(1) changes in anterior-posterior diameter: during inspiration, the ribs are elevated using the vertebral joints as fulcrums, thus causing the sternum to move forward and upward, increasing the anterior-posterior diameter of the rib cage.

(2) Changes in left and right diameters: the angle between the axis of motion and the sagittal plane, formed by the articulation of the thoracic vertebrae, thoracic vertebral bodies and transverse processes, is closer to a right angle compared to the higher ribs. Thus, the

① ①李宏奇.　科普：肺部结构[Z/OL]. 空总放疗科. 2006. https://www.sohu.com/a/57986244_104940.

elevation of the upper ribs leads to an increase in not only the anterior-posterior diameter of the ribs, but also triggers an increase in the lateral (left-right) and anterior-posterior diameters of the lower ribs.

(3) Changes in longitudinal diameter: the diaphragm is distributed over the lumbar vertebral body, the lower rib cartilage, the lower sternum, the 11th and 12th ribs, and runs mainly over the central tendon. The overall upward motion of the upper ribs and sternum causes an upward expansion of the thoracic cavity, while the contraction of the diaphragm triggers a downward expansion of the thoracic cavity. However, this movement is limited by the organs in the abdominal cavity, so that when the downward movement stops, the diaphragm will try to pull the ribs upward, and this will lead to an increase in the anterior-posterior and lateral diameters of the lower ribs.

(4) Reduction of the thoracic cavity: the internal retraction of the ribs and sternum and the upward movement of the diaphragm will lead to a reduction in the volume of the thoracic cavity.

Under normal inspiratory conditions, the ribs expand, the diaphragm goes down, the anterior thorax extends forward, the lateral thorax extends outward, and the abdomen bulges; during exhalation, the ribs contract, the diaphragm goes up, and the external observation is the opposite of that during inspiration[①].

Breathing is regulated by a combination of factors[②]. The inspiratory and expiratory centers exert opposite influences on the respiratory muscles; the inspiratory center stimulates

---

① Hixon, T. J. , Goldman, M. D. , & Mead, J. . (1973). Kinematics of the chest wall during speech production: volume displacements of the rib cage, abdomen, and lung. *Journal of Speech Language and Hearing Research*, 16(1), 78.

② Andoh H . Anatomy and Kinesiology of Respiratory System[J]. Journal of Japanese Physical Therapy Association, 1978, 4.

the inspiratory muscles and inhibits the expiratory muscles, whereas the expiratory center stimulates the expiratory parasympathetic center and inhibits the inspiratory muscles; in addition, during expiration, the extension of the alveolar wall excites the vagus nerve, which in turn inhibits the inspiratory center through the expiratory center; the vasodilation or constriction of blood vessels due to changes in blood pressure also causes reflex respiratory control. Finally, stimulation of the cerebral cortex also allows us to freely control breathing in states such as vocalization and snoring.

In summary, the diaphragm, as the main respiratory muscle, cooperates with other internal organs of the abdomen in respiratory activity, but each retains a certain degree of freedom of movement. it is not easy to determine the contribution of an active and passive respiratory organ to changes in lung volume during respiration alone, but it is perfectly possible to measure the relative motion of the thorax, septum and abdomen during respiration by means of controlled variables. Konno & Mead (1967) recorded changes in the anterior-posterior diameters of the ribs and abdomen on a direct-written X-Y recorder axis to measure the relationship between changes in motion of the two as influenced by respiration and changes in cavity volume in the presence of a fixed lung volume[①]. The MLT1132 piezoelectric respiratory band sensor manufactured by AD Instrument in Australia is based on this research idea. In addition, there is also the Phonatory Aerodynamic System, which measures the sound pressure level, EEG, and air flow during respiration, a measurement tool that can be used in the medical field in addition to phonological research.

Ji-deok Yin and Jiangping Kong(2007) investigated the respiratory rhythmic features of

---

① Konno K , Mead J . Measurement of the separate volume changes of rib cage and abdomen during breathing.[J]. Journal of Applied Physiology, 1967, 22(3):407.

the Korean read-aloud corpus using respiratory belts and compared them with the results of a study on the respiratory rhythm of the Chinese news read-aloud corpus in Jingjing Tan and Jiangping Kong(2006), and found that the respiratory reset point and speech pause position remained basically the same in the Korean news read-aloud corpus, while Chinese showed a mismatch in the co-occurrence of the two, and the reason for this is attributed to the difference in the rhythmic structure between Korean and Chinese. Haihong Qu and Jiangping Kong (2012) examined the abdominal breathing characteristics of the art of announcing and found that announcing with emotion has longer breathing duration and richer and more varied breathing resetting levels than reading without emotion. However, in the actual sampling process, it is difficult to guarantee the degree of "emotion" in pronunciation, so how to ensure the stability of this variable in the experimental process is a matter of concern, which also provides a reference and reflection for our experimental design[①]. Chunlian Zhang and Jiangping Kong (2013) analyzed the breathing rhythm of news reading, and found that there are three levels of breathing rhythm in Chinese news reading, in which the duration of breathing resets in each level is more stable, and the F0 resets in the first level of breathing resets are not significantly different from the F0 resets in the second and third levels of breathing resets, and the F0 resets and breathing resets are not necessarily co-occurring. Most of the studies on breathing rhythm in Peking University focus on artistic phonetion and investigate the relationship between breathing resets and speech & voice parameters such as fundamental frequency and pauses.

Jinshang Zhang, Hu A-Xu, and Hongzhi Yu (2010) provided a more detailed statement of

[①] This is the result of a discussion with Jie Yang, a PhD student in the Phonetics lab at Peking University.

the physiological mechanisms of verbal breathing and how they apply in physiological phonological studies. In the same year, Yonghong Li, Jinshang Zhang, Shuwen Wang, and Hongzhi Yu (2010) conducted a multidimensional acoustic analysis of Tibetan news reading based on respiratory signals and found that Tibetan news reading was also divided into three levels of respiratory resetting, and that the number of words in the clause before and after the respiratory resetting had an effect on the magnitude of respiratory resetting. Subsequently, Yangrui Yang, Hongzhi Yu, and Yonghong Li (2010) also examined the high correlation between breath-reset duration and reset amplitude, using Tibetan poetry as the object of study. Xuan Liu, Huaping Fang, and Yonghong Li (2013) investigated the respiratory unit and rhythmic characteristics of the Mongolian Short-tone folk song "Toasting Song" and found that the rate of rise and fall of the thoracic respiratory signal during singing was much greater compared to that of the ventral respiratory signal, which mainly helped to coordinate thoracic gas regulation, and that the resetting of the thoracic respiratory signal lagged behind the resetting of the ventral respiratory signal, which was the opposite of the respiratory resetting characteristics during normal speech . In the same year, Xuan Liu and Yonghong Li (2013) added two factors to the study based on Mongolian short folk songs, namely gender differences and differences in sung texts, and found that female singers tended to use thoracic breathing for articulation, while male singers tended to use combined thoracic and abdominal breathing, and both male and female singers had a slightly earlier resetting position of abdominal breathing than thoracic breathing during singing. Xuechen Yin, Liling Duan, Yonghong Li, and Hongzhi Yu (2014) examined the rhythmic characteristics of breathing in Song lyrics and used a confidence tool to compare the relationship between breathing duration and corresponding rhythmic unit duration in the silent and speech segments. Xuechen Yin, Shiliang Lu, Liling Duan, and Hongzhi Yu (2014) provided another review of the physiological mechanisms of respiration and the current state of research, and offered some new ideas for future research in this area. Yonghong Li (2015) gave a detailed description of the self-written and designed respiratory and airflow signal analysis system, which is similar to the speech respiratory rhythm analysis system written by Feng Yang (2012) at Peking University, both of which are based on the MatLab programming language to simultaneously realize the simultaneous analysis of speech, thoracic and abdominal respiratory signals, and the program adds airflow velocity analysis, air pressure analysis, and frequency domain map display three analysis channels. Xiaomei Gao (2019) conducted a rhythmic acoustic analysis of different genres of Arabic recitation, focusing on extracting three parameters of F0, power, and duration to perform acoustic analysis of Arabic poetry, prose, and macramé. Northwest Minzu University has long been concerned with the physiological phonology of speech and breath, and has taken advantage of two great advantages, namely the abundance of multi-ethnic speech resources and the analysis software edited in its own laboratory, to conduct in-depth analyses of Tibetan and Mongolian corpora, contributing to the preservation of ethnic languages and speech arts, and the discovery of corresponding features for better learning and preservation.

Jinyu Zhang (2014; 2016; 2017) successively investigated the breathing characteristics during Chinese singing, ordinary speech breathing, reading aloud, self-reading, and retelling states; she used a breathing acquisition tool other than the MLT respiratory belts, but the MP150 breathing data acquisition system produced by BIOPAC in the U.S. She found that the melodic group during singing influenced to a considerable extent the deployment of breathing and strategy selection; the correspondence between respiratory and rhythmic units was most regular during reading aloud, while the most complex respiratory states and the most disordered unit correspondence were found during self-reading. Dechong Deng (2017) conducted a comparative analysis of the phonetic and respiratory rhythms of both Cantonese chanting and reading aloud of metrical poetry, and when performing the extraction of respiratory parameters, Deng used the superimposed signal of thoraco-abdominal breathing to circumvent the differences in pronouncers' tendencies in the choice of pronunciation style, but because of the greater or lesser differences in the way pronouncers adopt thoraco-abdominal breathing grouping, this treatment is likely to have a masking effect on the relative variation of thoracic and abdominal breathing, and thus a more precise breathing pattern of Cantonese chanting cannot be examined.

In contrast to the domestic focus on speech art and oral culture, foreign scholars have focused on the relationship between speech and health, using the detection of respiratory signals as an indicator for diagnosing various medical conditions or natural human signs. Meanwhile, in the selection of individual research subjects, we can clearly observe the shadow of domestic scholars using as reference and conducting expansion or validation experiments.

Christopher Dromey et al. (1998) compared the effects of varying sound pressure level (SPL) and rate on respiratory, phonological, and articulatory behaviors during sentence production, and the authors' team collected data from multiple channels, including lung volume (LV), SPL, F0, semitone standard deviation (STSD), upper and lower lip displacement, and peak velocity. They found that loud speech resulted in increases in lung volume initiation, lung volume termination, F0, STSD, upper and lower lip articulatory displacement and peak velocity; and that the variability of these articulatory rhythms generally decreased with increasing sound pressure level; that F0 increased with rate in males and STSD in both sexes; and that the faster the speech rate, the smaller the lower lip displacement. Since SPL and speech rate variability are important concerns in the treatment of dysarthria, this data also provides a study of rehabilitation treatment options for such patients.

Kathryn Hird et al. (2001) explored the rhythmic and respiratory features of spontaneous conversational communication and found that pitch dip occurred at the end of almost all sentences, but that fundamental frequency resetting occurred only when the respiratory unit coincided with the rhythmic unit (in the context of normal communication situation). In the same year, David H. McFarland (2001) investigated the breathing movements of 20 subjects during recitation and spontaneous monologue, as-is dialogue, and spontaneous conversation, considered the correlation between inspiratory duration, expiratory duration, and respiratory reset duration and verbal rhythm, and found that inspiratory duration was the most consistent and sensitive parameter for distinguishing quiet breathing from verbal breathing.

Kathryn P. Connoghan et al. (2004) investigated the development of breath-driven vocalizations through longitudinal observations of chest wall movements in four normally developing children aged 9 to 48 months. She measured the relative contribution of thoracic and abdominal movements during vocalization (e.g., dentition or production of real words) and resting breathing every 3 months using the Respitrace and found that there was a difference between vocalized and resting breathing in terms of thoracic movements, with thoracic-abdominal coupling during vocalization diminishing with development, whereas coupling for resting breathing remained strong during the observation period , i.e., based on the fact that verbal breathing differs from resting breathing in thoracoabdominal respiratory grouping patterns, longitudinal changes in respiratory characteristics in children of different ages may be related to developmental changes such as changes in thoracic cavity shape and adaptation.

Sally Collyer et al. (2006) investigated whether the use of standard breathing and ventilated masks affected breathing performance in terms of lung volume (movement, onset and termination) or duration (inspiration and expiration) in both speaking and singing contexts, ultimately concluding that ventilated pneumotachograph masks do not affect breathing behavior in speech studies, a finding that also applies to singing during This finding is also applicable to the study of breathing behavior during singing. Subsequently, Sally Collyer et al. (2008) explored the effects of F0 and SPL range on breathing behavior in classical singing and found that the direction of the effect of F0 on left ventricular beat (LYE) varied between singers, but that SPL range appeared to be less important than duration for LYE. LYE was generally evenly distributed between crescendo and decrescendo. Each singer maintained her characteristic movement patterns, both in terms of F0 and SPL range, although these did affect aspects of RC and AB behavior. Considering the important role of breathing in classical singing, further work is needed to understand how singers develop their highly individualized breathing strategies.

There were also differences in the type of verbal respiratory movements that the human body tends to use at different ages and in the weighting of factors that influence the type of respiratory movements. Jessica E. Huber et al. (2008), on the other hand, explored the effects of natural aging on the process of respiratory support for speech production when speech length is constant, and she found that older individuals had higher lung volumes than younger individuals. There were no significant differences between older and younger women. Older adults tended to use more abdominal movements during loud speech compared to younger adults, especially when speaking in noisy environments. Some of the mechanisms used by older adults to support a greater response to signals were different from those used by younger adults. Age differences are more pronounced when longer utterances are produced. The authors suggest that reduced chest wall compliance, pulmonary elastic recoil and laryngeal closure may be used to explain these results, while the data may also be used to help distinguish whether the relevant changes are normal age-related or paradoxical disease changes. Subsequently, Corol A. Boliek et al. (2009) studied verbal breathing in a total of 60 healthy children aged 4, 5 and 6 years and found that children's breathing behavior was influenced by height and age, but not by gender, and that improvements in verbal breathing mechanisms were gradual, occurring roughly between 3 and 10 years of age. Douglas F. Parham et al. (2011), on the other hand, explored the breathing behavior of nine infants aged 53 ~ 90 weeks and found that 2-year-old infants showed significant differences in tidal breathing and breathing during speech, but not for syllable types with different articulatory requirements, and that older infants were more likely to maintain articulatory abilities associated with respiratory kinematics.

Rosemary A. Lester et al. (2014) analyzed the typical pattern of inspiration during speech breathing in healthy adults, as well as possible influencing factors, with subjects required to complete various speech tasks and collect nasal pressure, audio and video recordings. Inhalation patterns were categorized as nasal only, oral only, nasal and oral simultaneously, or alternating nasal and oral. Ultimately, it was found that the most common breathing pattern for healthy adults was simultaneous nasal and oral inhalation, across all speaking tasks, and independent of the nature of the speaking task or the speech environment before and after inhalation. However, this common feature of adult breathing patterns does not imply that the specific amplitude of the respiratory signal remains constant depending on the activity in which the human body is engaged. Raul I. Ramos-Garcia et al. (2016) approached this perspective and took the study further by exploring the feasibility of analyzing the temporal information of the respiratory signal through Hidden Markov Models (HMMs) to make a classification of human activity. HMMs are trained on five activities: sitting, walking, eating, talking and smoking, during which their airflow and temporal information of each respiratory segment are collected, their airflow signals are divided into smaller frames and the features of each frame are calculated; based on this, these frames are used as observations to model the state of HMMs by mixing Gaussian curves; finally using the leave-one-out method for cross-validation, the classification performance showed an average precision, recall and F-score of 60.37%, 67.01% and 62.78%, respectively. This study pushes the study of respiratory signals further to the application level, where it has the potential to be used as a primary or secondary source in the recognition of some human activities.

Susanne Fuchs et al. (2015) assessed the relationship between respiration and F0 decline and consistently showed that for spontaneous speech, the slope of thoracic motion was not related to F0 decline, that the slope of F0 during spontaneous speech output was smaller than the slope of speech F0 during reading aloud, and that F0 decline was influenced by sentence length but not by thoracic slope; also, although voiceless stops produces greater thoracic compression, it does not affect the F0 slope. These results suggest that although F0 slope decreases in many languages, it may not be due to purely physiological causes of breathing; other factors such as speech articulation planning and communication restrictions may interact with and ultimately cause F0 slope down along with physiological restrictions.

Mehrnaz Shokrollahi et al. (2016) proposed a novel neural network-based approach to classify snoring and non-snoring segments based on respiratory features, and their snoring detection algorithm was applied to nine patients with different levels of obstructive sleep apnea (OSA) severity. The sensitivity and specificity of this classifier reached 95.9% and 97.6%, respectively. The breathing feature-based detection approach can greatly help the accuracy of snoring detection and contribute to a more accurate diagnosis of sleep apnea.

Daniel J. Croake et al. (2018) quantified the interaction of the 3 vocal subsystems of breathing, phonation and articulation in 10 participants with healthy vocal cords before, during and after laryngeal disturbance (temporarily induced unilateral vocal cord paralysis) and found lower lung volume initiation and termination and longer respiratory excursions during laryngeal disturbance; airflow rate increased, while subglottal pressure remained constant.

It can be seen that the respiratory signal has an important position in the physiological representation of speech, and it can well reflect either individual or group differences (depending on the research contents) exhibited by individuals in phonation and performing other vital activities, and is therefore of interest and use by scholars in an increasing number of fields.

1.3 Research Problem

In the process of screening nature documentaries, it serves as a channel of sound information to supplement the narrative images, embellish the textual language of audiovisual communication, glue the structure of the work, and express the consciousness of the creative subject. However, in this process, the dubbing and its accompanying narration itself emphasize the importance of less as more, not to take over the dominant position of the picture content, and to follow the picture in time, with the sound starting at the beginning of the picture and ending at the end. These are the differences from ordinary speech forms such as newscasts and text readings. However, no scholars have explored the acoustic characteristics of this art from the perspective of experimental phonetics and based on empirical principles so far; when media students study this area of expertise, they can only imitate famous dubbing artists or even rely on their own so-called "perception" to figure out how to dub the documentary well. However, the development of modern speech technology has made it possible to quantitatively analyze the relevant acoustical parameters and performance characteristics of specific speech forms. Since its establishment by Fu Liu, the *Laboratory of Speech and Music* at Peking University has always focused on the empirical study of speech and its related cultural phenomena and ethnic instruments, and is equipped with excellent experimental infrastructure and measurement instruments[①].

Based on this background, we selected an episode of a well-known domestic nature documentary program *Animal World*, discarded the content of too many empty frames and too long pauses, and selected the clip that satisfied the characteristic of "steady and variable speech speed according to the needs of the content, instead of flat and straightforward all the time" as the object of study, and collected the speech, voice and breathing signals of the same pronouncer when dubbing and reading the same script, and the data of ordinary pronouncers without any dubbing training. The data were analyzed in these three dimensions to investigate the differences in breath resetting, F0, open quotient, and speed quotient, and then to observe which speech parameters or the grouping between them play a key role in the differentiation of this particular art of voicing.

1.4 Research Purpose and Significance

This research project takes nature documentaries as the entry point, excluding the existence of animal anthropomorphic dubbing, real-life interview communication or only the simple video and soundtrack performance without dubbing, and selects nature documentaries that combine video, narration-type narration, and background music. This will further provide a research paradigm for documentary dubbing and other forms of speech research, and also lay the foundation for the final speech synthesis. On the other hand, the special speech art phonation of dubbing is often accompanied by different patterns from normal voice, which shows more diverse EGG signals, and this also expands the cognitive boundary

① 孔江平. 语音科学与言语文化—北大语音乐律研究介绍[C]// 第七届中国语音学学术会议暨语音学前沿问题国际论坛论文集. 2006: 1-3.

of emotional speech research while enriching the EGG signal bank.

Chapter 2 Experimental Description and Research Methodology

2.1 Speakers and Dubbing Material

Since the selected segment belongs to a high-quality CCTV program and the original voiceover artist Zhongxiang Zhao has passed away, we finally chose a senior undergraduate student (male) and a second-year graduate student (female) from Communication University of China as our official voiceover teachers after multiple contacts and voiceover sample auditions. They both majored in broadcasting and hosting, had taken voice acting courses, had a solid professional foundation, had participated in extensive extracurricular voice acting practice due to their high interest in voice acting, and therefore had some experience in voice acting. In addition, we had two non-news broadcasting majors from Peking University as the pronouncers of the general reading, one male and one female. All of them were in good health, had no history of smoking, had no throat lesions, and were of moderate build, which ensured that the respiratory signals would not be collected with insufficient response detection or comparable data due to physical characteristics. However, due to the time and effort constraints, we mainly selected three speech texts from the two male speakers for the respiratory and acoustic analyses, but for the summaries of their respiratory characteristics, we also referred to the overall characteristics of the respiratory signals of their corresponding female speakers to ensure the reliability and stability of the conclusions.

In the selection of dubbing materials, we examined a series of nature science documentaries produced by the BBC in the UK, as well as a number of nature documentaries produced by the *Animal World* and *Exploration-Discovery* programs in China. The documentaries were then played to elderly native Chinese speakers (over 65 years old, 6 subjects) who did not know any foreign language at all, while the Chinese documentaries were played to young foreign subjects (6 subjects) who hardly knew Chinese in audio mode, and the information of each subject was as follows.

Table 2.1 Table of subject information

Subject number	**Age**	**Gender**	**Native language**
1	67	Female	Chinese
2	69	Female	Chinese
3	68	Female	Chinese
4	67	Female	Chinese
5	69	Male	Chinese
6	72	Male	Chinese
7	18	Female	Korean
8	21	Female	Korean
9	24	Female	English （Zimbabwe）
10	20	Male	Korean
11	20	Male	Korean
12	23	Male	Japanese

Then ask them, either in person or through online communication,

1. whether if the selection is similar to a nature documentary - yes or no (discrimination experiments).

2. if it is not like a nature documentary, what do you think the speaker is doing - open-ended question (confirmation experiment).

The results showed that the older native Chinese subjects and the young foreigners could make more accurate judgments based on the paralinguistic information at the phonological level, even though they could not grasp the semantic information to understand the language they were unfamiliar with. Of the 12 subjects, only a few answered negatively, and the overall accuracy rate of foreigners' judgments on "whether it is like a nature documentary dub" was 83.3%. Finally, based on the results of the survey and the proficiency of the speakers (if the speakers are not proficient enough in English, the accuracy of the breath signals will be seriously affected, because every pause or mispronunciation during the dubbing process will lead to abnormal fluctuations in the breath signals. Therefore, after consultation and communication with the pronouncer, we abandoned the dubbing of the English text), we chose an excerpt with the highest correct rate of foreigners as our dubbing material. The material is 10 minutes long and is about the migration of reindeer in North America and their efforts to escape from wolves, bears and other predators during the migration process, but their companions are still captured, however, the reindeer herd finally overcame all the obstacles to reach their migratory destination - the breeding area. The plot is tight, and the narration alternates between different types of descriptive expressions such as exposition, lyricism and description, which also facilitates the measurement of changes in breathing signals. Because the narration is created closely around the content of the picture, and the voiceover is a further phonetic processing of the narration, it is inevitably influenced by the content and potential emotional color of the narration, and the diversity of the types of narration descriptions has a great impact on the high quality of this speech, voice, and especially breath signal acquisition.

2.2 Experimental equipment and data collection

The recording of this experiment was conducted on November 8, 2020 in a professional recording studio in the Chinese Department of Peking University pair. The main equipment for data collection and the connection methods are as follows.

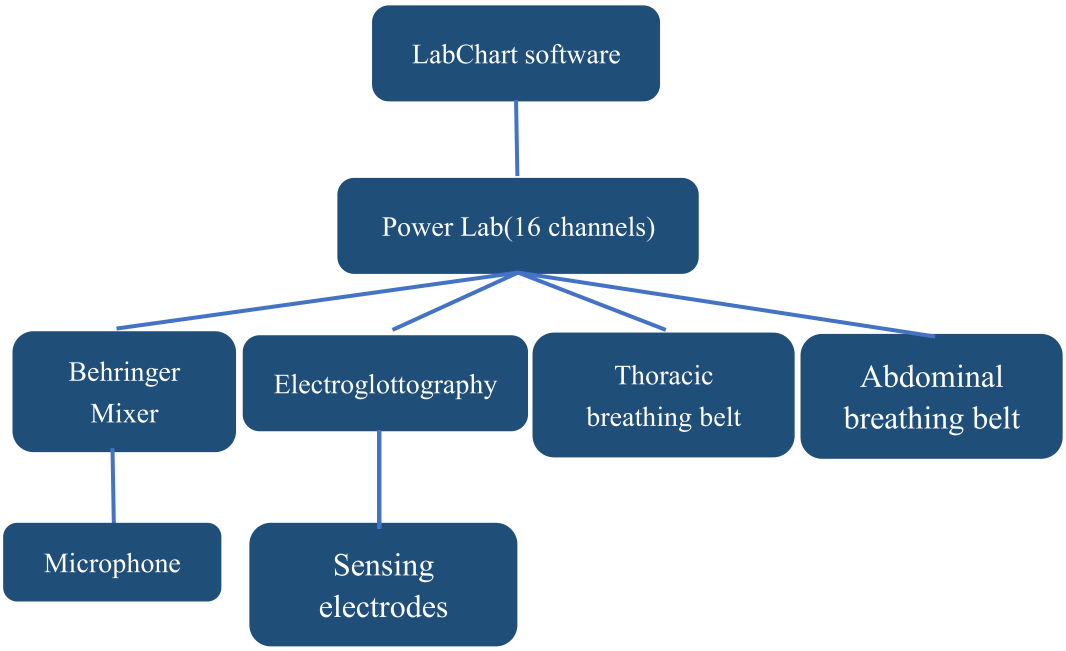


Figure 2.2 Schematic diagram of device connection

The speech signal in this study is collected by Sony condenser lavalier microphone, which is small, light and has a relatively uniform frequency response; the mixer is a XENYX502 analog mixer produced by Behringer, Germany, with the Sony microphone signal input from the first channel, BNC interface output, and access to the 9th channel of the PowerLab sixteen-channel collector.

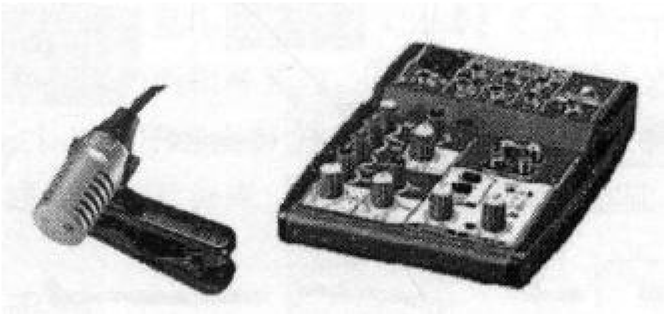


Figure 2.3 Sony microphone and Behringer XENYX502 type mixer

The laryngograph used in this experiment was an electroglottography 7050 manufactured by VoceVista of the Netherlands, which contains two main parts: the machine and two sensing electrodes.

