

语音乐律研究报告

Status Report of Phonetic and Music Research

2015



北京大学中文系语言学实验室

Linguistics Lab

Department of Chinese Language and Literature

Peking University

目录

1. Jiangping Kong. "The Physiological Aspects of Phonetics", Journal of Chinese Linguistics, Volume 43, Number 1B, January, pp.265-267	3
2. Kong, Jiangping. "A dynamic glottal model through high-speed imaging.", Journal of Chinese Linguistics 43.1: 311-336.	6
3. Zhang, Ruifeng, R. J. Baken, and Jiangping Kong. "Vocal Attack Time of Different Pitch Levels and Vowels in Mandarin." Journal of Voice 29.5: 542-547.	32
4. Yang, Feng. "A study on the features of chest and abdominal breathing between reciting and chanting Chinese poetry." Journal of Chinese Linguistics 43.1:399-410.	38
5. Wang, Gaowu, and Jiangping Kong. "An articulatory model of standard Chinese using MRI and X-ray movie." Journal of Chinese Linguistics 43.1: 269-294.	50
6. Li, Yinghao. "Prosodic boundaries effect on segment articulation in standard Chinese: An articulatory and acoustic study." Journal of Chinese Linguistics 43.1: 364-398.	75
7. Wang, Feng. "Variations of laryngeal features in Jianchuan Bai." Journal of Chinese Linguistics 43.1: 434-452.	110
8. 张春连, 含灯大鼓的声学研究, 硕士毕业论文.....	126
9. 吴韩娜, 基于呼吸的汉韩韵律对比研究, 博士毕业论文.....	202
10. 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, NCMSC, 2015 天津...	338

THE PHYSIOLOGICAL ASPECTS OF PHONETICS