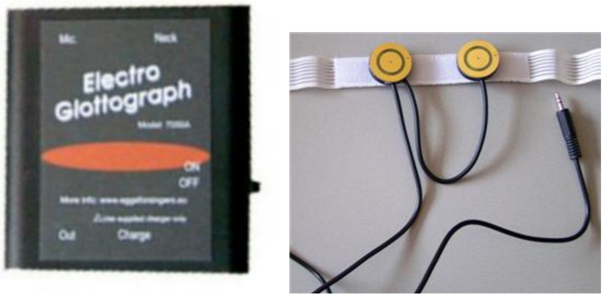


Figure 2.4 Electroglottography[1] and sensing electrodes[2]

The basic principle of the electroglottography is to fix two sensing electrodes on each side of the laryngeal node. In the working condition, the current will be sent from one electronic induction piece and received by the other, when the opening and closing of the vocal cords will cause the impedance of the current to change, and then this machine will record the measured current changes. At the same time, the non-invasive voice signal acquisition tool avoids the physiological pain caused to the subject during the study and also ensures the normal reading and analysis of the measurement results. The body of the laryngeal instrument was plugged into the PowerLab channel 10 through a BNC interface.

---

[1] Photo source VOCE VISTA official website, http://www.vocevista.com/electroglottograph/.

[2] Figure source of electroglottography and sensing electrode product details chart.https://b2b.baidu.com/m/land?id=76c1a33800860851e7bd5be77c602eab10.

The respiratory signal acquisition for this experiment was done through two respiratory belt sensors plugged into PowerLab channels 15 and 16 for acquisition and final recording into the computer. The respiratory belts were MLT1132 piezoelectric respiratory belt sensors manufactured by AD Instrument, Australia. The respiratory belts were connected to the PowerLab bioelectric collector manufactured by the company and used in conjunction with LabChart software for a total of four channels of bioelectric signal acquisition. The front panel of the sixteen-channel collector is a 16 BNC input port, which is output via a USB connection cable and connected to a computer.

Figure 2.5 Sixteen-channel collector① and breath belt sensor②

In this experiment, two breathing bands are tied to the chest and abdomen of the articulator (the abdominal breathing belt is placed three fingers below the articulator's umbilicus). When the articulator dubs or reads, the breathing will cause the chest and abdomen to contract and expand, so the change of the breathing band length during articulation can be detected by the piezoelectric device to determine the signal parameters related to exhalation and inhalation and obtain the breathing rhythm signal. Reflected on the LabChart display, the horizontal axis is time and the vertical axis is amplitude. The change of amplitude corresponds to the change of respiration: a rising respiration curve is inspiration, and a falling respiration curve is expiration.

① Photo source AD Instrument official website, https://www.adinstruments.com/products/powerlab-daq-hardware.
② 吴君如. 呼吸带说明.

The accompanying display and analysis software of PowerLab is LabChart7. In this speech experiment, a total of four signals are collected: channel 9 is the voice signal of dubbing or reading through microphone and mixer; channel 10 is the voice channel of speech and artistic articulation collected through laryngeal instrument; channel 15 is the signal of thoracic breathing collected through MLT1132 respiratory band sensor; channel 16 channel is the abdominal respiration signal collected through the MLT1132 respiratory band sensor, as shown in Figure 2-2-4. The sampling frequency is 20 kHz.

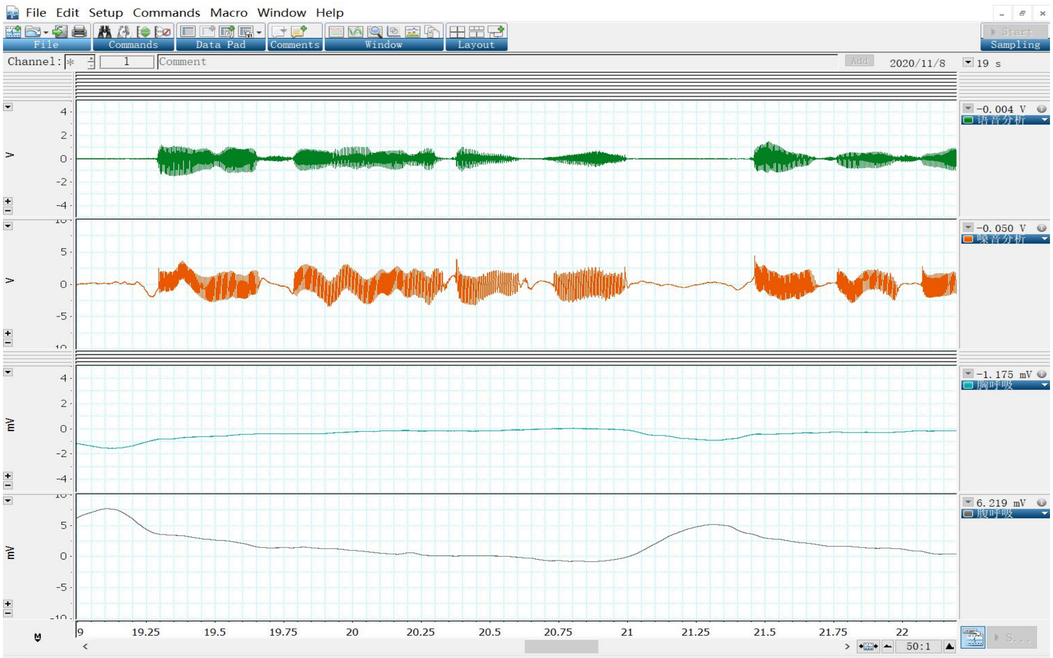


Figure 2.6 Four-channel signal acquired with LabChart 7

2.3 Research Methodology and Data Analysis

The four signals collected with Labchart7 are *.adicht format files, which cannot be directly used for refined analysis and statistics, so we exported them to *.wav format and read them into the speech rhythm breathing analysis system written by Dr. Feng Yang of Peking University in phonetics under the guidance of Professor Jiangping Kong using

Matlab①, as shown in Figure 2.6.

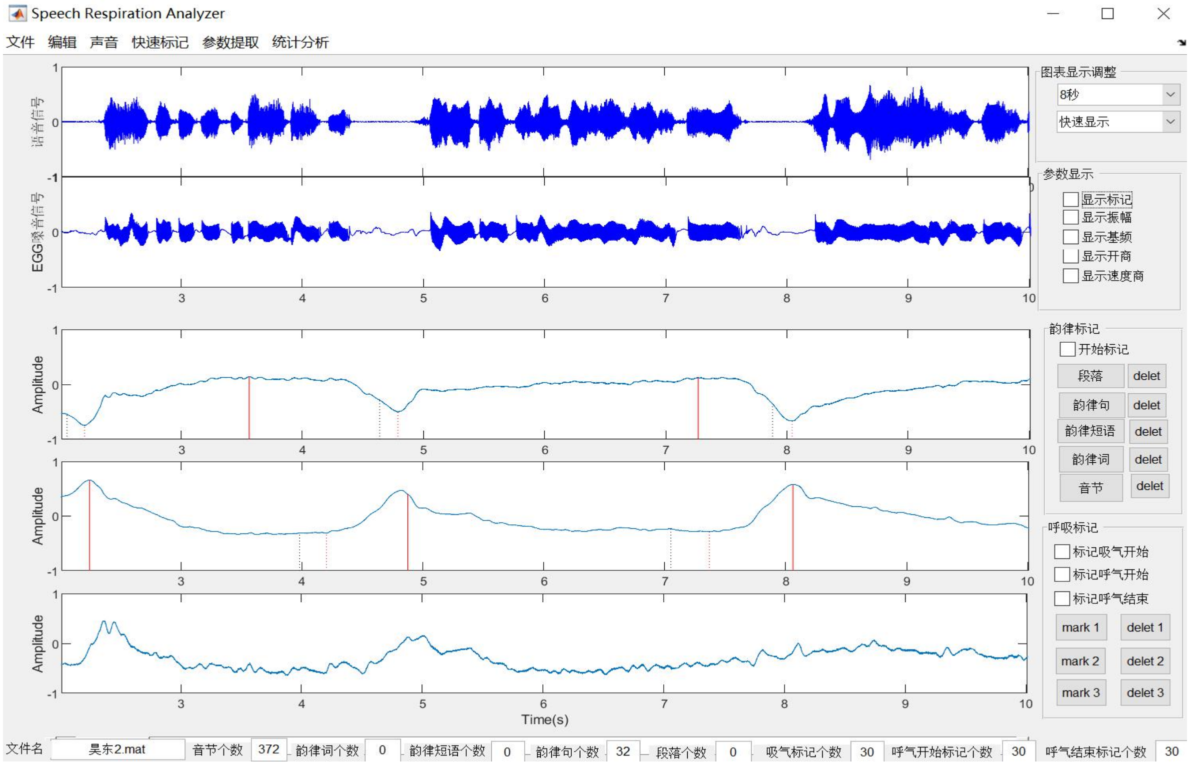


Figure 2.7 Speech Breath Analyzer

The program can realize simultaneous operation and comparative study of four signals, which provides great convenience for the processing of multiple corpus research objects. The speech waveform is displayed in the first channel from top to bottom, and we can first perform syllable tagging on the voice signals of dubbing and reading aloud, divide the rhyme units such as rhyme phrases and rhyme sentences according to the results of temporal clustering analysis of the pause duration between syllables, and output the parameters such as duration, fundamental frequency, amplitude and rhyme boundary duration of each rhyme unit to an Excel table for statistical analysis; meanwhile the calculation of rhyme units such as syllables and amplitude The display can be selectively presented in the speech waveform,

① The software has been applied for copyright protection by the creator Yang Feng: *Speech Breath Analyzer*, Registration No. 2012SR024971 Certificate No. 0393007; *Prosodic Marker*, Registration No. 2012SR041601 Certificate No. 0409637.

as shown in Figure 2.7, where the short green line indicates the syllable marker, the long green line is the left mouse button selected marker, the long black dotted line indicates the rhyming phrase marker, the short black dotted line is the rhyming phrase marker, and the red line is the amplitude marker (the program can also import the corresponding text library to mark the Chinese characters corresponding to the syllable segment at the bottom of the corresponding waveform. However, since there is too much data to be processed, importing the text would greatly slow down the program, so we did not import the text one by one, but only marked the data in the Excel table after extraction).

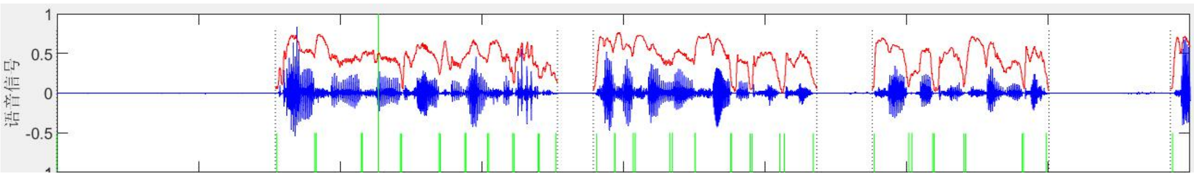


Figure 2.8 Program interface labeling display for speech signal

The second channel shows the EGG signal, which can be synchronized with the speech signal to extract the parameters of its corresponding syllable segment such as F0, open quotient, and speed quotient. As shown in Figure 2.8 below, the vertical green line is still the syllable marker, the red line indicates the fundamental frequency, the black line is the open marker, and the horizontal green line indicates the speed quotient. After the syllables are marked, they can be saved as *.mat files and the parameters of the F0, open quotient, and speed quotient before and after the normalization of each syllable can be output at any time.

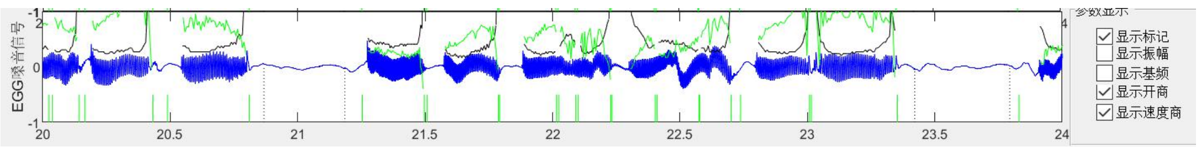


Figure 2.9 Program interface labeling display of EGG signal

The third, fourth and fifth channels show the thoracic respiration signal, abdominal respiration signal and superimposed thoracic and abdominal respiration signal respectively, where the red dotted line indicates the start of inspiration, the red solid line marks the start of expiration and the black dotted line marks the end of expiration, as shown in Figure 2.9. The analysis platform first needs to manually mark the time points of thoracic, abdominal or superimposed thoracoabdominal respiratory signals according to the study requirements, and then the system will automatically find the maximum value (for expiration start time point and expiration end time point) or minimum value (for inspiration start time point) in the 0.1s range at the marked time points. This "manual + automatic" processing mode maximizes the accuracy of the marker and reduces subsequent analysis errors caused by marker errors. During the tagging process, you can simultaneously pay attention to the statistical display of the number of inspirations, the beginning of expirations and the end of expirations in the bottom row of the program to ensure the reliability of data tagging by comparing the two. If an error occurs during the respiratory signal labeling process, you can use the left mouse button to click on the left side of the wrongly labeled line segment (shown on the platform as a solid green line with five channels of the same frequency), right click on the right side of the wrongly labeled line segment (shown on the platform as a solid red line with five channels of the same frequency), and ensure that there is one and only one wrongly labeled line segment within the selected range (i.e., the part sandwiched between the solid green and solid red lines) then click on the corresponding "Delete 1" (thoracic respiratory signal), "Delete 2" (abdominal respiratory signal), or "Delete 3" (thoracic and abdominal respiratory overlay) for the signal channel in which it is located. " (thoracic and abdominal respiratory overlay signals) to delete the selected segment. After the marking work is completed, the system will calculate and output the inspiratory phase and expiratory phase amplitude, duration, slope and other parameters of the thoracic respiration, abdominal respiration and their superimposed signals to Excel. In particular, it is important to note that the inspiratory and expiratory phase amplitudes are not the respiratory reset amplitude, but the specific values of their corresponding vertical coordinates, so the difference between the value of the vertical coordinate of the expiratory time point and the value of the vertical coordinate of the inspiratory time point should be used as the respiratory reset amplitude.

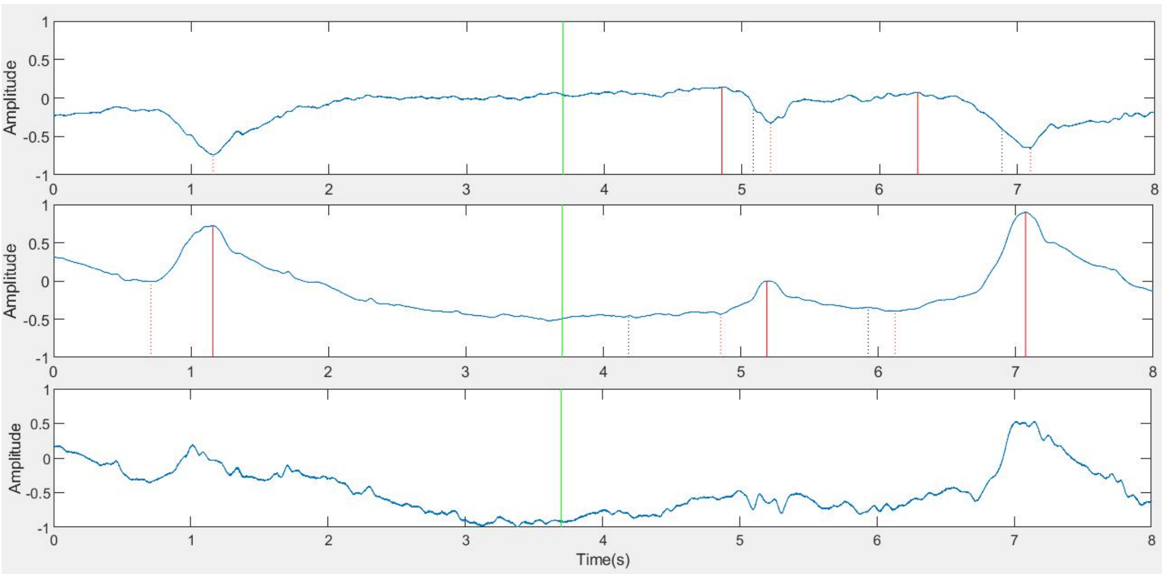


Figure 2.10 Program interface labeling display of thoracic respiration signal, abdominal respiration signal, and superimposed signal of thoracoabdominal respiration

In addition, since the looseness of the breathing belt straps affects the amplitude value of the respiratory signal display, the program performs amplitude-based normalization of both thoracoabdominal respiratory signals, thus ensuring comparability between the data results collected from different batches of experiments. Based on this, we obtained the respiratory rhythm characteristics of different pronouncers when dubbing or/and reading aloud based on the derived respiratory resetting durations and amplitudes, and observed how they differed from the artificially pre-defined rhythmic units based on semantics, based purely on the division of respiratory units during pronunciation and the location of respiratory resetting between professional dubbers and ordinary pronouncers. The EGG signal was also used to measure the F0 high and low point values, open quotient, and speed quotient of syllables before and after the breath boundary, and to compare and analyze the data with parameters such as the amplitude of breath resetting, the duration of specific rhythmic units before and after the breath boundary, and the slope and area of the expiratory and inspiratory segments. The data were also tested and analyzed using SPSS software.

2.4 Parameter Definition

The three voice parameters of F0, open quotient and speed quotient used in this paper are not defined based on the glottal airflow signal, but on the EGG signal. The former is based on the integral form of the voice, while the latter is based on the integral form of the glottal impedance, both of which can reflect the phonations from different perspectives[①]. Figure 2.11 shows a typical EGG signal.

[①] 孔江平. 语言发声研究及相关领域[A]. 中国中文信息学会.新世纪的现代语音学——第五届全国现代语音学学术会议论文集[C].中国中文信息学会:中国中文信息学会,2001:8.

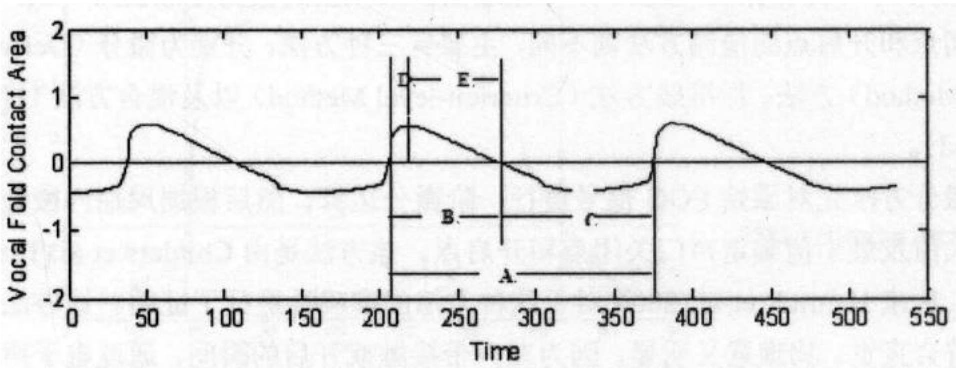


Figure 2.11 Typical glottis impedance signal parameter definition diagram[①]

The horizontal axis represents the time and the vertical axis represents the contact area of the vocal cords. *A* is a period of voice cycle, i.e., the time interval from the previous glottis closing point to the next glottis closing point; *B* and *C* are a closed phase (the time interval from the glottis closing point to the glottis opening point) and an open phase (the time interval from the glottis opening point to the next glottis closing point), respectively; *D* and *E* are the closed phase (the time interval when the glottis is gradually closing) and open phase (the time interval when the glottis is opening). Based on this, we can define the three parameters based on EGG signal.

Fundamental frequency (Pitch) = 1/Voice period (A).

Open quotient (Open Quotient/OQ) = open phase (C) / voice cycle (A) × 100%.

Speed Quotient/SQ = open phase (E) / closed phase (D) × 100%.

The OQ and SQ parameters obtained using the EGG signal can be used to distinguish the glottis movement during speech articulation. The analysis platform used in this paper adopts the method of artificially delineating a 25% standard line in calculating the relevant parameters of the EGG signal, which means that the glottis closure and opening points are positioned at the 25% horizontal line of the difference between the maximum and minimum

[①] 杨锋.中国传统吟诵研究[D]. 北京大学, 2012：28-32.

values in each signal period.

The respiratory parameters extracted in this experiment are mainly the respiratory reset duration and reset amplitude, and Figure 2.12 shows a schematic diagram of the respiratory signal variation.

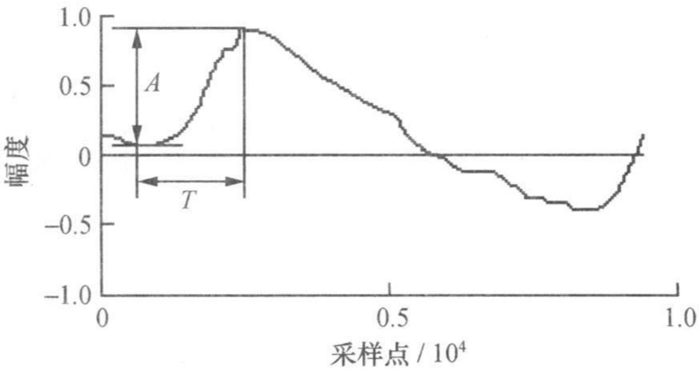


Figure 2.12 Schematic diagram of respiratory signal changes

The horizontal axis represents the time, and the vertical axis represents the voltage change due to elongation and contraction collected by the respiratory belt, which is shown as a curve in the schematic diagram, i.e., the respiratory curve, where a rising respiratory curve indicates inspiration and a falling respiratory curve indicates expiration. The main respiratory parameters extracted in this experiment are the reset amplitude, the reset duration, and the slope of the inspiratory segment and the slope of the expiratory segment, which are defined as follows.

Respiratory reset amplitude (REA): the longitudinal amplitude from the point where inspiration begins to the point where exhalation begins, i.e., the value of *A* in Figure 2-4-2.

Respiratory reset duration (RED): the lateral time interval from the point where inspiration begins to the point where exhalation begins, i.e., the *T* value in Figure 2-4-2.

Inspiratory segment slope (ISS): the slope of the respiratory curve between the point at which inspiration begins and the point at which expiration begins, as a positive value.

Exhalation segment slope (ESS): the slope of the respiratory curve between the point where exhalation begins and the point where exhalation ends, as a negative value.

Chapter 3: A Comparative Study of Dubbing and Reading

3.1 The Characteristics of Speech rhythm and Breathing Rhythm in Dubbing

In nature documentaries, many shots showing the survival and living conditions of wild animals are directly tracked from a distance or close up and edited in post, making the images more focused and easier to be understood by the audience, while also achieving the effect of using the least amount of screen language to explain the most meaningful content. In this process, narration is often added to explain the context, and the details involved before and after are linked into a complete and vivid story. Although the current documentary documentaries are increasingly emphasizing the importance of "letting the images speak"[①] and making the audience understand and judge the joy, sorrow, right and wrong based on the language of the images, this is certainly a conscious approach to the pursuit of objectivity in documentaries. However, if nature documentaries only have natural and realistic images, their expressions are often too general and ambiguous, making it difficult for people to understand the deeper meaning of the documentary. In practice, narration mostly plays the function of linking shots, sublimating themes, describing narratives, and suggesting supplements. This requires the dubbers to communicate in detail

[①] 董春晓. 纪录片制作教程[M]. 浙江大学出版社, 2014: 131-132.

with the choreographer and other workers before the actual dubbing, to refine the emotions to be expressed in each segment of the content, and to be ready to adjust their voices to better fit the specific images. Therefore, in the process of dubbing, professional dubbers will constantly adjust their pronunciation with the content of the picture and the need to express emotions, and with the regular grouping of thoracic and abdominal breathing, to master as much as possible the skills of audible speech expression, so as to finally make the best adaptation to the documentary content and the establishment of the character's perspective.

This section will summarize the regular features of the professional voiceover narration from two perspectives: speech rhythm and respiratory rhythm, and specifically examine whether the rhythmic units formed based on semantic hugging are consistent with those delineated based on respiratory resetting, and what intricate correspondence exists between them.

3.1.1 Speech Rhythm of Dubbing

By tagging the collected speech data with syllables, rhyming phrases and rhyming sentences, we obtained the basic data of the rhyming units in the dubbing and reading corpora based on the pause duration, including syllable duration, syllable pause duration, rhyming phrase duration, rhyming phrase interval duration, rhyming sentence duration, rhyming sentence interval duration, and the 10-point output of each syllable normalized by the F0, OQ, and SQ. In the study of Deng (2019) on the comparison of phonological rhythm and respiratory rhythm, he defined a phonological rhythm as "a phonological unit with the maximum phonological stop at the boundary" and a rhyming phrase as "a phonological unit with the maximum phonological stop at the boundary"[①]. This definition is imprecise because "maximum" is an extreme category, and the end of a sentence is not marked only by the one with the maximum stopping time[②]. Other sentences do not meet the constraints of the definition of "maximal", but it is still possible to determine whether a sentence is a phonological-rhythmic unit based on the meaning of the sentence and the pause. Likewise, the concept of "larger" phonological pauses is more ambiguous. At present, the academic community has interpreted the rhyme unit from various perspectives, including semantic and syntactic. It is generally accepted that there are three rhythmic units in Chinese, namely, rhyming words, rhyming phrases, and intonation phrases[③]. The rhyming phrase can express a more complete semantic meaning than the rhyming phrase. According to this formulation, there still seems to be a partial crossover between rhyming phrases and rhyming sentences. For example, if a sentence consists of only one rhyming phrase, should we classify it as a rhyming phrase or a rhyming sentence? In this paper, all rhyming components containing only one rhyming phrase are treated as rhyming sentences, and only when a sentence contains more than one rhyming phrase component with obvious pauses, we cut the rhyming sentences internally and continue to divide the rhyming phrases further. Since we did not examine respiratory resetting at the syllable level in conjunction with the observation of thoracoabdominal respiratory signals, syllable-based respiratory rhymes are not discussed. In fact, in the actual data annotation, it is difficult to divide the rhythmic units based on syllable

① 邓德崇. 格律诗词的粤语吟诵和朗读的语音及呼吸韵律研究[D]. 暨南大学，2019：20-21.

② The corpus used by Deng is poetry, and the stopping and rhythmic features of poetry are more stable than those of audio-media languages. Therefore, although Deng's definition is inaccurate, it is not too problematic when applied to this particular textual corpus.

③ 曹剑芬. 汉语韵律切分的语音学和语言学线索[A]. 中国中文信息学会.新世纪的现代语音学——第五届全国现代语音学学术会议论文集[C].中国中文信息学会:中国中文信息学会,2001:4.

stopping in most cases, and the semantic syntax always interferes with the division of units to a greater or lesser extent. Meanwhile, the rhyming word part has different manifestations in different corpora, sometimes clearly distinguishable and sometimes vague, so it is impossible to make a cross-sectional comparison based on statistical data, therefore, we exclude the rhyming word partition when dividing the rhyming units, and only include the rhyming units of syllables, rhyming phrases and rhyming sentences, and especially examine the F0, OQ and SQ at the end of rhyming sentences in the acoustic study.

In the dubbing, the syllables' duration and interval duration are normally distributed, as shown in Figure 3.1 and Figure 3.2 ("hxd" in the title of the figure is the initial letters of the pronouncer's name, the same below), and the average syllable duration is 0.174s with a standard deviation of 0.093. The average syllable interval length was 0.054s, and the actual average syllable interval (i.e., the average of the syllable intervals after excluding the syllable interval at the boundary of the rhyme scheme) was smaller than that measured so far because the program automatically included the syllable interval length at the boundary of the rhyme scheme during the statistical process.

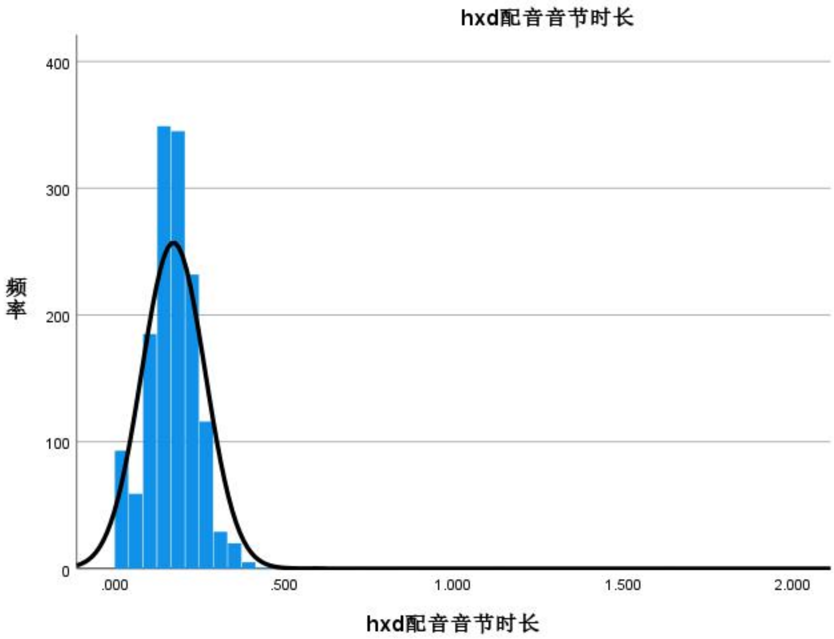


Figure 3.1 Histogram of frequency distribution of dubbing syllable duration

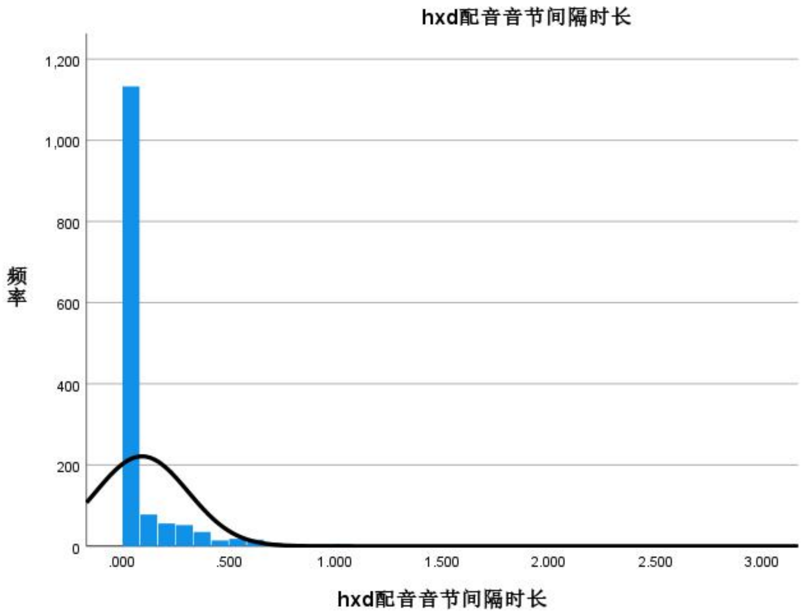


Figure 3.2 Histogram of frequency distribution of dubbing syllable interval duration

The rhyme phrase duration and rhyme phrase interval duration still conform to normal distribution, as shown in Figures 3.3 and 3.4. In the dubbing corpora, the mean value of rhyme phrase duration is 1.285s, and the standard deviation is 0.511. The varying length of rhyme phrase height shows the characteristics of the rhythmic variability of the dubbers during dubbing, and the majority of rhyme phrase duration is 3 times to 20 times higher than that of a single syllable, which is due to the fact that, according to the expression of the picture content, the voiceover artist often needs to pour in the narration within a specified duration, which greatly compresses the time for pauses between rhyming words, while the internal rhyming phrases mainly meet the overall requirement of a soothing tone through the accent and the extension of the end syllables of individual rhyming words, so as not to give the audience a hard and anxious listening experience. The average interval between rhyming phrases is 1.859s, which almost touches the upper limit of the length of pause at the boundary of rhyming phrases. The excessively long interval of rhyming phrases appears in the statistics because the interval of rhyming phrases at the boundary of rhyming sentences is also counted, so that the dubbing does not pause too long at the boundary of rhyming phrases in terms of the simple interval of rhyming phrases, otherwise it would be difficult to guide the audience to express the complete meaning of the picture.

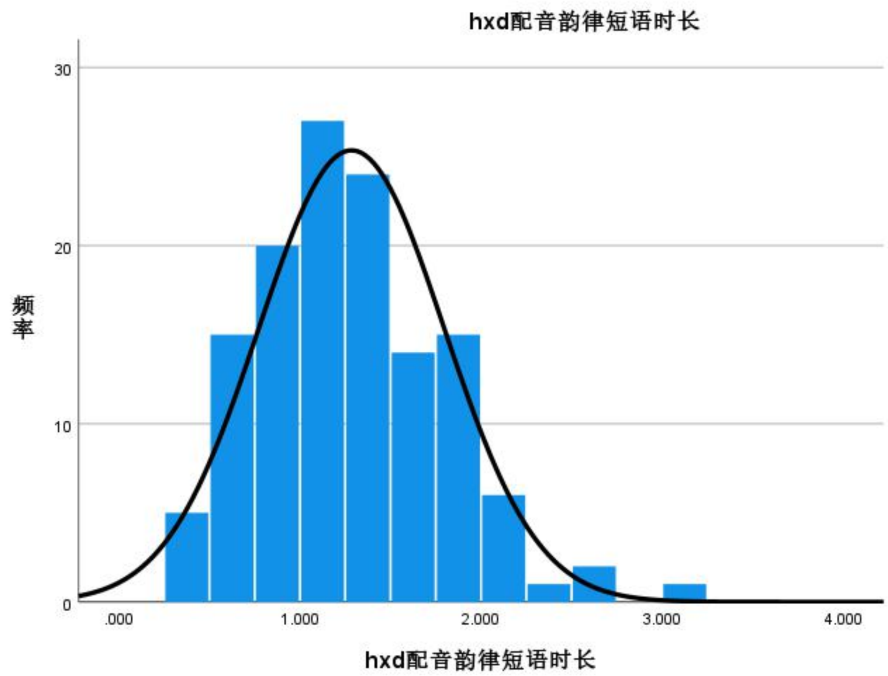


Figure 3.3 Histogram of frequency distribution of dubbing rhyming phrase duration

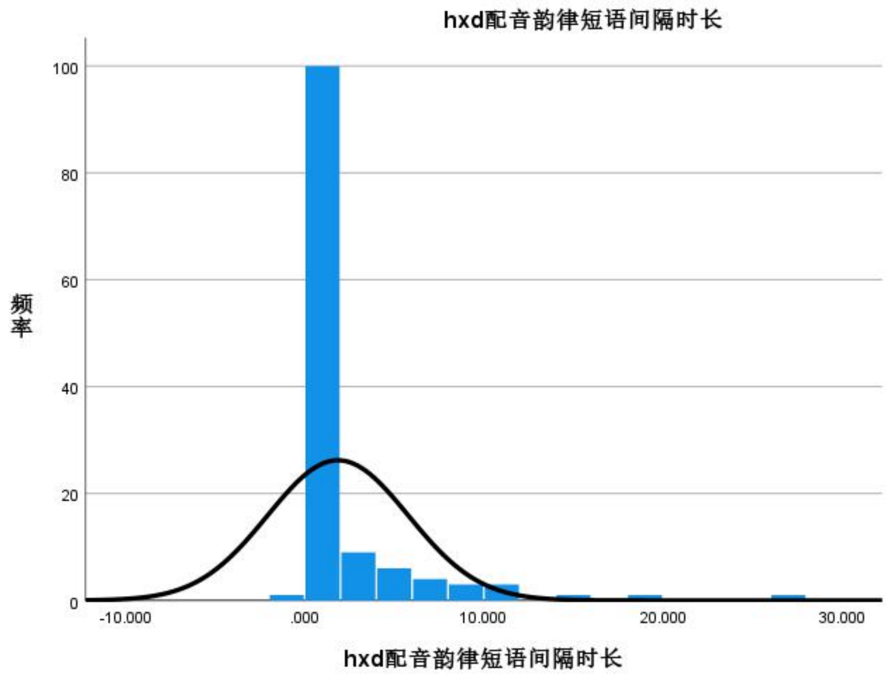


Figure 3.4 Histogram of frequency distribution of interval duration of dubbing rhyming phrases

The average length of rhyming sentences in dubbing is 2.422s, with a standard deviation of 1.101. The rhyming sentences are mainly concentrated in the interval of 2~3s, and according to the dubbers' processing, each sentence contains about 11 syllables on average, which is still in line with our listening habit, i.e., about 4 Chinese syllables are output to the outside world per second, and there is little internal variation in the length of rhyming sentences, which is not only the customary rhyming unit of the dubbers' physiological pronunciation This is not only the customary rhythmic unit of the voiceover artist's physiological pronunciation, but also invariably takes care of the audience's understanding under the traction of the narration. The average pause between sentences is 0.664s, with a standard deviation of 0.579. This is due to the fact that the voice-over artists have to arrange the specific position of the narration words according to the presentation of the picture content, so the average pause length is relatively long and the length of the interval varies.

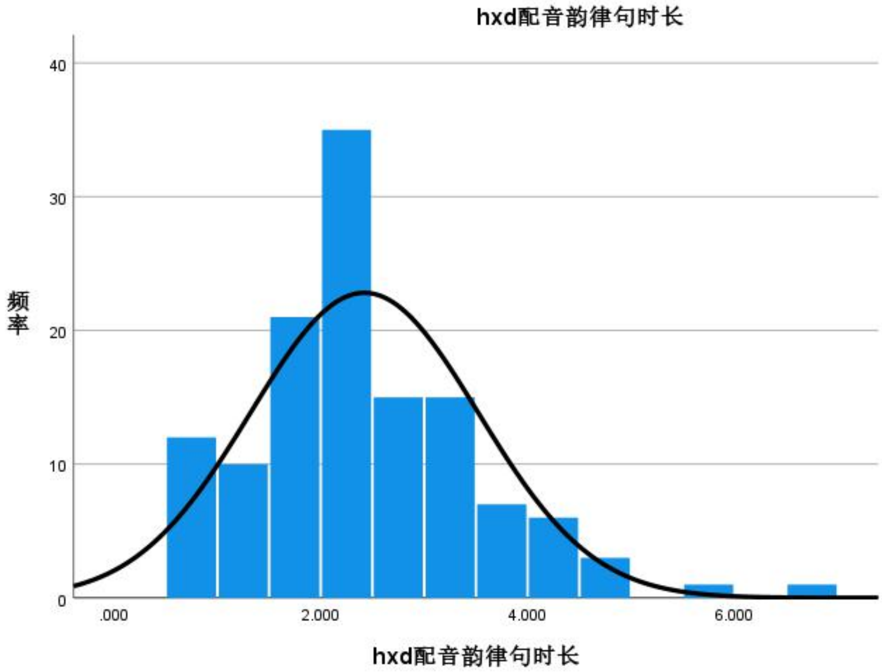


Figure 3.5 Histogram of frequency distribution of dubbing rhyming sentence duration

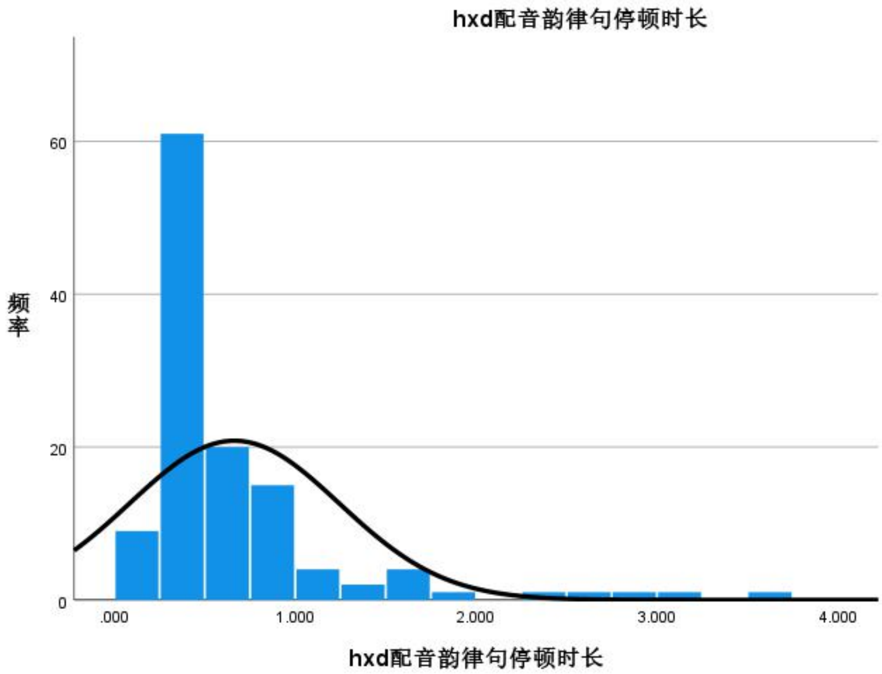


Figure 3.6 Histogram of the frequency distribution of pause duration in dubbing rhyming sentences

3.1.2 Breathing Rhythm of Dubbing

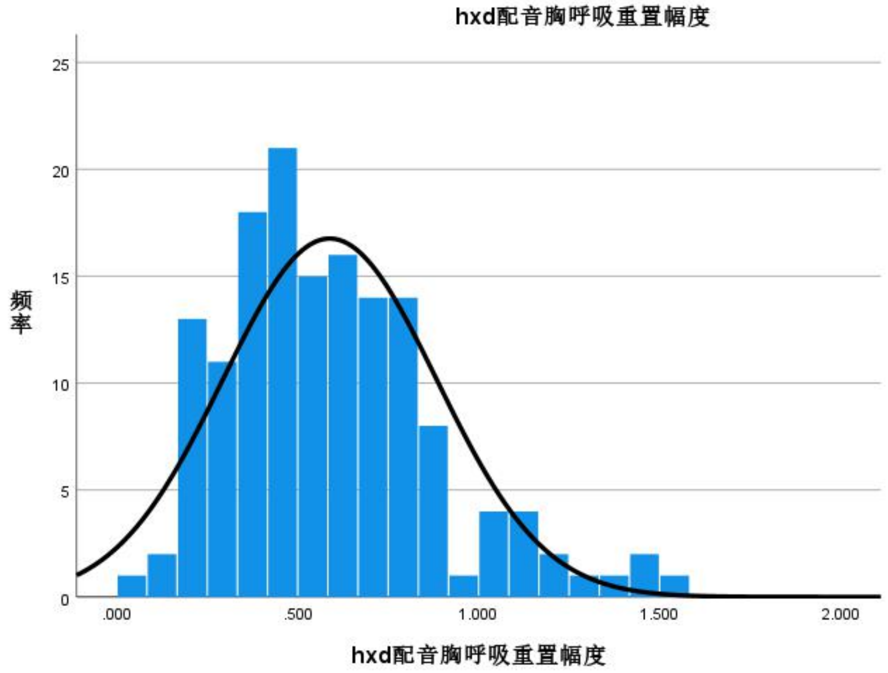


Figure 3.7 Histogram of frequency distribution of dubbing thoracic breathing reset amplitude

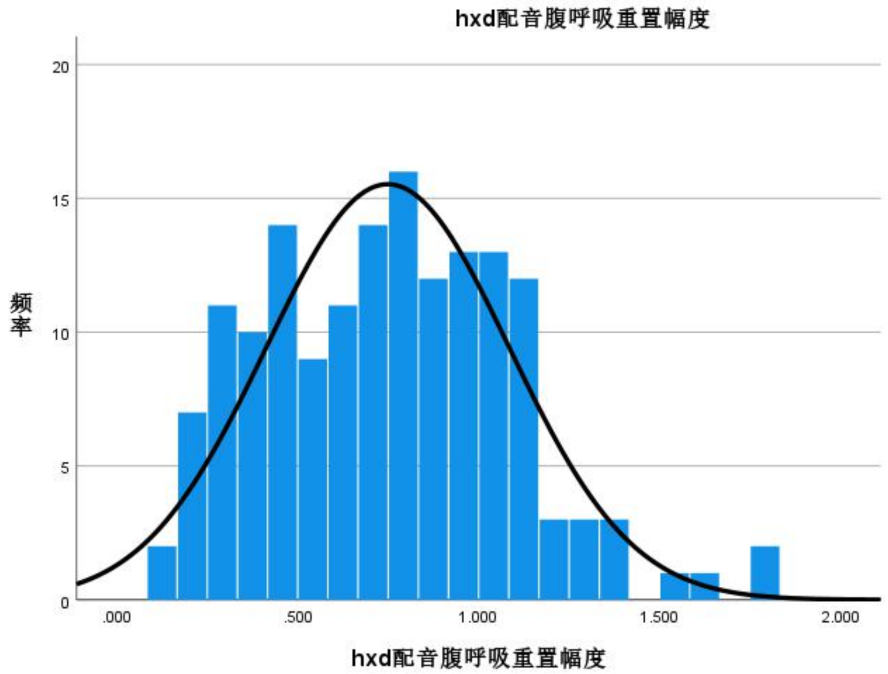


Figure 3.8 Histogram of frequency distribution of dubbing abdominal breathing reset amplitude

As shown by the histogram of frequency distribution of thoracic and abdominal respiratory resetting amplitude of the dubbed corpora above (a total of 149 for thoracic respiration and 157 for abdominal respiration), the respiratory resetting amplitude during dubbing basically conforms to a normal distribution, with a large peak at 0.5 for thoracic respiratory resetting amplitude and a more stable respiratory resetting at the two amplitude intervals of 0.3~0.8 and 1.0~1.6, which indicates that the thoracic breathing resetting in voiceover can be roughly divided into two levels. Corresponding to the dubbing style, i.e., small respiratory resets occur in each rhyming phrase, but larger respiratory resets occur only when there are longer silent segments in the picture, when the picture shot changes, or when the introduction object changes. The abdominal breathing reset amplitude also basically conforms to the normal distribution, and the average abdominal breathing reset amplitude is 0.749, which is higher than the thoracic breathing reset amplitude of 0.589. The abdominal breathing reset can also be divided into two levels, 1.3~1.8 for the primary breathing reset and 0.3~1.2 for the secondary breathing reset. The average amplitude and number of resets of abdominal breathing resets were higher than those of thoracic breathing, which reflects that voiceover artists can flexibly mobilize abdominal breathing to inhale more oxygen when performing voiceover and provide sufficient power for prolonged verbal expressions. We also noted that voiceover artists could sustain up to 35s of airflow exchange using mainly abdominal breathing after a large thoracic breath inhalation, when a very weak respiratory reset occurred only at the rhyming sentence boundary. As shown in the figure.

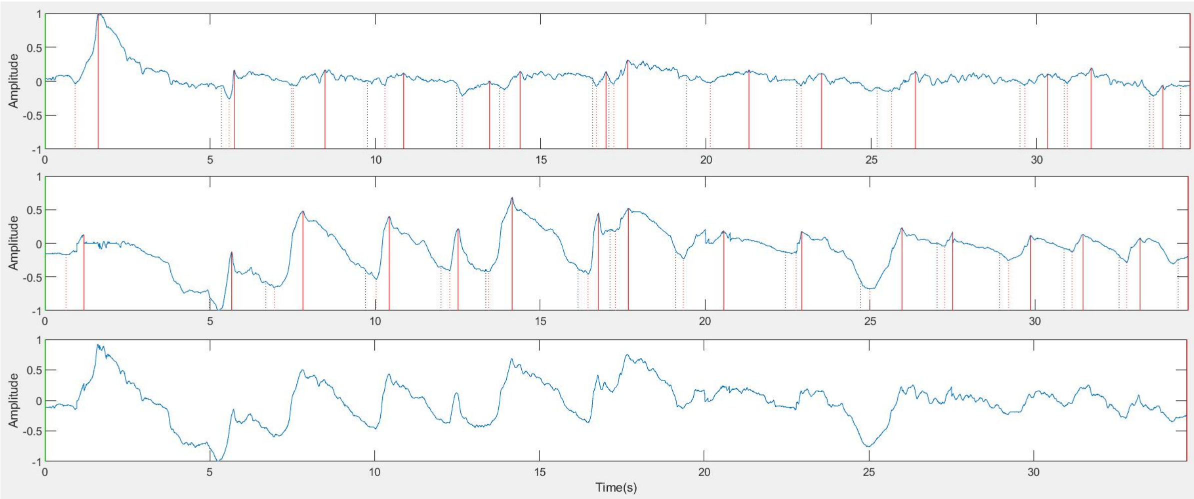


Figure 3.9 Schematic diagram of the dubbing thoracic and abdominal breathing signal

After the selection on the picture, there is a long silent segment, because the dubbers have previewed the whole narration content and its matching picture many times in advance, so they may be prepared for the breathing, and store a lot of power for pronunciation through thoracic breathing before the first rhyming sentence starts, while the abdominal breathing only shows a very small rise of inspiration. Then, during the subsequent dubbing process, the voice is gradually dubbed in accordance with the content of the picture, the rise of the thoracic breathing curve is weak, while the abdominal breathing takes over the task of providing the main power for pronunciation, and finally the first paragraph of the picture is dubbed completely. Professionally trained voice-over artists can flexibly deploy thoracic breathing and abdominal breathing to serve the stable pronunciation for as long as 35s, as we will see in the subsequent study, which is far beyond the reach of ordinary pronouncers. This breathing mode is very common for professional dubbers in the process of dubbing or announcing, and it requires high physiological quality of the body. On the one hand, this breathing pattern can avoid the occurrence of theft of breath caused by insufficient inhalation before the end of a sentence (this is the most taboo in the broadcasting genre, it tends to make the voice raw or unnaturally low, which affects the professionalism of the broadcast and the integrity of the expression); on the other hand, students of broadcasting are often involved in the learning and training of this breathing pattern in their daily professional training, which is also a preparation for the future broadcasting on camera: more reliance on abdominal breathing to provide power for speech pronunciation, and trying to maintain a more stable amount of airflow in and out of the chest after inhalation can make the announcer's shoulders and neck remain stable, and not easily affected by the influence of spitting back to the voice and causing the body above the chest to be upright and hold the weight, which can reap better visual aesthetic effects in the recorded program, and the audience will not be distracted from the content of the broadcast because the announcer's body is shaking[①].

In terms of the resetting time of thoracic and abdominal breathing, the mean value of the resetting time of thoracic breathing (0.786) was also lower than that of abdominal breathing (0.843), but the frequency peaks of both thoracic and abdominal breathing occurred around 0.3~0.7s, which indicated that the resetting time of this range might be the most consistent with the physiological needs of voiceover operators in both thoracic and abdominal breathing. Meanwhile, the standard deviation of the resetting duration of abdominal breathing was 0.96, while that of thoracic breathing was 0.578, which also indicated that the resetting duration of abdominal breathing was more regular than that of thoracic breathing, and the resetting duration of thoracic breathing was less stable than that of abdominal

---

[①] The analysis of the use of this breathing pattern by broadcast students in relation to on-camera requirements is from an exchange between the author and Dr. Jie Yang, a PhD student in the Phonetics Lab at Peking University. I would like to thank Jie Yang along with the analysis used above.

breathing.

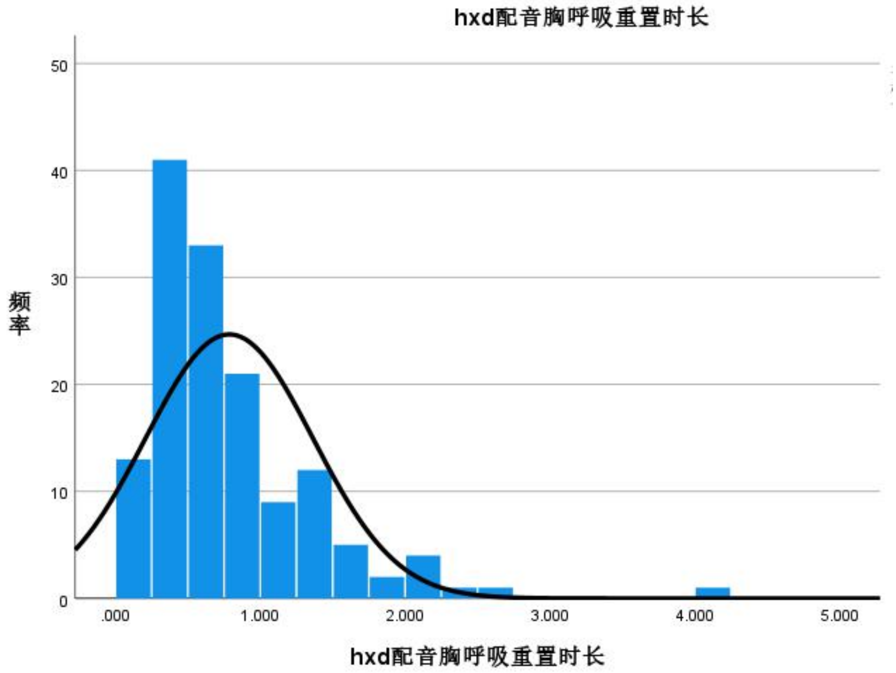


Figure 3.10 Histogram of frequency distribution of dubbing thoracic breathing reset duration

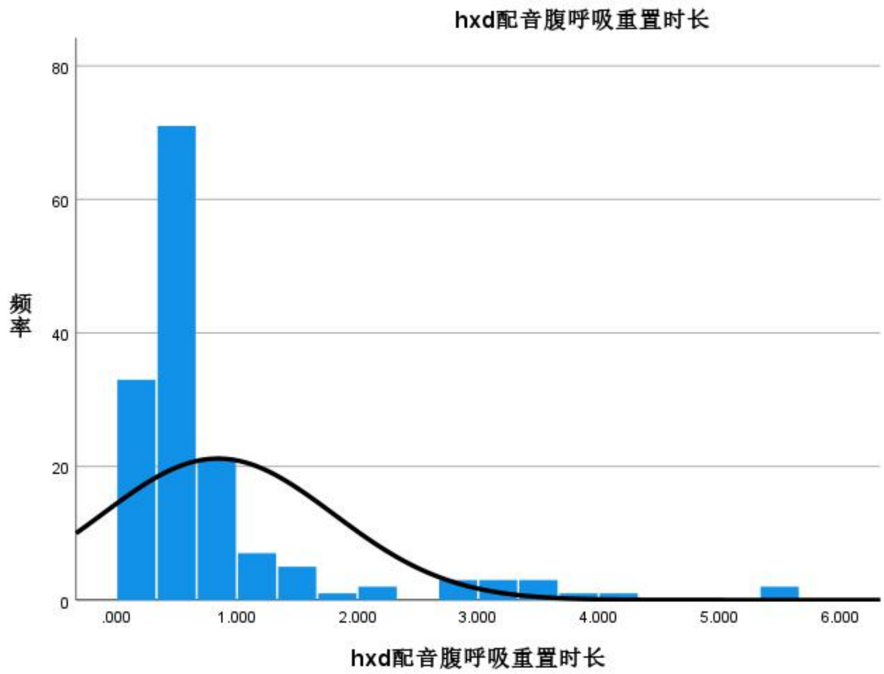


Figure 3.11 Histogram of frequency distribution of dubbing abdominal breathing reset duration

We used SPSS Statistics27 to perform a one-to-one correlation test for the duration of breath resets, reset amplitude, and slope of the inspiratory segment during the dubbing, and obtained the following results.

Table 3.12 Correlation statistics of hxd dubbing reset duration, reset amplitude and slope of inspiratory segment

相关性

			hxd配音重置幅度	hxd配音重置时长
肯德尔 tau_b	hxd配音重置幅度	相关系数	1.000	.239**
		显著性（双尾）	.	.000
		N	149	149
	hxd配音重置时长	相关系数	.239**	1.000
		显著性（双尾）	.000	.
		N	149	149
斯皮尔曼 Rho	hxd配音重置幅度	相关系数	1.000	.334**
		显著性（双尾）	.	.000
		N	149	149
	hxd配音重置时长	相关系数	.334**	1.000
		显著性（双尾）	.000	.
		N	149	149

**. 在 0.01 级别（双尾），相关性显著。

相关性

			hxd配音重置幅度	hxd配音吸气段斜率
肯德尔 tau_b	hxd配音重置幅度	相关系数	1.000	.231**
		显著性（双尾）	.	.000
		N	149	149
	hxd配音吸气段斜率	相关系数	.231**	1.000
		显著性（双尾）	.000	.
		N	149	149
斯皮尔曼 Rho	hxd配音重置幅度	相关系数	1.000	.341**
		显著性（双尾）	.	.000
		N	149	149
	hxd配音吸气段斜率	相关系数	.341**	1.000
		显著性（双尾）	.000	.
		N	149	149

**. 在 0.01 级别（双尾），相关性显著。

相关性

			hxd配音重置时长	hxd配音吸气段斜率
肯德尔 tau_b	hxd配音重置时长	相关系数	1.000	-.380**
		显著性（双尾）	.	.000
		N	149	149
	hxd配音吸气段斜率	相关系数	-.380**	1.000
		显著性（双尾）	.000	.
		N	149	149
斯皮尔曼 Rho	hxd配音重置时长	相关系数	1.000	-.490**
		显著性（双尾）	.	.000
		N	149	149
	hxd配音吸气段斜率	相关系数	-.490**	1.000
		显著性（双尾）	.000	.
		N	149	149

**. 在 0.01 级别（双尾），相关性显著。

After a two-tailed t-test, we found that there was a significant positive correlation between the reset duration and the reset amplitude of breathing in voiceover at the 0.01 level, and the greater the reset duration, the greater the reset amplitude; at the same time, the reset duration and the reset amplitude were also significantly correlated with the slope of the inspiratory segment (at the 0.01 level), i.e., the longer the reset duration, the bigger the reset amplitude and the slope of the inspiratory segment. This finding again supports the high correlation between the reset duration and reset amplitude, indicating that the breath reset duration and amplitude maintain a stable relative relationship regardless of the dubbing genre or the prose, news, reading, and singing genres that have been investigated previously.

3.1.3 Correspondence Pattern between Speech Rhythm and Respiratory Rhythm of Dubbing

The long black dashed line marks the rhyming phrase, the short black dashed line marks the rhyming phrase, the short green solid line marks the syllable, the black wavy dashed line marks the thoracic respiration signal, and the black wavy solid line marks the abdominal respiration signal. We can see that the abdominal respiratory resets are slightly earlier than the thoracic respiratory resets in the dubbing, and the beginning of the exhalation of the abdominal respiratory signal corresponds exactly to the beginning of the phonetic waveform of the rhyming phrase, while the thoracic respiratory signal curve shows several respiratory resets of smaller amplitude, and some of the thoracic respiratory resets do not occur synchronously with the abdominal respiratory resets or correspond exactly to the rhythmic boundaries. We counted the correspondence patterns between the 1443 syllables and 130 rhyming phrases and the thoracic respiratory signal resets in all the dubbed texts, and found that 42% of the thoracic respiratory resets in non-rhyming sentences did not correspond to the boundaries of rhyming phrases, and among these 42% of irregular respiratory rhyming units and irregular correspondence of phonological rhyming units, as many as 78% of the thoracic respiratory resets occurred in contexts where the immediately following syllables were falling tone (the 4th tone in Mandarin)[①]. Therefore, we speculate that the tone of the Chinese may have an effect on the thoracic respiratory resetting during dubbing. In general, the thoracic and ventral respiratory signals do not always change in the same direction; in general, the thoracic and ventral respiratory signals are both decreasing or increasing at the same time, but the starting point of the changes is not the same, which provides room for the asynchronous rise and fall of the thoracic and ventral respiratory signals.

① Secondly, a weak thoracic breathing reset also tended to occur when the tone of the following syllable was the 1st tone of Mandarin, a feature that occurred only in dubbing and voice-over reading, and mainly in the former, but not in regular pronouncers. We also found similar problems in the dubbing of female dubbers. We consider that this may be related to the size of the inspiratory volume that needs to be mobilized when pronouncing the Chinese tones, but this hypothesis has yet to be tested in a larger scale test of the breathing characteristics of the Chinese tones; at the same time, we cannot exclude the influence of individual pronouncers on the experimental results due to the insufficient sample size.

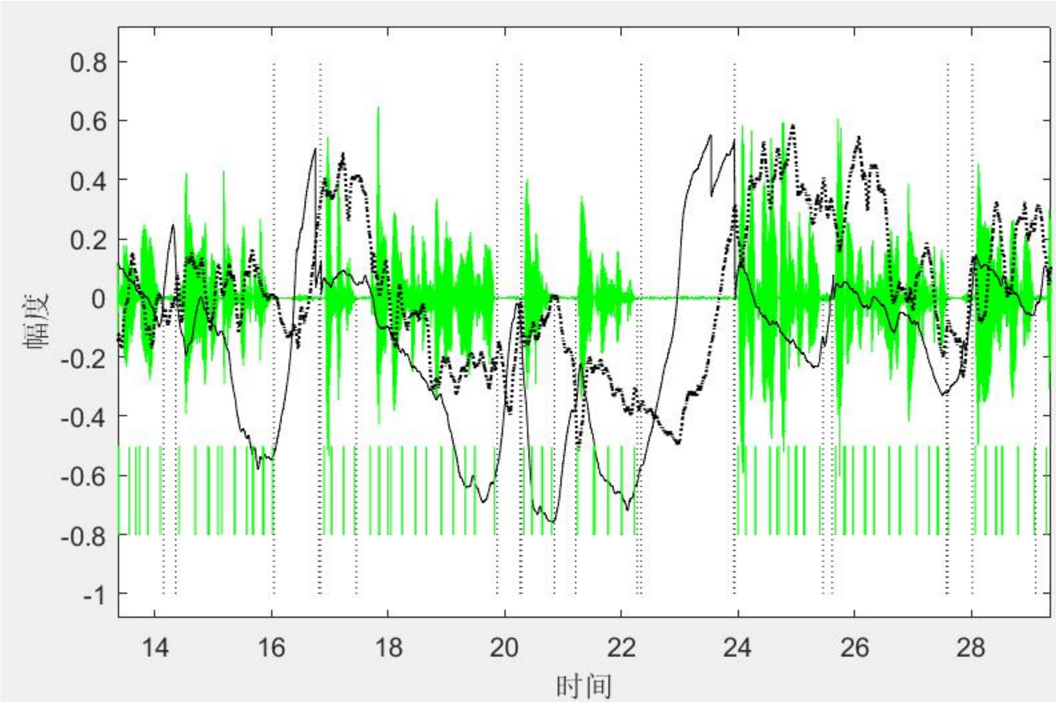


Figure 3.13 Superimposed schematic diagram of dubbing speech waveform and breathing signal

3.1.4 Respiratory Rhyme Analysis of the Speech types of Dubbing

Mingjun Wang and Xueliang Yan (2014) divided the speech styles of documentary narration into five types: argumentative, expository, lyrical, vernacular, and declarative. The argumentative narration has a more solemn and serious voice, without falsetto, with strong pronunciation and rounded and full pronunciation, mostly seen in the production of large-scale political or historical documentaries; the expository narration is not as solemn as the expository narration because its purpose is to expound some non-common knowledge to the audience, with a calm and firm rhythm; the lyrical narration is mostly descriptive, with obvious lyricism, often with some rhetorical style to help express, "round and soft, full of emotion and sincere"[①]; vernacular narration, that is, daily spoken, easy to understand, and unrefined voice, this mode of dubbing is mostly seen in biographical documentaries, in

① 王明军，阎亮. 影视配音实用教程[M]. 中国传媒大学出版社, 2014：4-14.

nature documentaries It is usually found in the description of the psychological activities or scenes of wilderness explorers or other staff members filmed in the documentary, but it does not appear in the segment selected in this paper, so we do not consider it as one of the objects to be examined; finally, it is a declarative mode, mostly used in cultural special programs, with a plain voice but not thin and dull, and the rhythm of the statements is not too big.

We used SPSS Statistics27 to statistically analyze the different language styles of the above dubbed expressions from the perspective of breathing reset level and speech rhythm, and found that in this excerpt of the nature class documentary, the frequency distribution of the combined reset duration and reset amplitude, regardless of the speech style, can be divided into the two levels of thoracic breathing reset and two levels of abdominal breathing reset, and the abdominal breathing reset point is slightly earlier than the thoracic breathing reset point. When the speech style changed, the breathing pattern did not change, but continued with the breathing pattern of the previous speech, and the pause interval between syllables, rhyming phrases and rhyming sentences did not change significantly due to the change of speech style. Even when the its shift occurred after a fairly long period of silent segments, the voiceover's breathing signal was only expressed as a single larger breath reset. In the documentary production process, the director will first do the dubbing according to the content of the script after finishing the editing and assembling of the footage, placing the specific position of the particular segment, and then hiring a professional artist to do the dubbing, which means that the priority of the dubbing artist's differentiation between silent and voiced segments should be much higher than the differentiation of the speech style within the voiced segments according to the difference of emotional expression. Therefore, there is no significant difference in the breathing patterns or phonetic rhythms of the voice-over artists, but this does not mean that there is no significant difference in other phonetic characteristics, and we still need to further combine multidimensional analysis from the perspective of emotional speech to make a more detailed examination and judgment.

In addition, the division of the language style of voiceover narration in media textbooks still has its application value. In the process of specific dubbing, the distinction between the five categories of argumentative, vernacular, expository, lyrical and declarative relies first on the overall grasp of the documentary's subject matter and text type, so as to set the emotional tone of the dubbing; then on the control of the specific details within the text, and also on the need to pinpoint the size of the ups and downs of the emotional transitions when the language style changes, in order to avoid too great a gap between the front and back, which creates a sense of disconnection between the viewer and the content of the picture. Although the analysis of the breath rhythm characteristics based on the speech style shows no significant difference between them, it also suggests that the analysis of emotional speech should be further refined, and we may even rule out the possibility that the results of each dimension in the multidimensional acoustic analysis need to be integrated and their respective weights quantified in order to better locate and simulate a specific phonation paradigm.

3.2 Phonetic Rhythm and Breathing Rhythm Characteristics of Dubber’s Reading

We also collected four signals from the same voice-over artist reading the same text, and compared them with the voice rhythm and breathing characteristics of the dubbing to identify the differences and better understand the acoustic and breathing characteristics of the two different speech styles.

3.2.1 Speech Rhythm of Dubbers' Reading

The syllable durations and syllable interval durations of the dubbers' reading are normally distributed as in the dubbed speech, as shown in the figure. The average syllable duration of 0.148s is shorter than the average syllable duration of dubbing; the standard deviation is 0.45, mainly between 0.08s and 0.21s; the average syllable interval duration is 0.054s, both parameters are much lower than the average syllable duration and syllable interval duration of dubbing, which indicates that the average speech speed of reading is faster than that of dubbing, and dubbing requires more emotion to achieve a more compelling expression. The dubbers also pay attention to the progressiveness of the beginning and end according to the presentation and transformation of the content of the picture, while these points do not require much attention in the reading style.

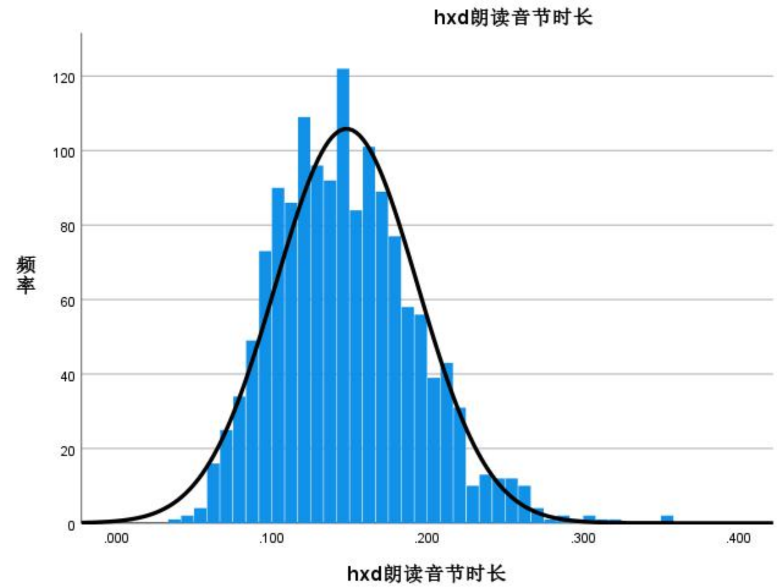


Figure 3.14 Histogram of the frequency distribution of syllable duration in dubber's reading

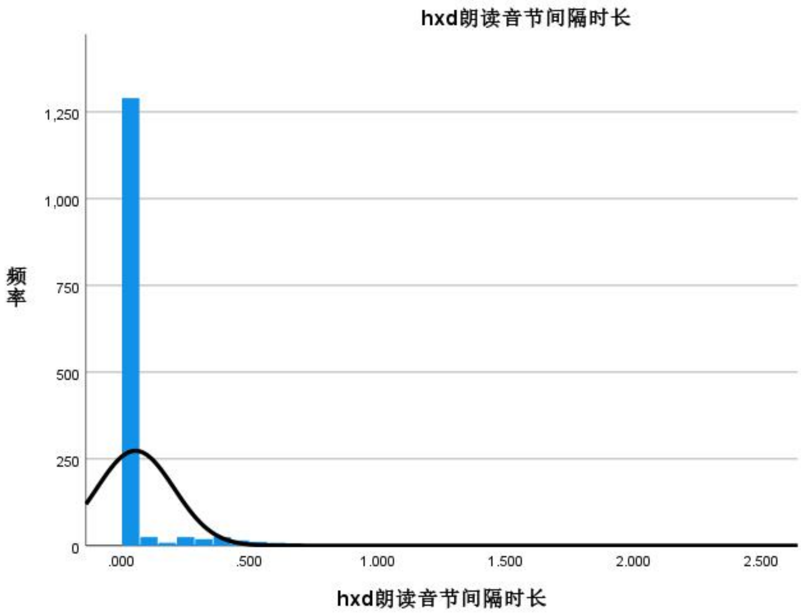


Figure 3.15 Histogram of the frequency distribution of the duration of syllable intervals in dubber's reading

The rhyming phrase duration and rhyming phrase interval duration during reading aloud still conformed to the normal distribution, as shown in the figure below. The average length of rhyming phrases during reading is 1.1s, which is slightly lower than the average length of rhyming phrases during dubbing. However, the number of rhyming phrases divided in reading aloud is much smaller than that in dubbing, with only 45 rhyming phrases. The small number of rhyming phrases also leads to a very long mean interval of 4.347, because there may be a long interval between adjacent rhyming phrases in which only rhyming phrases containing a rhyming phrase alternate (as mentioned above, we classified this part of the rhyming unit as a rhyming phrase).

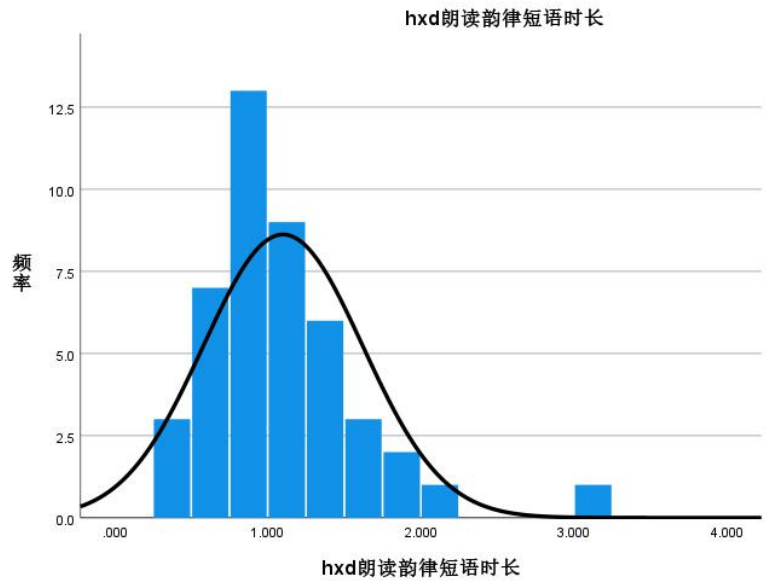


Figure 3.16 Histogram of frequency distribution of rhyming phrase duration in dubber’s reading

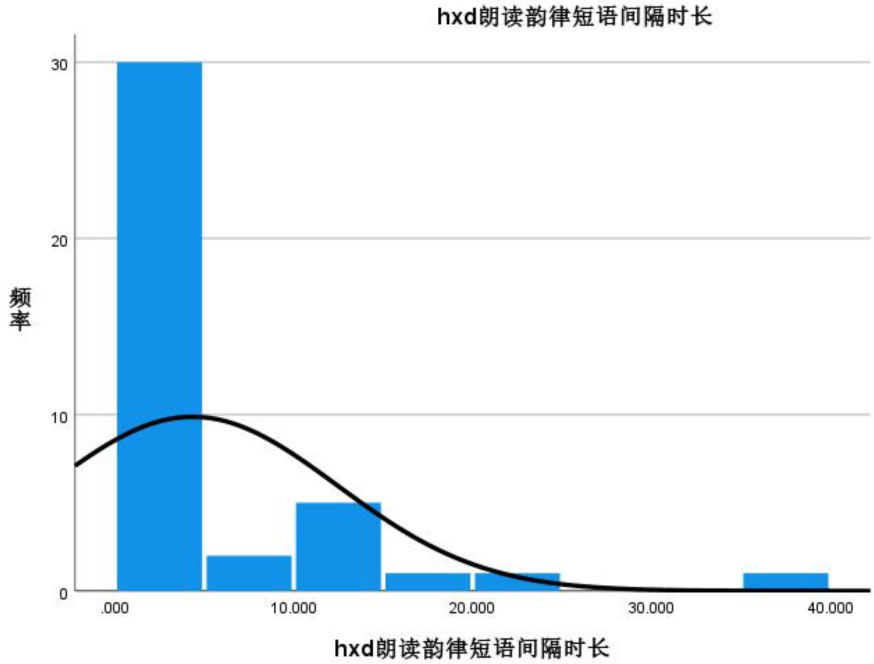


Figure 3.17 Histogram of frequency distribution of the interval duration of rhyming phrases in dubber's reading

The average duration of rhyming sentences in reading aloud is 1.751s, with a standard deviation of 0.769. This statistic is much lower than the average duration of rhyming sentences in dubbing, which is 2.422s; rhyming sentences are mainly concentrated in the range of 1s-2.4s, and the average pause between sentences is 0.389s, which is about 50% lower than the average pause in dubbing, which is also a chain reaction of the average speech speed in reading aloud being faster than that in dubbing. This is also a knock-on effect of the average speed of speech in the reading than in the voiceover.

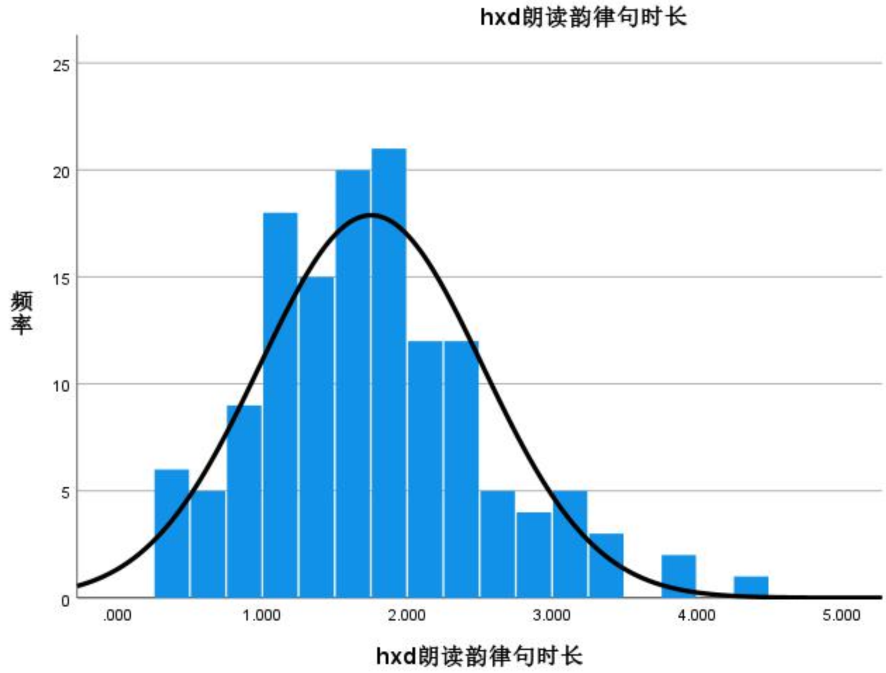


Figure 3.18 Histogram of frequency distribution of rhyming sentence duration in dubber's reading

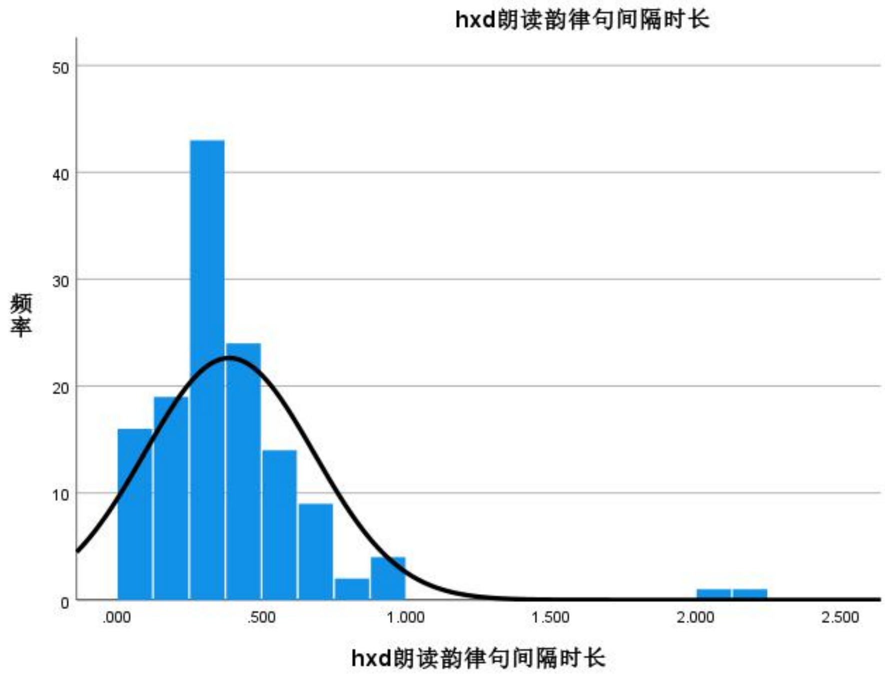


Figure 3.19 Histogram of frequency distribution of the interval duration of rhyming sentences in dubber's reading

3.2.2 Respiratory Rhythm of the Dubber's Reading

As shown in the histogram of the frequency distribution of thoracic and abdominal respiratory resetting amplitude of the read-aloud corpora (138 thoracic respirations and 130 abdominal respirations, both of which are less than the number of respiratory resets in the voiceover, but the relative relationship between the number of thoracic and abdominal respirations is the same as that in the voiceover, both of which show more thoracic than abdominal respiratory resets), the respiratory resetting amplitude in the read-aloud also basically conforms to the normal distribution, with a peak at A large peak was found at 0.8, while more stable respiratory resets were found at 0.6~1.0 and 1.1~1.3, which indicated that the thoracic respiratory resets could be roughly divided into two levels. We also note that the number of thoracic breathing resets in the interval 0.1-0.4 is much lower than the number of thoracic breathing resets during voiceover in the same interval, which is due to the fact that a significant portion of the smaller thoracic breathing resets during voiceover correspond to rhyming phrases, but the number of rhyming phrases is greatly reduced when the dubber is reading, and this change synchronously resulted in fewer small-amplitude thoracic breath resets observed during the reading.

The abdominal respiratory resets were also normally distributed, and the average abdominal respiratory reset was 0.781, which was not significantly different from the average thoracic respiratory reset. This is different from the relative relationship between the average resetting amplitude of thoracic and abdominal breathing during dubbing. There were two peaks of abdominal breathing around 0.3 and 0.8, and there were relatively stable segments in the intervals of 0.1-0.5, 0.6-0.8, and 0.9-1.5, which indicated that it was more appropriate to divide the abdominal breathing of the voiceover artist during reading aloud into three levels of breathing resets.

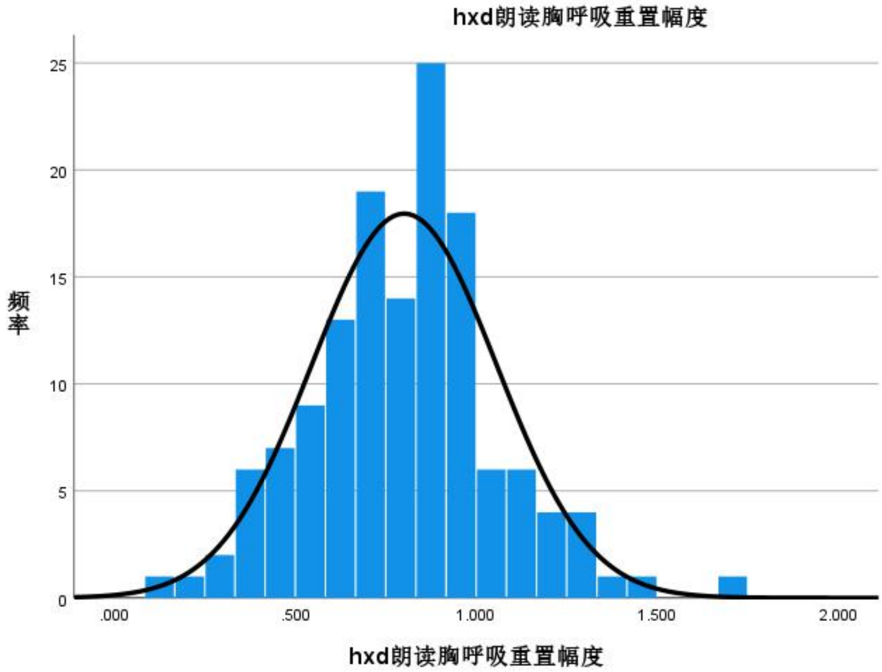


Figure 3.20 Histogram of frequency distribution of thoracic breathing resetting amplitude in dubber's reading

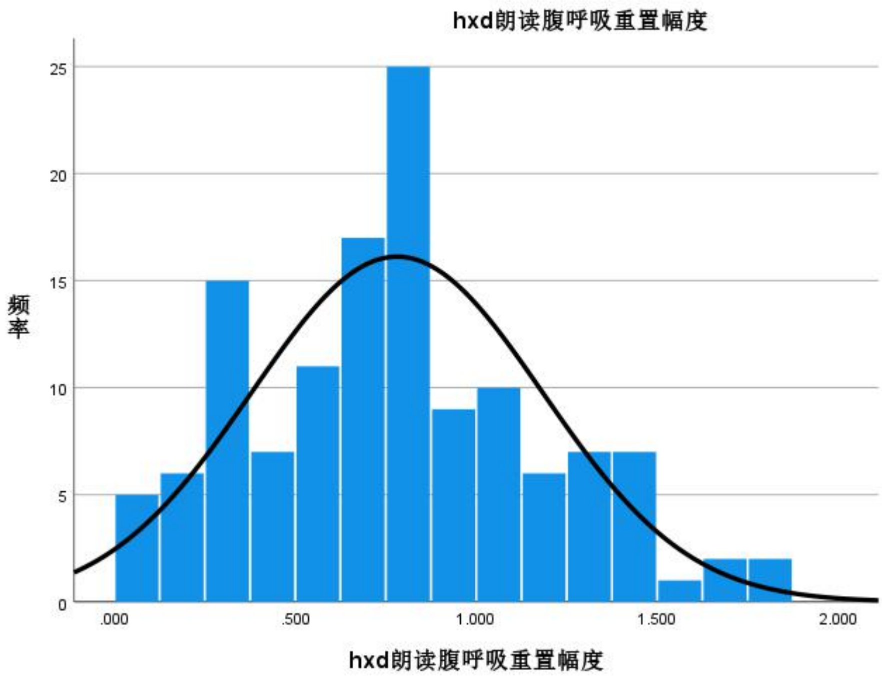


Figure 3.21 Histogram of frequency distribution of belly breathing resetting amplitude in dubber's reading

Combined with the duration of the thoracic and abdominal respiratory resets, the thoracic respiratory resets were stable and high in the 0.3s-1.5s range, and there were also a considerable number of prolonged respiratory resets in the 1.6s-2.2s range, which confirmed that it was appropriate to classify the thoracic respiratory resets as level 2 during the reading. At the same time, the frequency distribution of abdominal respiratory resets showed that there was a peak at 0.4s and then remained at a relatively high frequency range until 1.5s; then the frequency distribution of abdominal respiratory resets decreased linearly. However, the frequency of the resetting amplitude of the abdominal breaths, small breath resetting occupies a large part of the frequency chart, and this frequency distribution pattern of amplitude and duration reflects the characteristics of small and medium breaths resetting more and large breaths resetting less during the dubber's reading.

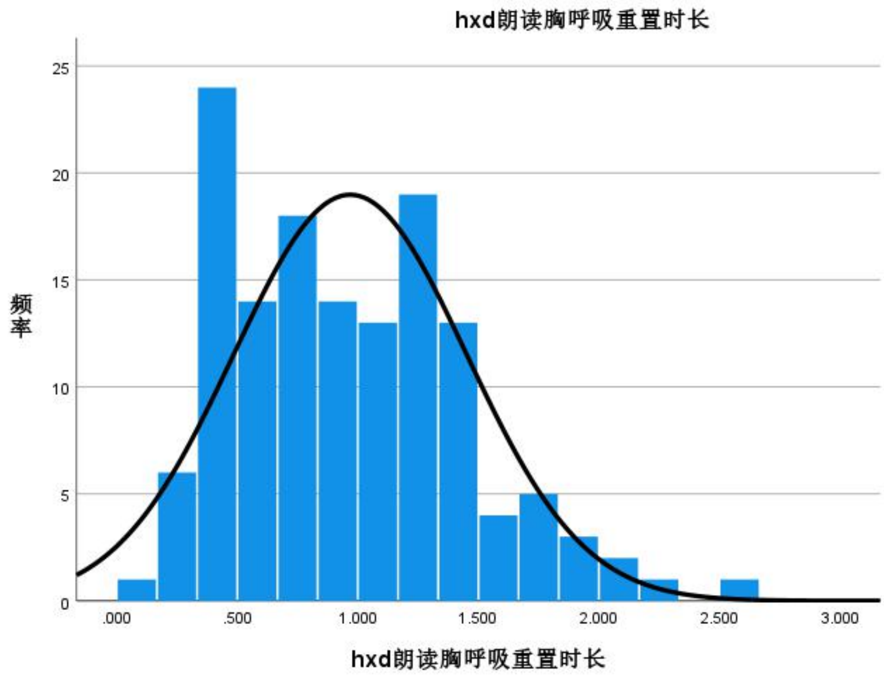


Figure 3.22 Histogram of the frequency distribution of the duration of thoracic breath resetting in dubber's reading

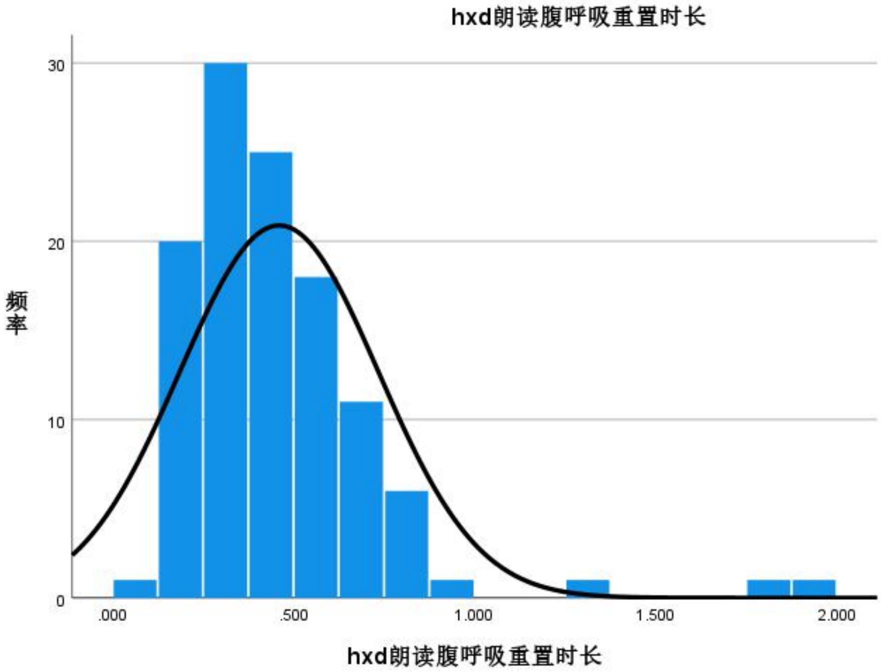


Figure 3.23 Histogram of frequency distribution of belly-breath resetting duration in dubber's reading

We likewise conducted a correlation analysis of the breath-resetting parameters when read aloud by professional voiceover artists, with the following results.

Table 3.24 Correlation statistics of *hxd* read aloud reset duration, reset amplitude and slope of inspiratory segment

相关性

			hxd朗读重置幅度	hxd朗读重置时长
肯德尔 tau_b	hxd朗读重置幅度	相关系数	1.000	.446**
		显著性（双尾）	.	.000
		N	101	101
	hxd朗读重置时长	相关系数	.446**	1.000
		显著性（双尾）	.000	.
		N	101	101
斯皮尔曼 Rho	hxd朗读重置幅度	相关系数	1.000	.617**
		显著性（双尾）	.	.000
		N	101	101
	hxd朗读重置时长	相关系数	.617**	1.000
		显著性（双尾）	.000	.
		N	101	101

**. 在 0.01 级别（双尾），相关性显著。

相关性

			hxd朗读重置幅度	hxd朗读吸气段斜率
肯德尔 tau_b	hxd朗读重置幅度	相关系数	1.000	.433**
		显著性（双尾）	.	.000
		N	101	101
	hxd朗读吸气段斜率	相关系数	.433**	1.000
		显著性（双尾）	.000	.
		N	101	101
斯皮尔曼 Rho	hxd朗读重置幅度	相关系数	1.000	.600**
		显著性（双尾）	.	.000
		N	101	101
	hxd朗读吸气段斜率	相关系数	.600**	1.000
		显著性（双尾）	.000	.
		N	101	101

**. 在 0.01 级别（双尾），相关性显著。

相关性

			hxd朗读吸气段斜率	hxd朗读重置时长
肯德尔 tau_b	hxd朗读吸气段斜率	相关系数	1.000	-.107
		显著性（双尾）	.	.113
		N	101	101
	hxd朗读重置时长	相关系数	-.107	1.000
		显著性（双尾）	.113	.
		N	101	101
斯皮尔曼 Rho	hxd朗读吸气段斜率	相关系数	1.000	-.147
		显著性（双尾）	.	.141
		N	101	101
	hxd朗读重置时长	相关系数	-.147	1.000
		显著性（双尾）	.141	.
		N	101	101

We found that the respiratory reset duration and amplitude during voiceover reading still showed a significant correlation, but there was no correlation between reset duration and inspiratory segment slope, and there was instead a significant correlation between reset amplitude and inspiratory segment slope. The asymmetry of this correlation relationship not only reaffirms the close relationship between reset duration and amplitude, but also implies that the relationship between reset duration and inspiratory segment slope may not be stable, and an increase in reset duration does not necessarily lead to an increase in inspiratory slope.

3.2.3 Correspondence Pattern between Speech Rhythm and Respiratory Rhythm of Dubber’s Reading

Combining the speech waveforms and the thoracic and abdominal respiratory signals, the abdominal respiratory resetting during reading aloud was slightly earlier than the corresponding thoracic respiratory resetting, as was the abdominal respiratory resetting during voiceover. Similarly, the onset of exhalation of the abdominal respiratory signal also corresponds exactly to the onset of the phonological waveform of the rhyming phrase, while the thoracic respiratory signal curve shows a number of smaller respiratory resets, and some respiratory resets are not synchronized with the abdominal respiratory resets or correspond to the rhythmic boundaries. However, unlike the voiceover performance, the abdominal and thoracic respiration patterns were basically opposite to each other, with the abdominal respiration tending to remain in the exhaled state when the thoracic respiration was inhaled and vice versa. In contrast, the voice-over artist's thoracic and abdominal breathing often appear to be in the same frequency. This reflects that the voiceover artist is using a different way of thoracic and abdominal breathing grouping when he or she is doing voiceover and when he or she is reading aloud. However, regardless of the breathing pattern, the pause in speech is inevitably accompanied by the resetting of the thoracic or abdominal breathing signal; while the resetting of the thoracic or abdominal breathing signal is not always observed at the same time as the pause in speech. This rule also applies to the dubbing style.

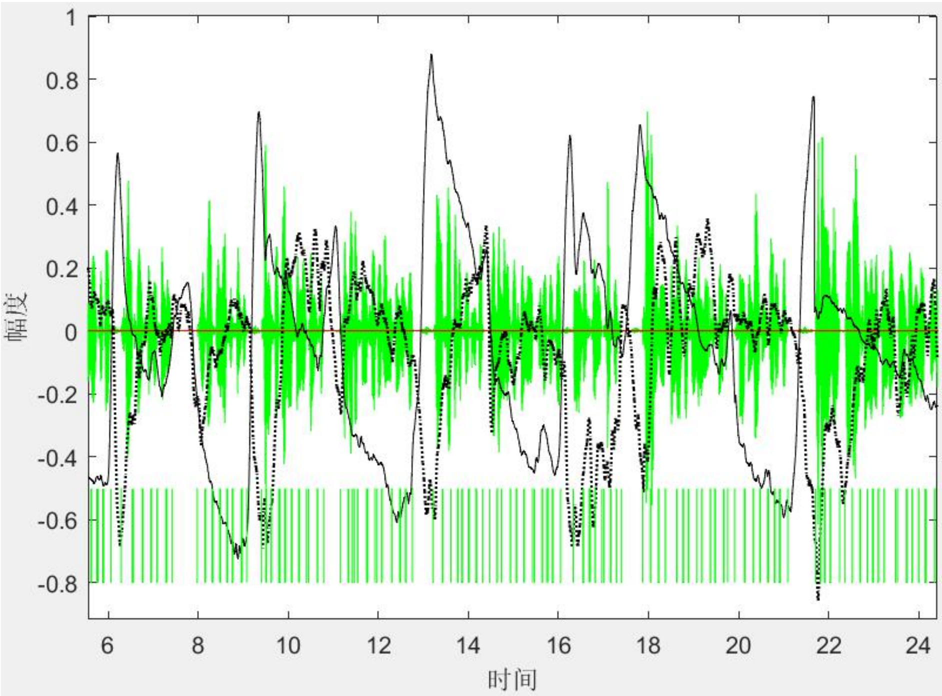


Figure 3.25 Superposition of voice waveform and thoracoabdominal respiratory signal in dubber's reading

3.3 Speech Rhythm and Respiratory Rhythm Characteristics of Common Pronouncer's Reading

From the above analysis, it can be seen that even the same pronouncer may adopt different breathing and vocalization styles in the process of dubbing and reading. Next, we will investigate the speech and respiratory rhythms of the common pronouncers and observe the similarities and differences between the pronunciation of the common pronouncers, the dubbing and the dubber's reading.

3.3.1 Speech Rhythm of Common Reading

The average duration of syllables in the pronouncers' speech is 0.187s, mainly in the range of 0.1s to 0.25s. The average syllable duration is similar to that of the dubbed speech, but slightly longer than that of the professional dubbers. We believe that the reason for this situation is not only the individual differences of the pronouncers, but also the fact that the dubbers recorded both the dubbed and the read speech consecutively during the data collection, so that the dubbers themselves kept the maximum difference in the perceptual range for the pronunciation of these two styles. The dubbers themselves, in order to maintain the greatest degree of difference in the perceptual range of pronunciation between the two genres, may therefore adopt a vocalization and breathing style that deviates more from the dubbing than the daily reading in the read-aloud genre performed after the dubbing. This treatment will affect the subsequent changes of a series of rhythmic parameters. Therefore, this also reminds us that when collecting the corpus in the field, we should not only ensure the normal use of software and hardware, but also pay attention to the possible psychological cues of the pronouncers, and try to avoid the influence of the pronouncers' psychological cues on the experimental results by means of the experimental process settings adjustment. The average length of the syllable interval of the common pronouncer was 0.07s, which was higher than the average length of the syllable interval of the voice-over speaker but lower than the average length of the syllable interval of the voice-over speaker. This is also due to the fact that the dubbing needs to be presented slowly with the picture content. This is supported by the comparison of the data between the two reading and dubbing genres.

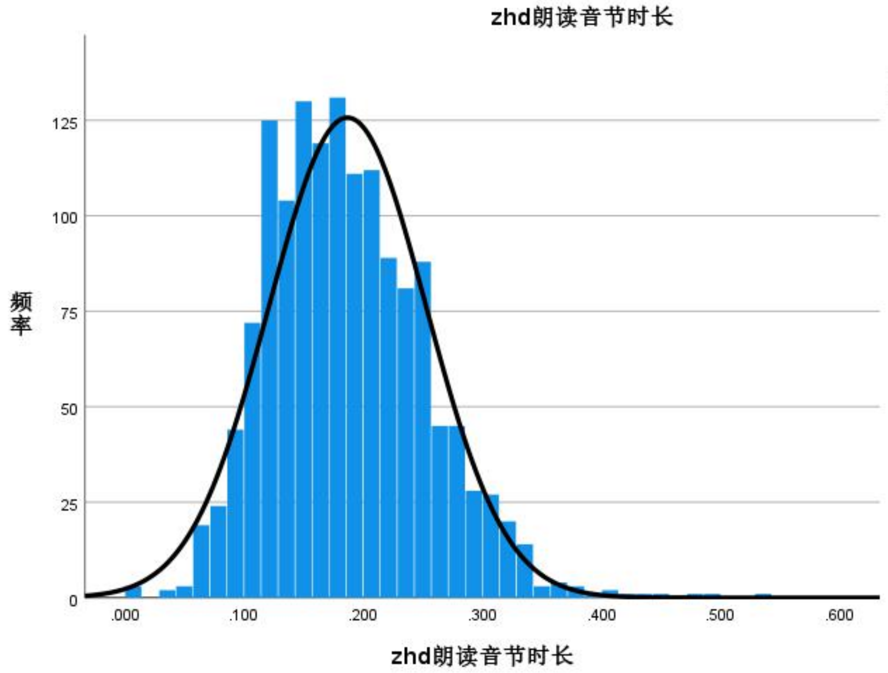


Figure 3.26 Histogram of frequency distribution of syllable duration in common reading

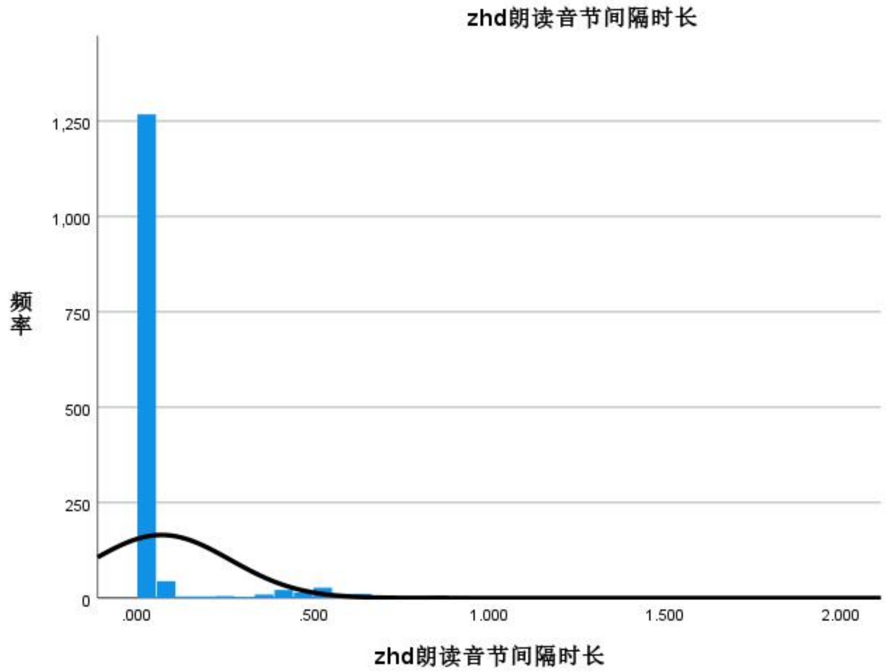


Figure 3.27 Histogram of frequency distribution of the syllables' interval duration in common reading

The average length of rhyming phrases is 2.638s, which is much higher than the average length of rhyming phrases in dubber's dubbing and reading. This is due to the fact that the average pronouncer reads aloud with a constant rhythm, without adding any emotion, and with a superficial understanding of the text, thus being almost unaffected by the semantic and grammatical units, and just flatly converting the text into speech and then emitting it.

In dealing with the rhythmic unit of rhyming sentences, common pronouncers also show characteristics that are different from those of dubbers. The average length of rhyming sentences for common speakers was 2.219s, which was lower than that of dubbing but higher than that of dubbing, but it was mainly stable in the range of 1s to 2.8s for all three speech forms. The average length of rhyming sentence interval is 0.492s, which is also lower than the average length of rhyming sentence interval in dubbing but higher than the average length of rhyming sentence interval in dubber's reading.

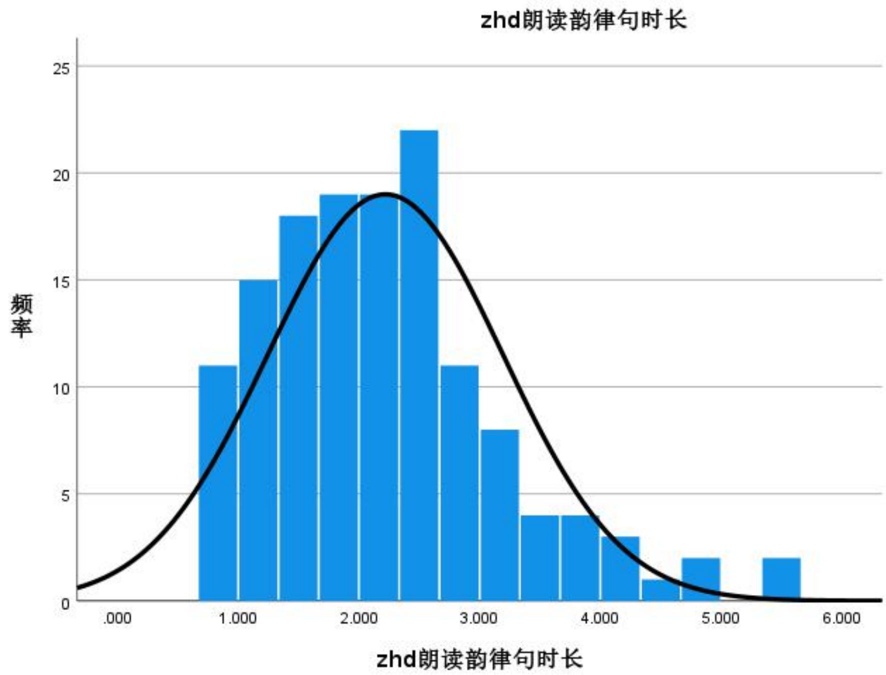


Figure 3.28 Histogram of frequency distribution of the rhyming sentences duration in common reading

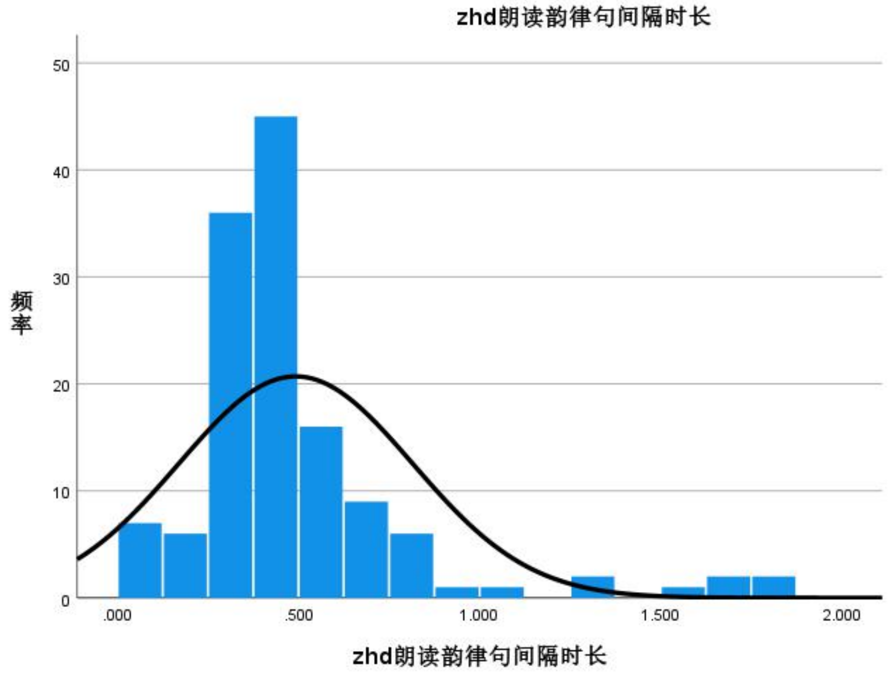


Figure 3.29 Histogram of the frequency distribution of the rhyming sentences' interval duration in common reading

3.3.2 Respiratory Rhythm of Common Reading

The histogram of the frequency distribution of thoracic and abdominal respiratory resetting amplitude of the common pronouncers reading aloud shows (133 thoracic respirations and 133 abdominal respirations, which are less than the number of respiratory resetting amplitude of the voiceover, and not much different from the number of respiratory resetting amplitude of the voiceover). The average thoracic respiratory resetting amplitude was 0.674, which was larger than the average thoracic respiratory resetting amplitude in dubbing but lower than the average thoracic respiratory resetting amplitude in dubbing, and the average abdominal respiratory resetting amplitude was 0.737, which was almost equal to the average abdominal respiratory resetting amplitude in the other two languages. The frequency distribution of thoracic and abdominal respiratory resetting amplitude was basically the same, except for a peak amplitude around 0.6, and a relatively stable interval between 0.8 and 1.0. Therefore, we classified the thoracic and abdominal respiratory signals of the common pronouncers as secondary respiratory resetting.

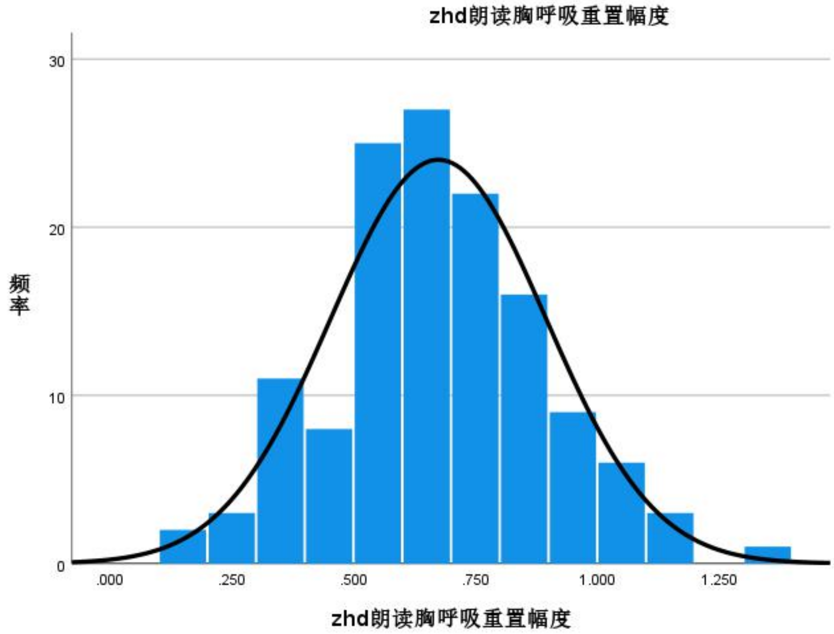


Figure 3.30 Histogram of the frequency distribution of the thoracic respiration resetting amplitude in common reading

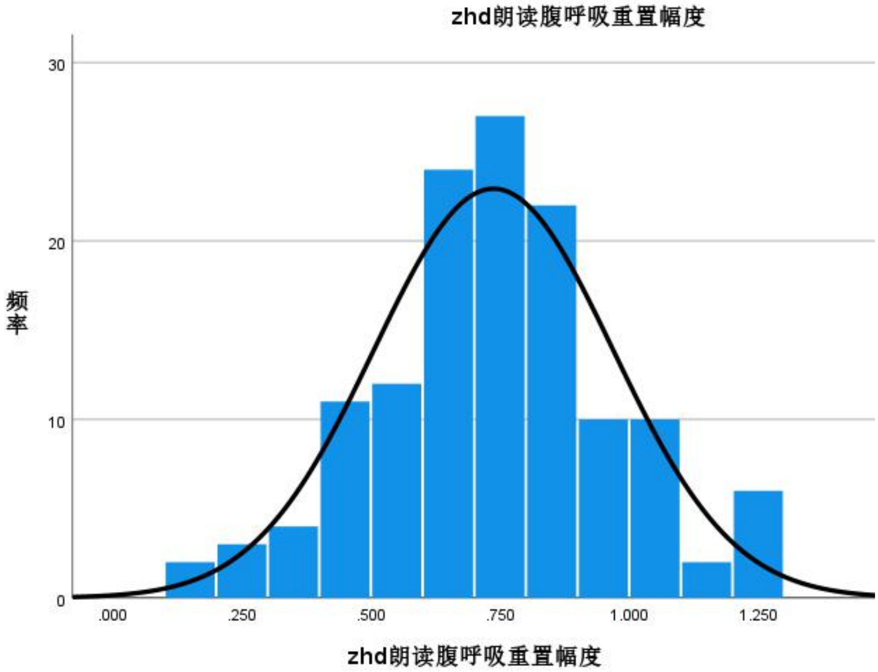


Figure 3.31 Histogram of the frequency distribution of abdominal respiration resetting amplitude in common reading

The number of thoracic and abdominal respiratory resets of the common articulators is completely equal, and the butterfly-shaped distribution of thoracic and abdominal respiratory resets is clearly seen in the speech waveform and the schematic diagram of thoracic and abdominal respiratory signals, which is completely different from the dubbers' dubbing and reading. At the same time, the thoracic and abdominal respiratory signals of the common pronouncers are very smooth, especially the thoracic respiratory signals, and there is almost no respiratory resetting at the rhyming phrases based on semantic division compared with the former two.

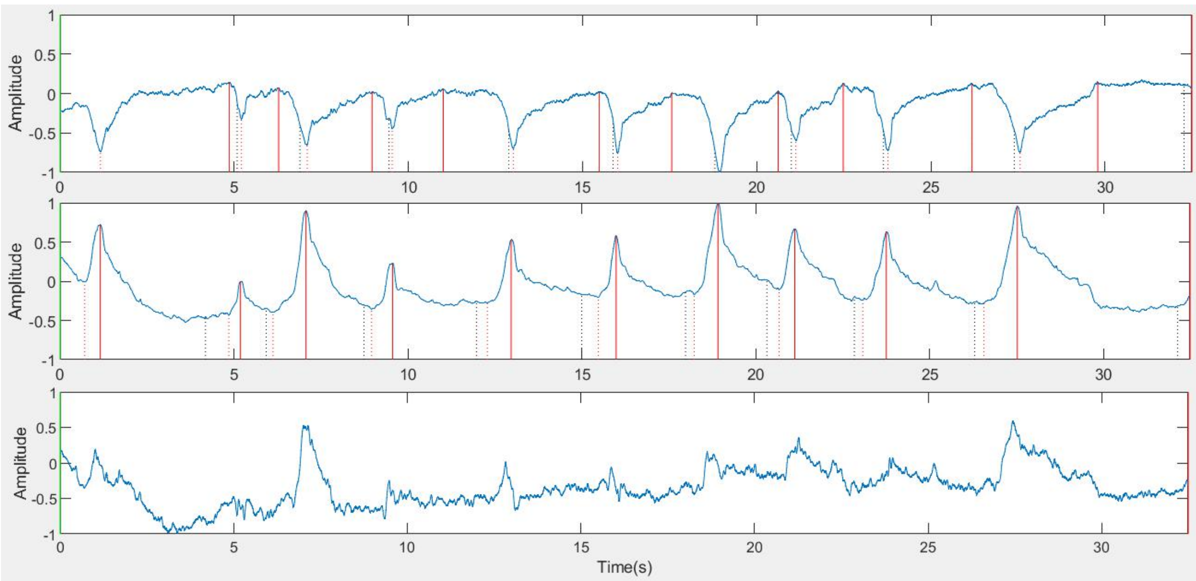


Figure 3.32 Schematic diagram of thoracic and abdominal breathing signals in common reading

In terms of the resetting duration of thoracic and abdominal breathing, there is a relatively stable interval between 0.7s and 2.5s for the thoracic breathing signal of ordinary pronouncers reading aloud, and the existence of this stable interval indicates that the thoracic breathing resetting time is longer in ordinary reading aloud; at the same time. There is a peak length of abdominal breathing around 0.7s, and this peak is also the high incidence of abdominal breathing resetting in dubbing and dubbers' reading aloud, which is essentially equivalent to a secondary abdominal respiratory reset. This is due to the fact that most of the rest of the passage, except for the primary respiratory resets at the beginning of the passage, are secondary respiratory resets, and especially the secondary respiratory resets are the most frequent, thus showing a higher stable segment around 0.7 seconds in the frequency distribution of the resets. This time also indicates that the human body may prefer to use this range of abdominal breath resets, which is more in line with our normal breathing rhythm, ensuring oxygen intake without affecting articulation too much, and also taking into account the coherence between sentences.

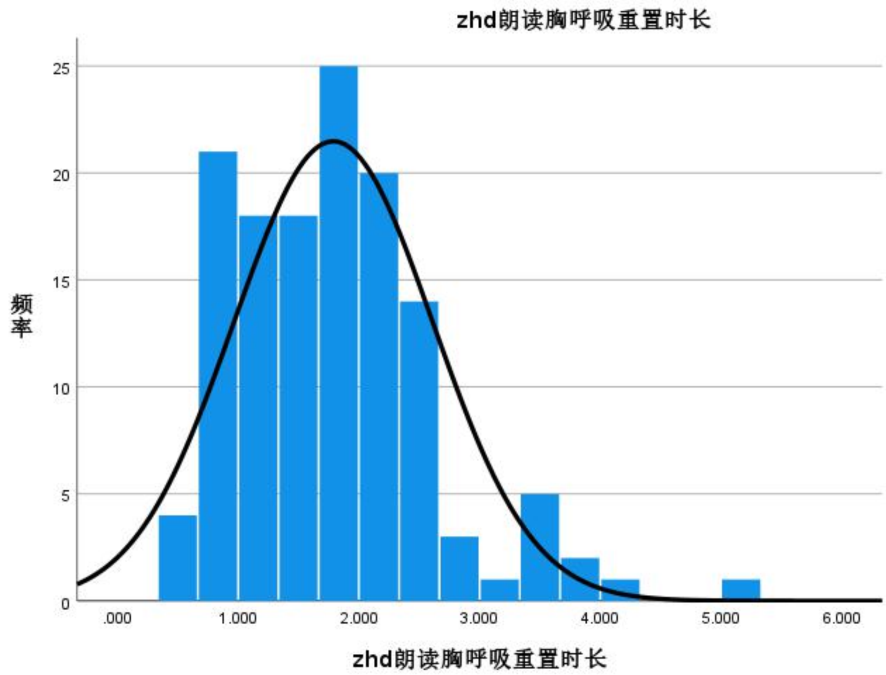


Figure 3.33 Histogram of the frequency distribution of the thoracic respiratory reset duration in common reading

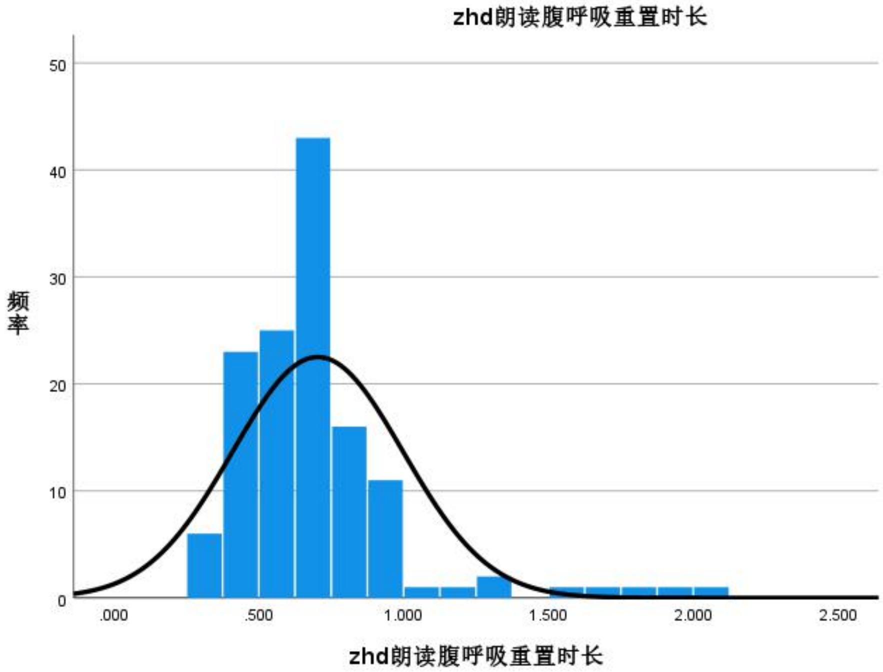


Figure 3.34 Histogram of frequency distribution of the abdominal respiratory reset duration in common reading

We performed a statistical analysis of the respiratory resetting characteristics of common speaker with the following results.

Table 3.35 Correlation statistics of *zhd* respiratory reset duration, amplitude and slope of inspiratory segment

相关性

			zhd朗读重置幅度	zhd朗读重置时长
肯德尔 tau_b	zhd朗读重置幅度	相关系数	1.000	.514**
		显著性（双尾）	.	.000
		N	133	133
	zhd朗读重置时长	相关系数	.514**	1.000
		显著性（双尾）	.000	.
		N	133	133
斯皮尔曼 Rho	zhd朗读重置幅度	相关系数	1.000	.696**
		显著性（双尾）	.	.000
		N	133	133
	zhd朗读重置时长	相关系数	.696**	1.000
		显著性（双尾）	.000	.
		N	133	133

**. 在 0.01 级别（双尾），相关性显著。

相关性

			zhd朗读重置时长	zhd朗读吸气段斜率
肯德尔 tau_b	zhd朗读重置时长	相关系数	1.000	-.390**
		显著性（双尾）	.	.000
		N	133	133
	zhd朗读吸气段斜率	相关系数	-.390**	1.000
		显著性（双尾）	.000	.
		N	133	133
斯皮尔曼 Rho	zhd朗读重置时长	相关系数	1.000	-.542**
		显著性（双尾）	.	.000
		N	133	133
	zhd朗读吸气段斜率	相关系数	-.542**	1.000
		显著性（双尾）	.000	.
		N	133	133

**. 在 0.01 级别（双尾），相关性显著。

相关性

			zhd朗读吸气段斜率	zhd朗读重置幅度
肯德尔 tau_b	zhd朗读吸气段斜率	相关系数	1.000	.062
		显著性（双尾）	.	.292
		N	133	133
	zhd朗读重置幅度	相关系数	.062	1.000
		显著性（双尾）	.292	.
		N	133	133
斯皮尔曼 Rho	zhd朗读吸气段斜率	相关系数	1.000	.090
		显著性（双尾）	.	.303
		N	133	133
	zhd朗读重置幅度	相关系数	.090	1.000
		显著性（双尾）	.303	.
		N	133	133

Compared with the dubbing and reading by professional voice-over artists, the breathing resetting length and resetting amplitude of ordinary speakers also showed a high positive correlation (0.01 level), but their resetting amplitude did not correlate with the inspiratory slope. However, the statistical results of the voice-over corpus and the sample size still need to be expanded, so we can only conservatively assume that the slope of the inspiratory segment during breathing may not have a significant relationship with the length and amplitude of breath resets.

3.3.3 Correspondence between Speech Rhythm and Respiratory Rhythm of Common Reading

Combining the phonetic waveforms and the thoracic and abdominal respiratory signals, the abdominal respiratory resetting and abdominal respiratory resetting of the common pronouncers can be neatly corresponded to each other, and the relative relationship between the thoracic and abdominal respiratory resetting time points is similar to that of the dubbing and dubber's reading, where the abdominal respiratory resetting is slightly earlier than the corresponding thoracic respiratory resetting. Similarly, the onset of exhalation of the abdominal respiratory signal in the normal reading also corresponds exactly to the onset of the phonetic waveform in the rhyming phrase. However, the respiratory signal curve is more stable in the common reading, without too many small respiratory resets, which is especially evident in the thoracic respiratory signal. Both the abdominal and the thoracic respiration patterns were basically relative to each other, with the thoracic respiration curve rising and the abdominal respiration curve falling, and the vice versa, while the thoracic and thoracic respiration curves of the voiceover artists were often in the same direction. This reflects the fact that there are indeed differences in the respiratory grouping between dubbing and reading. However, the regularity of the resetting of the respiratory signal in common reading corresponds to the beginning and end of rhyming sentences, so this also leads to a primary or secondary resetting of both the thoracic and abdominal respiratory signals at the pauses in speech, i.e., where there is a pause there is necessarily a respiratory reset, and where there is a respiratory reset there is necessarily a pause in speech. The finding of respiratory resetting in common speakers complements previous studies that have considered the relationship between phonological pauses and respiratory resetting, in other words, there is indeed a co-occurrence of the two in specific speech expressions[①].

3.4 Summary

Through the above comparison of the phonetic and respiratory characteristics of the dubbing, dubber’s reading, and common reading, we found that the abdominal respiratory resetting was always slightly earlier than the thoracic respiratory resetting in all three cases, and the beginning of the exhalation of the abdominal respiratory signal could correspond exactly to the beginning of the phonetic waveform of the rhyming phrase, while the thoracic respiratory signal curves of the dubbing and dubber’s reading showed several respiratory resetting of smaller magnitude, and some of the respiratory resets were not synchronized with the abdominal respiratory resets or corresponded exactly to the rhyme boundaries. The tone of the Chinese may have an effect on the thoracic respiratory resetting during dubbing. The thoracic and abdominal respiratory signals do not always move in the same direction, and often the thoracic and abdominal respiratory signals both remain in a downward or

① It is also possible that this feature is not only related to the "speech expression". In fact, although many scholars have analyzed "reading" as a reference quantity in previous studies on breath resetting, few of them have explicitly stated whether the "reading" is emotionally charged or not. Even if pronouncers read, it is very difficult to measure the amount of emotion of different pronouncers in the field. Since the common pronouncers’ reading, the dubbers reading, and the dubbing showed a stepped distribution in the emotional speech performance, we can basically assume that the emotion will affect the breathing and speech acoustic features during phonation, and the more complex and stronger the emotion carried, the greater the influence on the respiratory signal and speech signal.

upward trend, but they do not start at the same point. However, in both reading styles, abdominal and thoracic respiration are basically opposite to each other, when the thoracic respiration is inhaled, the abdominal breath tends to remain exhaled, and vice versa. The dubber's thoracic and abdominal respiration often appear to be in the same frequency. This reflects the fact that the voiceover artist uses a different set of thoracic and abdominal breathing when dubbing and reading. However, in dubbing and dubber’s reading, the resetting of the thoracic or abdominal respiration signal must be accompanied by a pause in speech; however, the resetting of the thoracic or abdominal respiration signal is not always observed at the same time as the pause in speech. This rule does not apply to the pronunciation style of ordinary speakers. Both voice-over and normal pronunciation have a secondary respiratory reset for thoracic and abdominal breathing, whereas when voice-over pronouncers read aloud, thoracic breathing is divided into a secondary reset and abdominal breathing is divided into a tertiary pattern. This may be related to the overkill of the dubbers' reading in order to improve the distinction from the dubbing. The duration of breath resetting and the amplitude of resetting were significantly and positively correlated in all speech styles, but the correlation statistics with the slope of the inspiratory segment showed that they might not be significantly correlated.

In addition, the end of the rhyming phrase of the three voices is basically in the normal voice range. The average F0 of the final rhyming phrases of dubbing is 64.07Hz, with a variation range of 110~224Hz, an average OQ of 42.07%, and an average SQ of 45.76%; in contrast, the average F0 of the final rhyming phrases in dubber’s reading is 86.53Hz, with a variation range of 39~289Hz. In contrast, the F0 of the final rhyming phrases in common reading was 86.53 Hz, with a range of 39-289 Hz, an average of 40.08% for the OQ and 43.49% for the SQ. The large range of variation in the F0 of the modals also leads to the fact that its range overlaps with the F0 of other phonations (for example, the modals crosses the range of variation in the F0 of the squeezed creaky voice in the high pitch segment), and this, together with the stable pitch dip at the end of rhyming sentences in the dubbing, all result in the appearance of the voice characteristic of the creaky voice at the end of rhyming sentences in dubbing, where the range of variation and standard deviation of the F0, OQ, and SQ are much larger than those of the dubber's and common reading, which means that the period variation of the electroglottography signal at the end of each sentence is larger in dubbing; in contrast, the common reading still maintains the modal pattern at the end of the rhyming sentence. This shows that the rhythmic modulation in voiceover is not only expressed by the change of the F0 of the same phonation, but also by the change to other phonation types at the boundaries of the original phonation's range, which further enriches the rhythmic expression in dubbing, a feature that is not available in either dubber's or common reading.

Chapter 4 Summary and Discussion

4.1 Research Results

The narration of nature documentaries needs to be closely matched to the content shown in the video, while always taking care not to overwhelm the audience, so it shows distinctly different characteristics from speech art forms such as reading and news broadcasting. In this paper, we take nature documentaries as the object of study and explore the specific performance of this type of documentary in terms of phonation, articulation and breath resetting, and compare the formal dubbing with the dubber's reading and common reading from the above perspectives. With the help of the speech rhythm analysis system built by Feng Yang (2012) based on Matlab platform, we collect speech signal, EGG signal, thoracic and abdominal respiratory signals to analyze the acoustic and breathing differences between dubbing and reading, and summarize the pronunciation and breathing methods of natural documentary dubbing, so as to better understand and learn this speech art, and provide useful reference for media teaching.

We summarized the breathing and phonetic characteristics of three speech styles: dubbing, dubber's reading and common reading, compared them from multiple dimensions, and the results are shown below.

Table 4.1 Summary comparison of phonetic and respiratory features of the three speech styles

	Contrast items	**Dubbing**	**Dubber's Reading**	**Common Reading**
Phonetic Rhyme	Average duration of rhyming units	Longest	Shortest	longer
	Average interval duration of rhyming units	Longest	Shortest	longer
	*[①] Characteristics of voice type changes at the end of rhyming phrases	Transition from a modal voice to a creaky voice	Modal voice	Modal voice
Breathing Rhythm	Thoracic breathing reset level	2	2	2

① "*"表示该对比项特征能够明显区分出配音和朗读两种语体的不同。

	Abdominal breathing reset level	2	3	2
	* Average reset amplitude, duration, and number of thoracic and abdominal breaths	All three parameters are higher in abdominal breathing than in thoracic breathing.	All three parameters are higher in abdominal breathing than in thoracic breathing; and all thoracic and abdominal breathing were less than the number of breath resets during dubbing.	The average amplitude of abdominal breath resets was higher than that of thoracic breaths, but the average duration and number of resets were equal to those of thoracic breaths. The number of thoracic and abdominal breaths did not differ significantly from the number of breaths in dubber's reading, and both were less than the number of breath resets in dubbing.
	Internal distribution characteristics of amplitude and duration of thoracic and abdominal respiratory resetting	Abdominal breathing is more regular and stable than thoracic breathing resets.	Abdominal breathing is more regular and stable than thoracic breathing resets.	The internal distribution of thoracic and abdominal respiratory resets was almost identical and was the most stable reset in all three speech styles.
	Peak duration interval of thoracic and abdominal respiratory reset	Both thoracic and abdominal breathing: 0.3s~0.7s	Abdominal breathing 0.3s~1.5s, thoracic	Abdominal breathing 0.5s~0.7s, thoracic breathing 0.7s~2.5s.

			breathing 0.2s~0.7s.	
	Correlation between respiratory reset amplitude and reset duration	Obvious positive correlation	Obvious positive correlation	Obvious positive correlation
	*Correlation between respiratory reset amplitude and slope of inspiratory segment	Obvious positive correlation	Obvious positive correlation	No obvious correlation
	*Correlation between respiratory reset duration and slope of inspiratory segment	Obvious positive correlation	No obvious correlation	Obvious positive correlation
	Does a breath reset necessarily occur at a voice pause	Yes	Yes	Yes
	* Does the abdominal breathing reset necessarily correspond to a voice pause	No	No	Yes
	* Whether tone has an effect on thoracic respiratory resetting	Yes	Yes(yet a little)	No
	Relative relationship between the thoracic and abdominal respiratory resetting time points	Abdominal breathing precedes thoracic breathing	Abdominal breathing precedes thoracic breathing	Abdominal breathing precedes thoracic breathing

	* The thoracic and abdominal respiratory signal trend	Basically, the changes appear to be homogeneous, but the difference in the starting point of the changes leads to the appearance of non-synchronous rise and fall.	Basically opposite, the inspiratory segment of thoracic breathing tends to correspond to the expiratory segment of abdominal breathing and vice versa.	Basically opposite, the inspiratory segment of thoracic breathing tends to correspond to the expiratory segment of abdominal breathing and vice versa.

Specifically, first of all, from the point of view of speech rhythm, rhyming phrase length and intervals are flexible and variable in dubbing, and their rhythm is mainly achieved through the appropriate extension or accent of syllables within the rhyming unit rather than simple changes in the length of the interval, which is because the dubbers need to tailor the rhythm of the linear expression of the narration according to the presentation of the picture content, so as to balance the overall requirement of a soothing tone of voice with the semantic requirement of guiding the audience to the full picture meaning. The rhyming phrase ends with a steady dip in pitch, and the phonation type may migrate from the modals range to a creaky voice, which makes the voice fit the picture better and reminds the audience of the change in the content of the narration with the switch of the picture. In contrast, the number of rhyming phrases in the dubber's reading and common reading is greatly reduced or even nearly zero, the rhythm is single, there is no obvious pitch dip at the end of the rhyming sentences, and the phonation remains in the modal range.

In terms of respiratory rhythm, the main differences were shown as follows.

(1) The thoracic and abdominal respiratory signals of both dubber's and common reading show two-level resetting, and the resetting duration and amplitude show a significant positive correlation, while the correlation between the two and the slope of the inspiratory segment is less obvious and stable; while the abdominal respiratory signal of the dubber's reading shows three-level respiratory resetting, and the thoracic respiration still showstwo-level respiratory resetting. This may be due to a deliberate shift made by the dubbers themselves to increase the differentiation between their breathing and vocalization styles during voiceover, but of course, we cannot rule out the possibility that the sample size is not large enough to cause the bias in the results.

(2) The average amplitude, duration, and number of resets of abdominal breathing were higher than those of thoracic breathing in dubbing. This implies that dubbers mobilize more abdominal breathing to power the phonation. A typical example is that after a large thoracic inhalation, the voiceover artist can maintain a maximum of 35s to use mainly abdominal breathing to power the phonation, while the thoracic breathing is weakly reset according to the division of the speech rhythmic units, which can greatly ensure a consistent and smooth breath and a natural expression of emotion. This is also true for the dubber's reading, but the number of breath resets is less than that in dubbing[①], which indicates that the breathing pattern of the same speaker is stable, and the breathing pattern used by the dubber migrates from the broadcast pattern to the non-broadcast style even when reading. The average

① The number of breath resets is higher during voice-over due to emotional expression and other attention allocation needs. Combined with the analysis of Jingyu Zhang (2014; 2016; 2017), the complexity and diversity of attention mechanisms and emotion expression deployment will have an impact on the breathing state of speech: the more attention needs to be allocated and the more complex and varied emotions are instilled, the higher the number of breath resets will be and the more complex the correspondence between breath reset units and rhythmic units will be.

amplitude of abdominal breathing resets was higher than that of thoracic breathing for common reading, but the average duration and number of resets were equal to those of thoracic breathing; the number of thoracic and abdominal breaths did not differ significantly from those of the dubber's reading, and both were less than the number of breathing resets in dubbing. The characteristics of this comparison term also reflect more obvious stylistic differences.

(3) The abdominal breathing was more regular and stable than the thoracic breathing resets in all three speech styles, among which the thoracic and abdominal breathing signals were the most stable in the common reading, which was also caused by the fact that the pronouncer did not consider the emotional infusion and change at all, and all attention was focused on the separate process of converting the text into speech and then issuing it, so the thoracic and abdominal breathing signals were pretty stable, and there were almost no more weak respiratory resets within a distinct respiratory reset.

(4) Breath resetting was always present at speech pauses in all three speech forms, but only in the common reading was it also always present at ventral breath resetting. This means that the abdominal breathing, which powers the articulation of speech, can often be used as a part of the thoracic breathing at the end of the rhythmic unit to complete the speech expression, thus ensuring a wavy flow of speech within the dubbing that does not stop suddenly or appear abruptly. It can be pronounced up and down in accordance with the needs of emotion and language block processing, and it can still be clear and natural when the speech speed is fast, and it can also handle the tendency of speech and the end of the rhythmic unit when the speech speed is slow.

(5) In the dubbing and dubber's reading, the tone has an effect on the resetting of thoracic breathing, and the weak resetting of thoracic breathing is most likely to be triggered when the following syllable is the 4th tone, followed by 1st tone, but it is more obvious in dubbing; the phenomenon is not observed in the common reading. This may be related to the difference in the amount of inspiration required for the pronunciation of the four Mandarin tones in emotional speech expressions.

(6) The thoracic and abdominal respiratory signals in dubbing generally show the same change, but the starting point of the change is not the same, which leads to a non-synchronous rise and fall of the two; however, in both dubber's and common reading, the two basically move in opposite directions, with the thoracic inhalation segment often corresponding to the abdominal exhalation segment and vice versa. The fact that the breathing patterns used by dubbers in dubbing and reading aloud are not the same also indicates that the differences in breathing patterns are much greater than the individual differences of the speakers.

(7) In addition, the breathing patterns of the different speech styles within the dubbing do not differ significantly in terms of breathing resetting grading, relative position of resetting time points, and breathing grouping.

In summary, the two speech styles showed more obvious differences in six dimensions: the type of voice change at the end of the rhyming phrase, the number of resets of thoracic and abdominal breathing, the correlation between the length and amplitude of breathing resets and the slope of the inspiratory segment, the relationship between breathing resets and voice pauses, the overall stability of the thoracic breathing signal curve, and the characteristics of the thoracic and abdominal breathing signal trend. We can use these six acoustic and respiratory characteristics to distinguish the dubbing and reading. However, at the same time, we also notice that the three speech styles, from dubbing to dubber's reading and then to common reading, show gradual changes in several dimensions due to the complexity and diversity of the speakers' attention and emotion deployment. As an intermediate stage, the dubber's reading often carries some of the archetypal characteristics of both dubber's and common reading, and is the most unstable in actual expression, and may appear to favor dubbing or common reading at times.

4.2 Problems and Prospects

The number of dubbing professionals selected in this paper is not enough and they are all broadcasting students during their schooling. Although we have tried to choose dubbers with previous experience when selecting the pronouncers, their voiceover skill is still not rich enough, which may affect the typicality of the collected content of voiceover part of speech and breathing data. Professional dubbers are scarce and most of them are busy, so it is not practical to invite them to a professional recording studio for a longer period of data collection, so the dubbers we chose are considered a compromise, and we will continue to expect dubbing experienced seniors to be discovered and communicated with, and to get valuable opinions about documentary dubbing.

In this experiment, four signals were collected, but due to the time constraint, we mainly analyzed the speech signal and the thoracic and abdominal breathing signal, and only did a statistical analysis of the voice signal as a whole, but lacked a detailed comparative study.

The findings on dubbing in different types and the distinction between pronunciation and breathing patterns in dubbing and reading point to an important role for emotional speech research in this area, but at the operational level we have not really included it in our study and provided a specific experimental design for it, and in fact it is very likely that a clearer explanation of the differences in phonological and respiratory rhythms between reading and dubbing can be obtained by using the paradigm of emotional voice.

References

孔江平. 实验语音学基础教程[M]. 北京大学出版社, 2015: 189-190.

朱景和. 纪录片创作[M]. 中国人民大学出版社, 2002: 4-8.

王明军, 阎亮. 影视配音实用教程[M]. 中国传媒大学出版社, 2014: 1-2.

董春晓. 纪录片制作教程[M]. 浙江大学出版社, 2014.

李潇.人文地理类纪录片解说配音中的问题与策略研究[D].河南大学,2017: 8-20.

巴尔诺. 世界纪录电影史[M]. 中国电影出版社, 1992.

[英] 克里斯廷·林克莱特 著/[荷] 安德烈·斯洛布 绘/彭莉佳 译, 表演练声课[M]. 后浪丨文化发展出版社, 2018.

高明敏.配音艺术——配音语言对角色的表现和塑造方式探析[D].上海师范大学,2020: 15-46.

魏惠敏.新闻播音员做好各类配音工作需要掌握的技巧[J].科技传播,2019,11(23):55-56.

吴峰.电视专题纪录片配音技巧的把控[J].记者摇篮,2020(04):112-113.

黄强, 黄平. "含灯大鼓"唱声声学特征的初步分析[C]// 新世纪的现代语音学——第五届全国现代语音学学术会议论文集. 2001.

于善英.不同歌唱类型歌手共振峰特征及音色形成的机理研究[J].音乐研究,2010(02):74-80+58.

韩启超.南昆念白声学实验分析(一)[J].中国音乐学,2014(04):102-114.

谢更高才旦. 玛曲弹唱口传文化的声学特征分析[D].西北民族大学, 2015:1-36.

陈彧.怀梆“挑后嗓”的音乐声学分析[J].天津音乐学院学报,2016(02):23-29.

曲春燕,刘永祥.京剧嗓音的声学分析及音域特征[J].中华耳鼻咽喉科杂志,2000(01):58-60.

尹基德.汉语韵律的嗓音发声研究[D]. 北京大学, 2010.

杨锋.中国传统吟诵研究[D]. 北京大学, 2012.

吉永郁代. 能剧嗓音发声研究[D]. 北京大学, 2012.

胡阿旭,格根塔娜.蒙古长调的嗓音特征声学分析[J].科技信息,2013(36):8.

董理,孔江平.昆曲闺门旦颤音的嗓音特征[J].清华大学学报(自然科学版),2017,57(06):625-630.

马睿启.山东快书的语音声学分析[D]. 北京大学, 2019.

陈玉东.传媒有声语言语段的构造和调节 [D]. 北京大学, 2004.

李晓华.新闻播音节律特征研究[D].北京大学, 2007.

谭晶晶.汉语普通话朗读时的呼吸节奏研究[D].北京大学, 2008.

钱一凡. 言语发生与艺术发声嗓音特性比较研究[D].北京大学, 2009.

段力, 庞宇, 贺焱秋. 基于数字滤波的呼吸信号去噪研究[J]. 数字技术与应用, 1007-9416（2016）02-0069-02.

李斯农. 人体生理信号中呼吸信号的研究发展综述[J]. 中国国际财经:中英文, 2017(11):286-288.

吴君如. 呼吸带说明.

曲海红. 播音语言的用声研究[D]. 吉林大学.

曲海红, 孔江平. 从实验语音学的角度考量播音艺术用声气息的变化[J]. 现代传播(中国传媒大学学报), 2012, 34(008):132-133.

曲海红, 王倩. 播音情韵语音中情感、呼吸与语音表现的关系[J]. 现代传播-中国传媒大学学报, 2014,

036(008):149-150.
邓德崇. 格律诗词的粤语吟诵和朗读的语音及呼吸韵律研究[D]. 暨南大学.
衡点. 汉语自然话语陈述句音高下倾研究[D]. 暨南大学.
藏语新闻朗读语音停顿和呼吸重置关系研究[J]. 西北民族大学学报(自然科学版), 32(01):20-24.
印雪晨, 吕士良, 段力令,等. 呼吸生理信号在实验语音学中的应用研究[J]. 生命科学仪器, 2014(3):24-27.
李广友. 浅谈播音员主持人的语音艺术[J]. 新闻传播, 2014(14):69-69.
李智. 现代汉语口语疑问句语调升降规律及其成因研究[D]. 暨南大学.
殷治纲. 汉语普通话朗读语篇节奏研究. 中国社会科学院研究生院, 2011.
王毓钧. 汉语学习者言语韵律与呼吸节律的交互关系研究[D]. 南开大学.
谭晶晶, 李永宏, 孔江平. 汉语普通话不同文体朗读时的呼吸重置特性[J]. 清华大学学报:自然科学版, 2008(S1):613-620.
张金爽, 胡阿旭, 于洪志. 呼吸生理在言语声学中的应用研究[J]. 生命科学仪器, 2010, 08(006):31-34.
许成龙. 关于播音主持艺术语音发声理论的几点思考——以几部主流播音主持语音发声教材为思考对象[J]. 声屏世界, 2020, No.469(16):39-41.
李永宏, 张金爽, 王姝文,等. 藏语新闻朗读的呼吸信号声学分析[J]. 西北民族大学学报(自然科学版), 31(02):17-21.
张春连, 孔江平. 新闻朗读的呼吸节奏与音高的关系初探[C]// 第十二届全国人机语音通讯学术会议(NCMMSC'2013)论文集. 2013.
杨阳蕊, 于洪志, 李永宏. 基于呼吸信号的藏语诗歌韵律研究[J]. 微计算机信息, 2010, 26(028):20-22.
印雪晨, 段力令, 李永宏,等. 宋词呼吸韵律结构模型研究[J]. 生命科学仪器, 2014(4):41-44.
尹基德, 孔江平. 韩语呼吸节奏与语调群的关系初步研究[C]// 第九届全国人机语音通讯学术会议(NCMMSC2007)论文集. 2007.
李永宏. 基于 MATLAB 的呼吸和气流信号分析系统[J]. 西北民族大学学报(自然科学版), 2015, 36(04):19-22.
吴韩娜, 孔江平. 汉语和韩语近体诗的呼吸节奏特征对比研究[J]. 中国语音学报, 2016(1):71-78.
张锦玉. 朗读、复述与自述状态下的言语呼吸研究[J]. 西北民族大学学报(自然科学版), 2016(2):10-16.
刘璇, 方华萍, 李永宏. 蒙古族短调民歌的呼吸生理信号分析[J]. 西北民族大学学报(自然科学版), v.34;No.89(01):13-17.
刘璇, 李永宏. 基于呼吸带的蒙古短调民歌呼吸信号比较分析[J]. 西北民族大学学报(自然科学版), 34(03):28-32.
张锦玉. 普通话自然言语的呼吸特征研究[J]. 西北民族大学学报:自然科学版, 2017, 038(001):38-47.
张锦玉. 歌唱呼吸特征的实验研究[J]. 浙江艺术职业学院学报, 2014, 12(002):75-82.
蔡青, 朱维箫, CAI-Qing,等. 台词教学中的语音发声问题刍议[J]. 郧阳师范高等专科学校学报, 2010, 30(2):113-115.
高晓梅. 阿拉伯语不同文体朗诵的韵律声学分析. 西北民族大学, 2019.
孔江平. 语言发声研究及相关领域[A]. 中国中文信息学会.新世纪的现代语音学——第五届全国现代语音学学术会议论文集[C].中国中文信息学会:中国中文信息学会,2001:8.
Hird K , Kirsner K . The relationship between prosody and breathing in spontaneous discourse.[J]. Brain

& Language, 2002, 80(3):536-555.

Stemple, Joseph C. Vocalization Subsystem Responses to a Temporarily Induced Unilateral Vocal Fold Paralysis.[J]. Journal of Speech, Language, and Hearing Research, 2018, 61.

Collyer S , Davis P J . Effect of Facemask Use on Respiratory Patterns of Women in Speech and Singing[J]. Journal of Speech Language & Hearing Research, 2006, 49(2):412.

Collyer S , Thorpe C W , Callaghan J , et al. The Influence of Fundamental Frequency and Sound Pressure Level Range on Breathing Patterns in Female Classical Singing[J]. Journal of Speech Language & Hearing Research Jslhr, 2008, 51(3):612.

Parham D F , Buder E H , Oller D K , et al. Syllable-Related Breathing in Infants in the Second Year of Life[J]. Journal of Speech Language and Hearing Research, 2011, 54(4):1039-1050.

McFarland, David H . Respiratory Markers of Conversational Interaction[J]. J Speech Lang Hear Res, 2001, 44(1):128-143.

Boliek C A , Hixon T J , Watson P J , et al. Refinement of Speech Breathing in Healthy 4-to 6-Year-Old Children[J]. Journal of Speech Language and Hearing Research, 2009, 52(4):990-1007.

Lester R A , Hoit J D . Nasal and Oral Inspiration During Natural Speech Breathing[J]. Journal of Speech Language & Hearing Research Jslhr, 2014, 57(3):734-42.

Proctor D F . Breathing, Speech, and Song[M]. Springer-Verlag, 1980.

Huber J E , Spruill J . Age-related changes to speech breathing with increased vocal loudness.[J]. Journal of Speech Language & Hearing Research, 2008, 51(3):651-668.

Connaghan KP, Moore CA, Higashakawa M. Respiratory Kinematics During Vocalization and Nonspeech Respiration in Children From 9 to 48 Months.[J]. Journal of speech, language, and hearing research. 2004;47(1):70-84.

Fuchs S, Petrone C, Rochet-Capellan A, Reichel UD, Koenig LL. Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands.[J]. Journal of phonetics. 2015;52(Sep):35-45.

Dromey C , Ramig L O . Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation[J]. Journal of Speech Language and Hearing Research, 1998, 41(5):1003-1018.

Shokrollahi M, Saha S, Hadi P, Rudzicz F, Yadollahi A. Snoring sound classification from respiratory signal. United States: IEEE; 2016.

Ramos-Garcia RI, Tiffany S, Sazonov E. Using respiratory signals for the recognition of human activities. United States: IEEE; 2016.

Ladefoged P. Some Physiological Parameters in Speech. Language and speech. 1963;6(3):109-19.

Hardcastle WJ, Laver J, Gibbon FE. The handbook of phonetic sciences. 2nd ed. Chichester, U.K;Malden, MA;: Wiley-Blackwell; 2010.

Harrington J, Cassidy S. Techniques in Speech Acoustics. Dordrecht: Springer Netherlands; 1999.

Klatt DH. Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America. 1980;67(3):971-95.

Rabiner LR, Schafer RW. Digital processing of speech signals. Englewood Cliffs, N.J: Prentice-Hall; 1978.

Renov M . Theorizing Documentary[M]. 1993.

Fant G. Speech sounds and features. Cambridge: MIT Press; 1973.

Appendix A

Animal World - *The Never Ending Footsteps* - Complete Narration

(Requirements: no excessive empty frames in the selection, no excessive pause time, and a steady pace of speech that varies according to the needs of the content, rather than being flat all the time so that we can analyze its control of rhythm and rhyme.)

狼群很快就认识到了这些驯鹿速度太快，体格太强壮了。作为这场会战的旁观者，我们对狼群能否取胜表示怀疑，但随着冬季越发的严寒，这一切也可能随之改变。在如此遥远的北极，冬季是一个黑暗的季节，太阳一连几周隐没在地平线之下，只留下北极光在黑暗的夜空中流光飞舞。

到了二月份，太阳重新开始照耀大地，雌鹿这个时候已经有了五个月的身孕，几个月来一直忍受着狼群的侵扰。冰封的湖面曾一度保护着他们，但现在他们必须去森林中觅食，而那里危机四伏。一头雌鹿带领大家来到一处食物丰盛之地，灵敏的嗅觉告诉他们，食物埋藏在雪下一米深的地方。每吃一口，都要扒开积雪，进食也就成了一种繁重的任务，因此整个冬天驯鹿的体重都在下降。有一种纤细的地衣，富含碳水化合物，又被称做驯鹿苔藓。驯鹿试图靠它来保持体重，但很难达到平衡。为了补充觅食时消耗的能量他们必须从食物中获得更多的营养。在母亲密切的监护下，这只九个月大的小鹿仔正学着进食。大自然将鹿角和荷尔蒙赐予怀孕的雌鹿，保证他们能顺利的度过严冬。鹿角将有孕在身的雌鹿与其他的驯鹿区分开来，在同伴们扒开积雪之后他们可以直接享用下面的食物。这为他们节省了宝贵的精力，在拼命逃生时这往往意味着生与死的差别。

在冬季，死亡的威胁从未走远。狼群可以趁驯鹿不备悄悄潜入森林。与在湖面上相比，他们此刻离驯鹿更近。现在的形势对狼群有利。在厚厚的雪层上，如果哪只驯鹿体力不支，一眼就能看出来。头狼以迅雷不及掩耳的速度迅速压紧包围群。一般而言，丢掉性命的驯鹿并非因为体力不支，而是仅仅犯了个错误——在大家都向右转的时候这头小鹿却向左转过了身。它想重新跟上大部队的步伐但那里还有头狼正等着它。又来了一头狼，这一切很快结束。在这个世界上，犯不起错，要想生存，必须敏捷强壮，而且永不犯错，差一点点狼都不会手下留情。我们觉得这群雌鹿有股毅然决然的

气魄，尽管失去了一头一岁大的小鹿仔，而这些母亲还是义无反顾，继续向前。

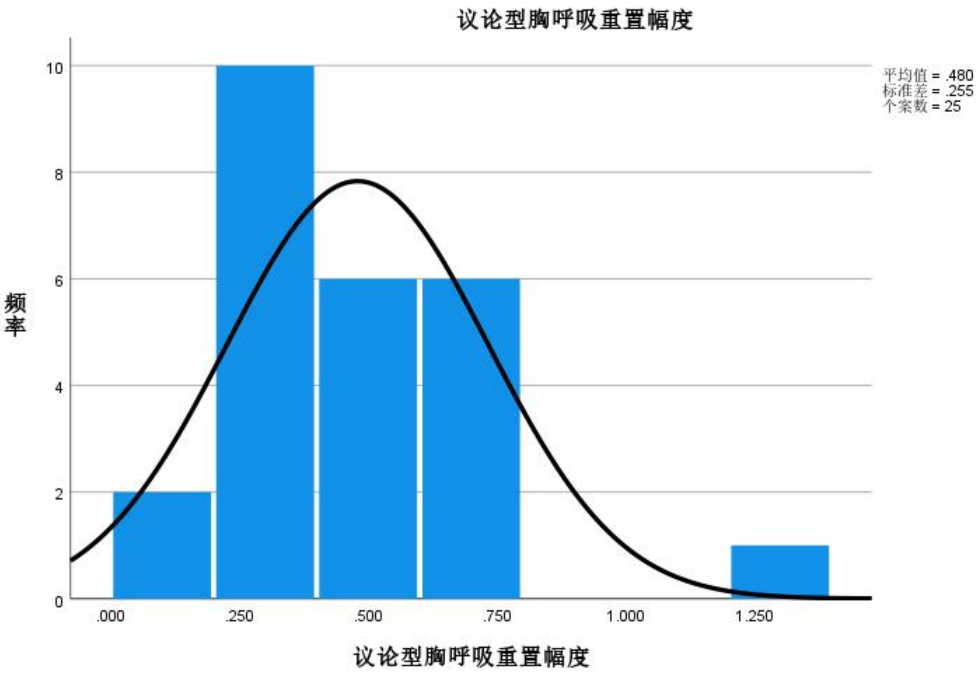
到了四月底，阳关灿烂，怀孕的雌鹿们开始焦躁不安，腹中的小生命一天天长大，他们开始向北方深情凝望，那里就是他们的繁殖区。但要抵达那里，他们必须离开森林的保护，来到广袤而荒凉的北极苔原。这是一场与时间的赛跑，在怀孕的最后一个月里平均下来这些雌鹿每天要在这片险恶的土地上跋涉二十公里。时间刻不容缓，必须在孩子出世之前抵达，那时候，冰雪消融，食物也将露出地面。这是一段漫长的旅程。在地球最为荒凉的版图上，他们要走上六百公里才能抵达位于巴瑟斯特角北冰洋海岸的传统繁殖区。大地开始回暖，此时走在齐胸深的冰雪间，每一步都充满了艰辛。湖面开始融化，到处都是烂泥，但他们没有时间绕道而行。

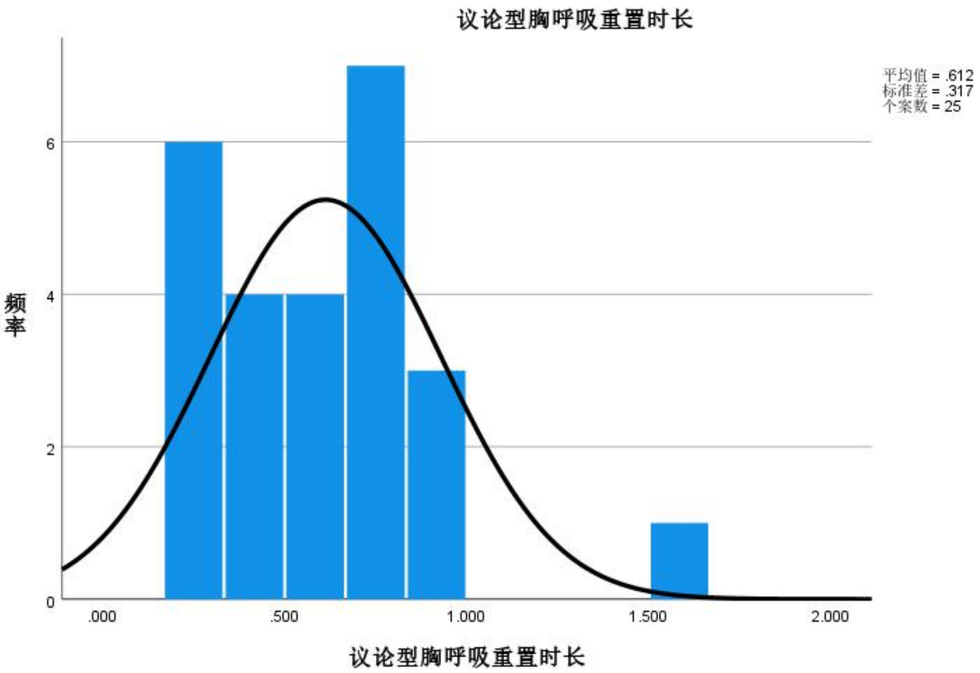
这一疯狂的迁徙之旅也带来了一个好处——他们把狼群远远甩在了身后。往南几百公里的地方，沿着这些多沙的山脊，狼群停下脚步开始产仔。他们需要多沙的泥土，以避开霜雪，挖掘洞穴。七年以来，一支狼群家族在这里繁衍壮大。接下来的三个月里，当幼崽慢慢长大的时候，这对首领夫妇也只好停留在原地，只能希望，当驯鹿下季迁徙的时候，可以离自己近一些。

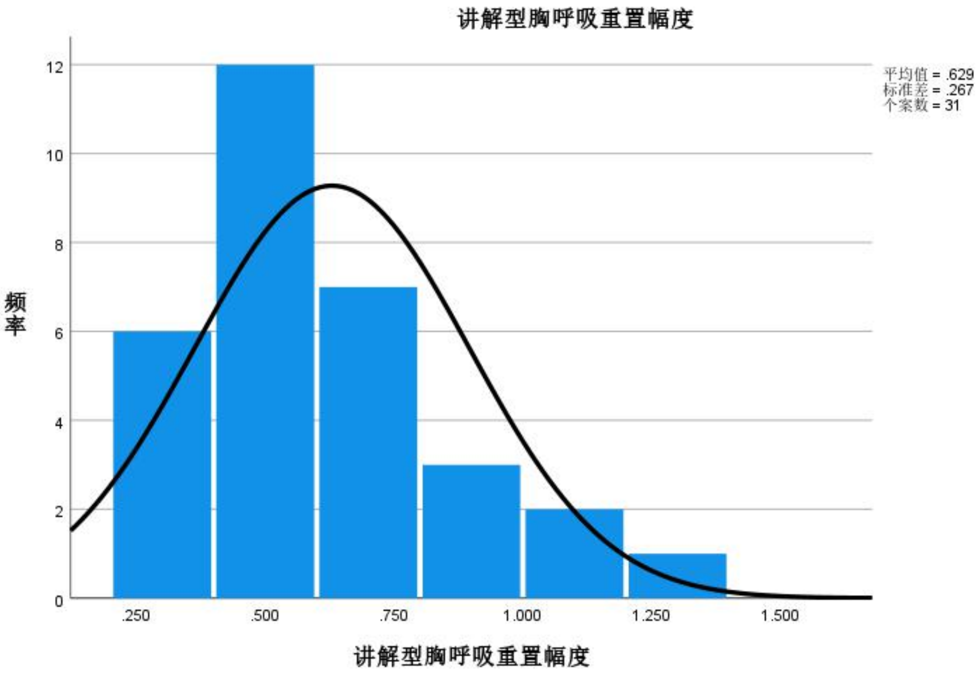
但此刻驯鹿仍然一路向北。六月初，怀孕的雌鹿最终抵达了繁殖区。越冬区和繁殖区构成了驯鹿的整个世界。他们的一生都在这二者之间来回无尽的迁徙。这头雌鹿顺利的完成了它的使命。这些雌鹿都已有孕在身，是什么促使他们踏上这一漫长而险恶的征程呢？此刻，只有在繁殖区才能找到营养丰富的食物。几个月来他们一直饱受饥饿的折磨，腹中的小生命也在一天天长大，他们现在急需补充营养。在继续向北迁徙的过程中，他们将绝大多数食肉动物甩在了身后。棕熊并不像狼群那样在驯鹿身后紧紧地跟随。他们常年在此地居住，时机成熟时才会出击。这只棕熊饥肠辘辘，甚至想对怀孕的雌鹿下手，他希望吃上刚出世的小鹿仔。

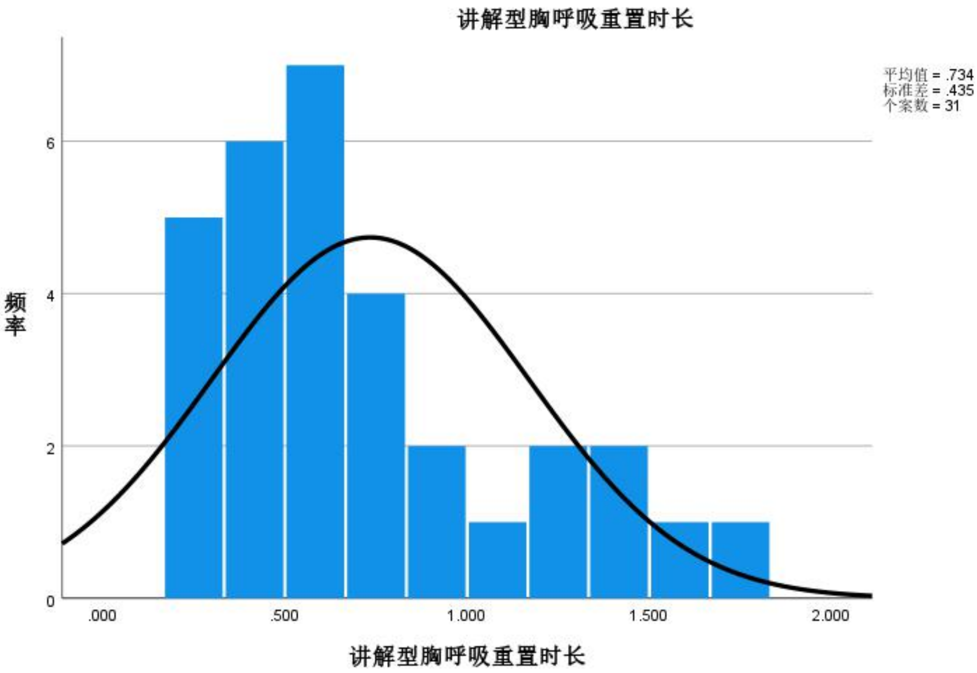
Appendix B

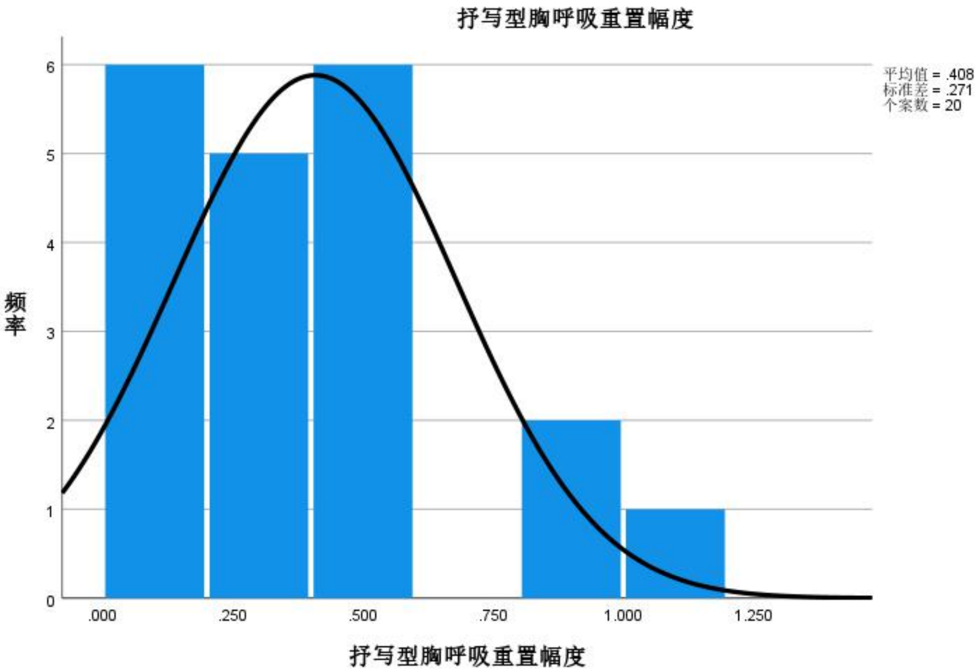
Histogram of frequency distribution of breathing parameters for different language styles of voiceover.

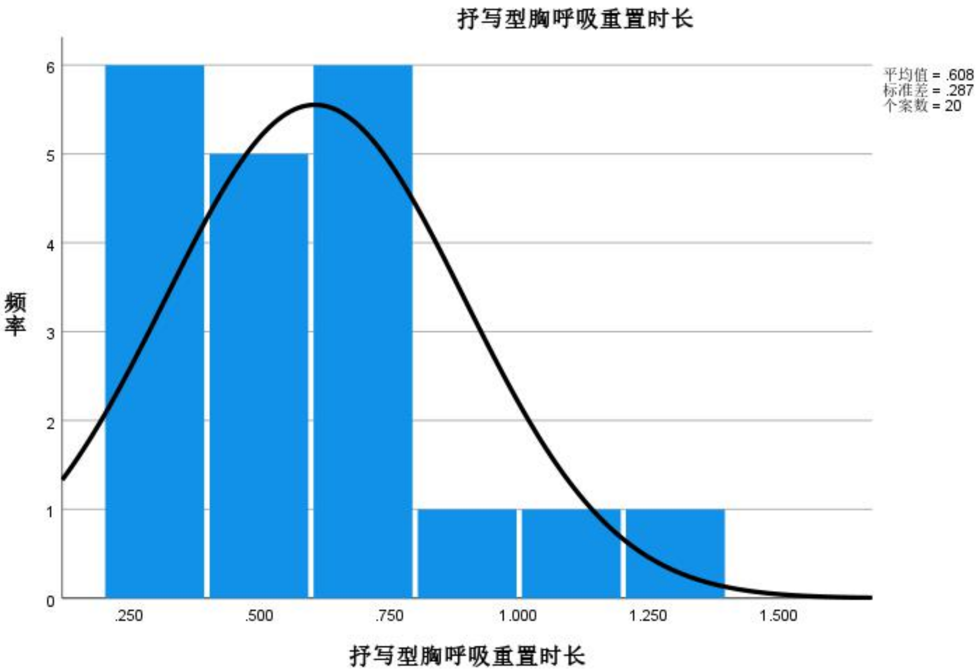


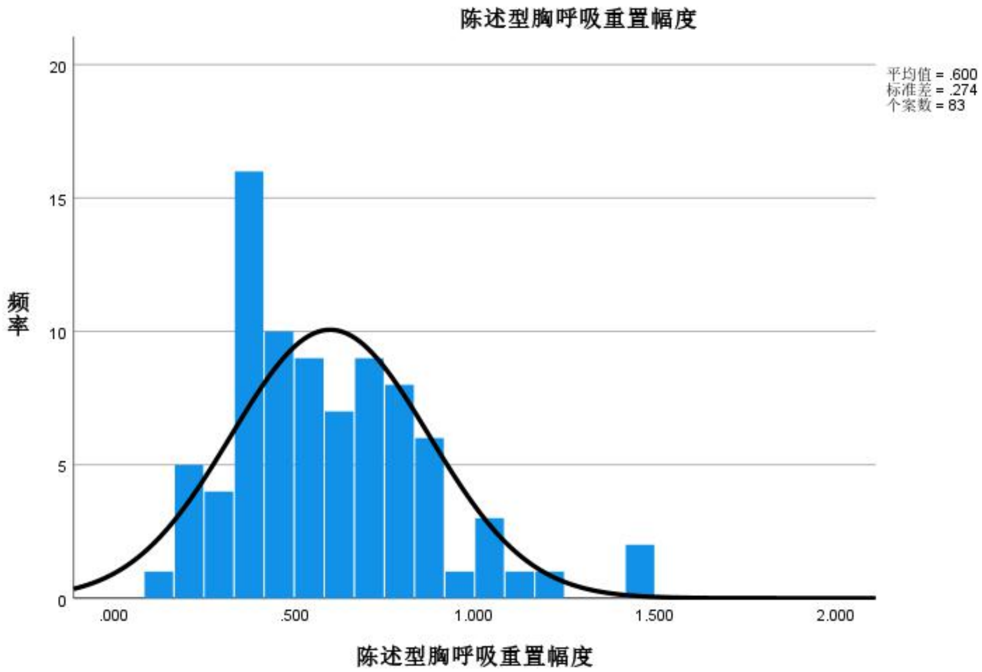


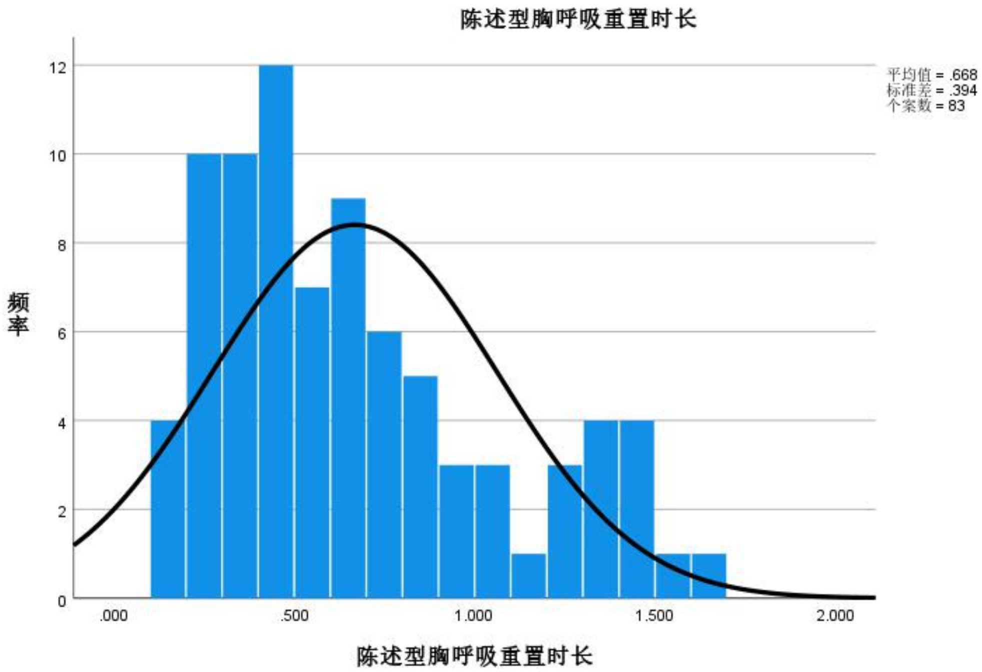

讲解型胸呼吸重置幅度
平均值 = .629
标准差 = .267
个案数 = 31
频率
12
10
8
6
4
2
0
.250
.500
.750
1.000
1.250
1.500
讲解型胸呼吸重置幅度

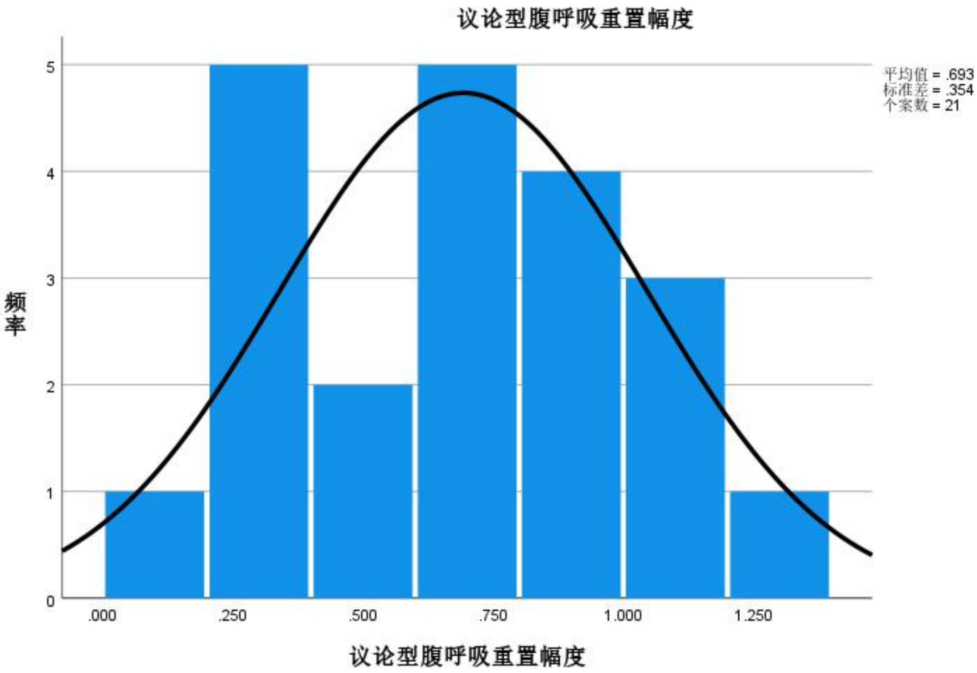

讲解型胸呼吸重置时长
平均值 = .734
标准差 = .435
个案数 = 31
频率
6
4
2
0
.000
.500
1.000
1.500
2.000
讲解型胸呼吸重置时长

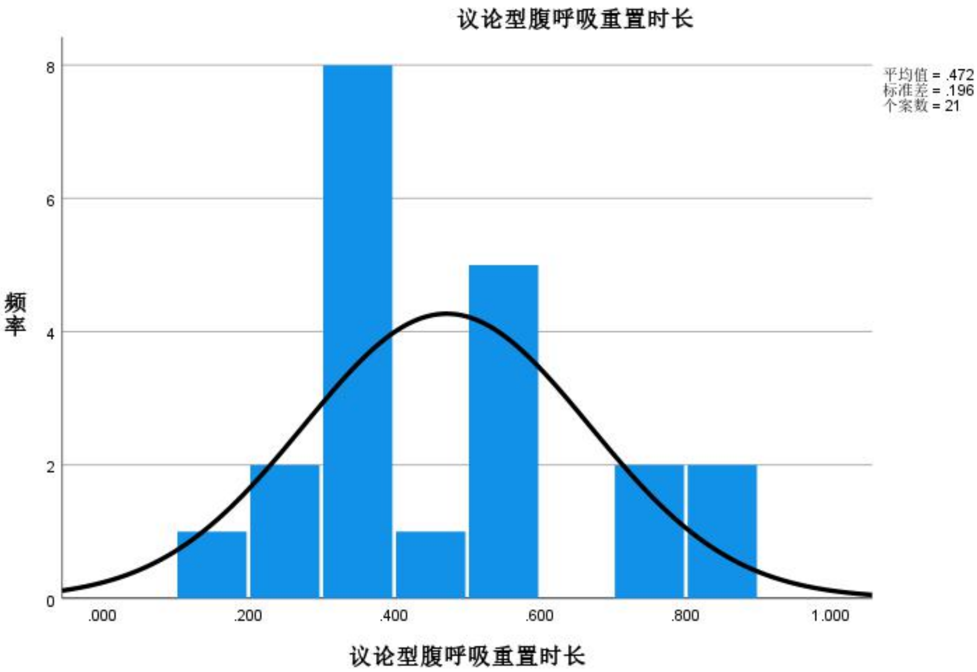

抒写型胸呼吸重置幅度
平均值 = .408
标准差 = .271
个案数 = 20
频率
0
1
2
3
4
5
6
.000
.250
.500
.750
1.000
1.250
抒写型胸呼吸重置幅度

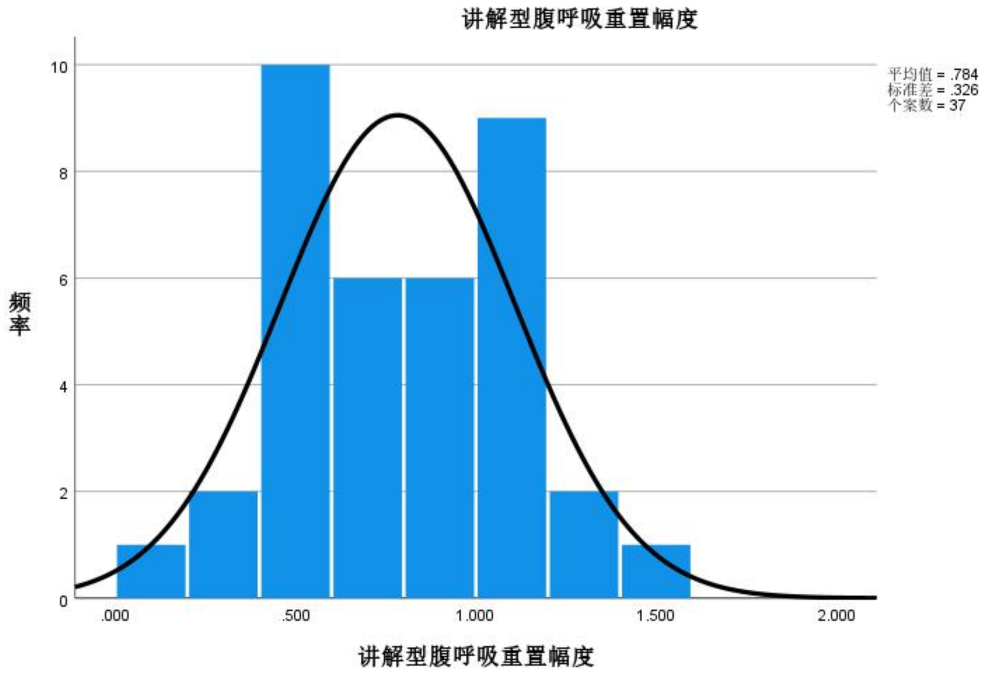

抒写型胸呼吸重置时长
平均值 = .608
标准差 = .287
个案数 = 20
频率
0
1
2
3
4
5
6
.250
.500
.750
1.000
1.250
1.500
抒写型胸呼吸重置时长


陈述型胸呼吸重置幅度
平均值 = .600
标准差 = .274
个案数 = 83
频率
20
15
10
5
0
.000
.500
1.000
1.500
2.000
陈述型胸呼吸重置幅度

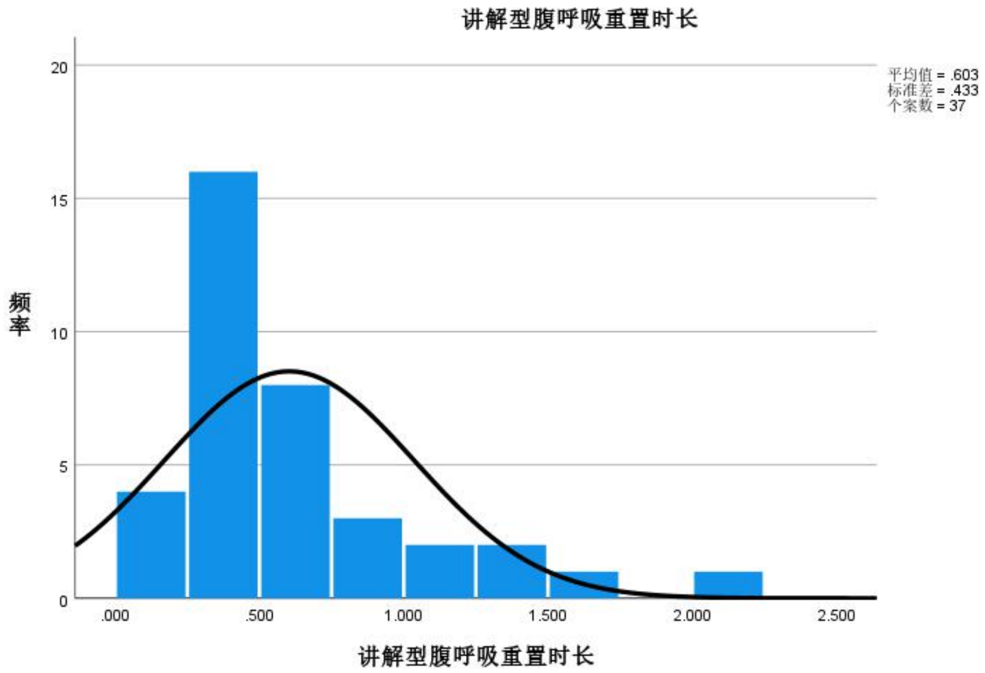

陈述型胸呼吸重置时长
平均值 = .668
标准差 = .394
个案数 = 83
频率
12
10
8
6
4
2
0
.000
.500
1.000
1.500
2.000
陈述型胸呼吸重置时长


议论型腹呼吸重置幅度
平均值 = .693
标准差 = .354
个案数 = 21
频率
5
4
3
2
1
0
.000
.250
.500
.750
1.000
1.250
议论型腹呼吸重置幅度

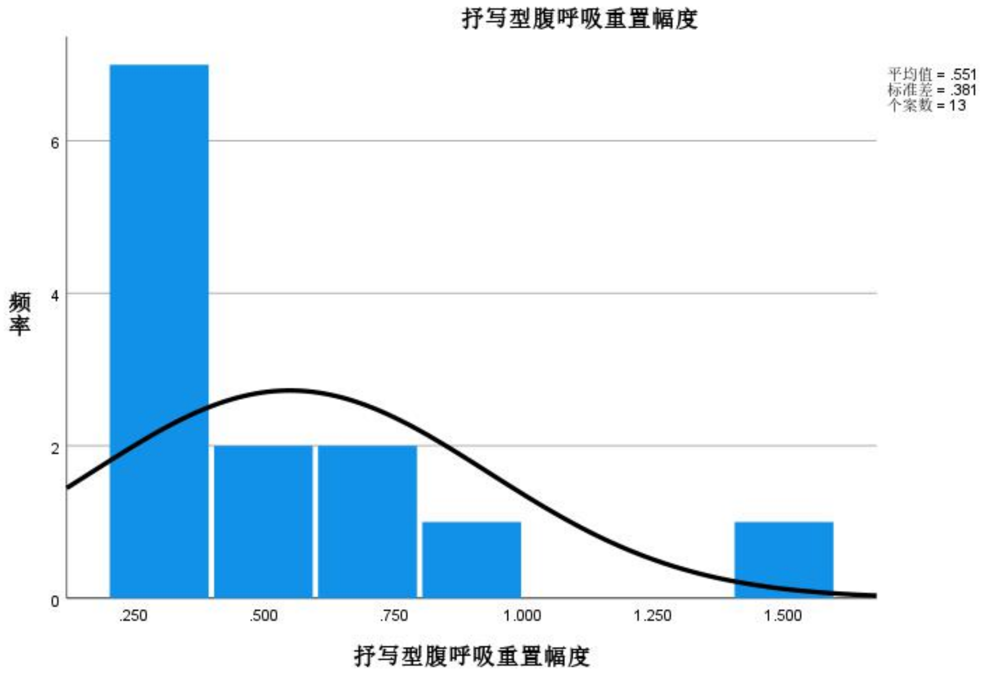

议论型腹呼吸重置时长
平均值 = .472
标准差 = .196
个案数 = 21
频率
8
6
4
2
0
.000
.200
.400
.600
.800
1.000
议论型腹呼吸重置时长

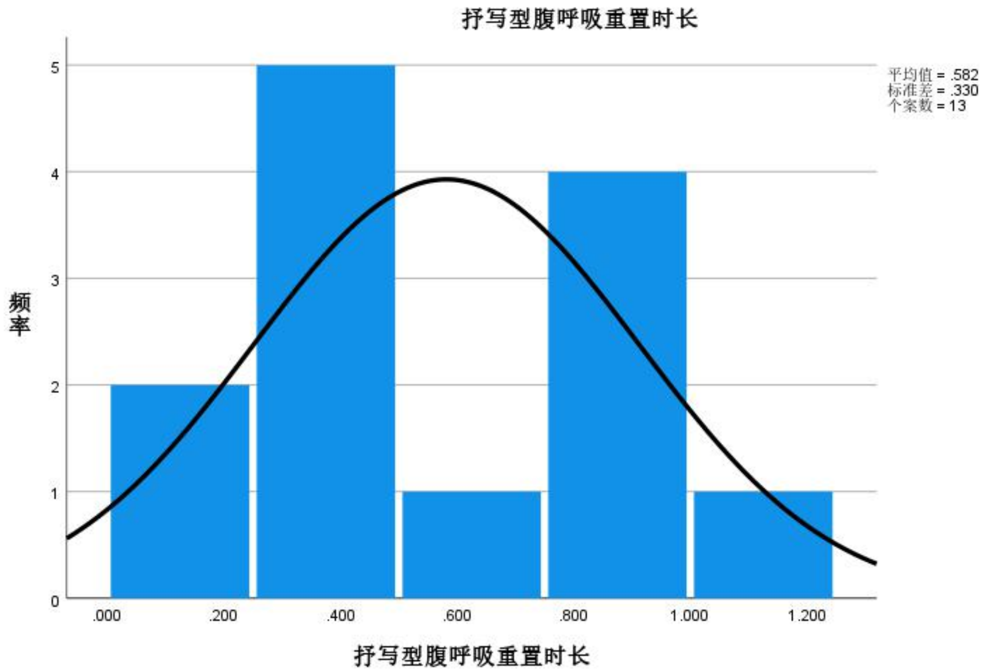


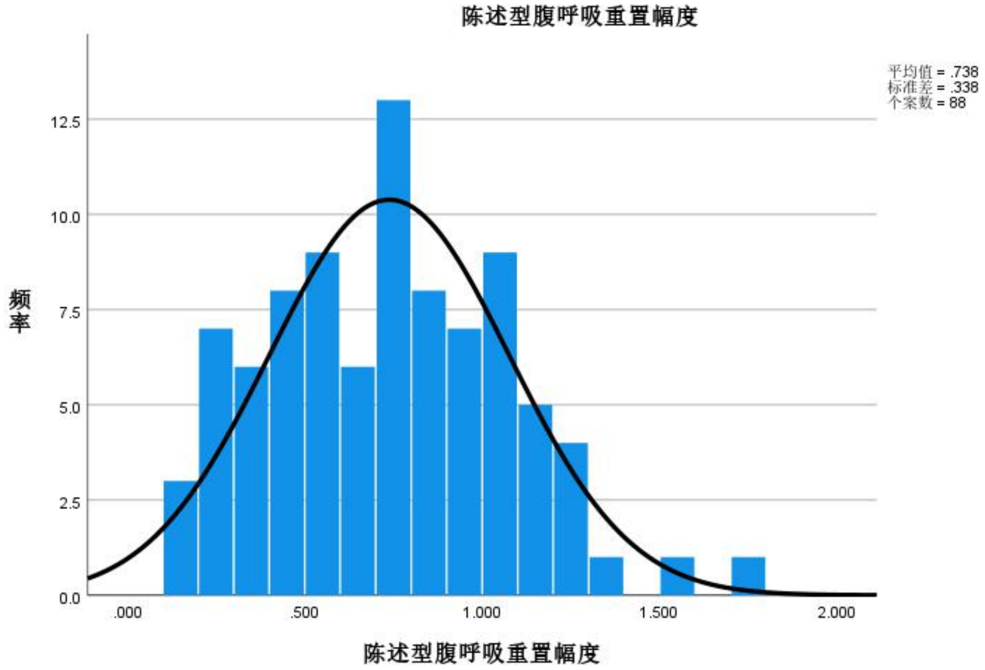
讲解型腹呼吸重置幅度
平均值 = .784
标准差 = .326
个案数 = 37
频率
0
2
4
6
8
10
.000
.500
1.000
1.500
2.000
讲解型腹呼吸重置幅度



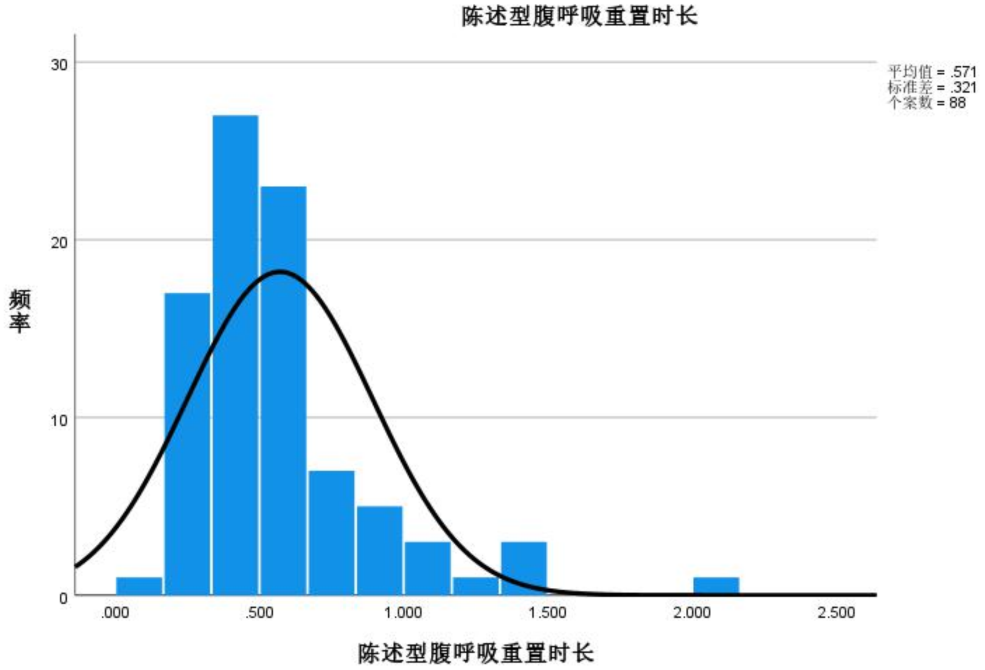
讲解型腹呼吸重置时长
平均值 = .603
标准差 = .433
个案数 = 37
频率
0
5
10
15
20
.000
.500
1.000
1.500
2.000
2.500
讲解型腹呼吸重置时长


抒写型腹呼吸重置幅度
平均值 = .551
标准差 = .381
个案数 = 13
频率
0
2
4
6
.250
.500
.750
1.000
1.250
1.500
抒写型腹呼吸重置幅度


抒写型腹呼吸重置时长
平均值 = .582
标准差 = .330
个案数 = 13
频率
0
1
2
3
4
5
.000
.200
.400
.600
.800
1.000
1.200
抒写型腹呼吸重置时长



陈述型腹呼吸重置幅度
平均值 = .738
标准差 = .338
个案数 = 88
频率
12.5
10.0
7.5
5.0
2.5
0.0
.000
.500
1.000
1.500
2.000
陈述型腹呼吸重置幅度



陈述型腹呼吸重置时长
平均值 = .571
标准差 = .321
个案数 = 88
频率
30
20
10
0
.000
.500
1.000
1.500
2.000
2.500
陈述型腹呼吸重置时长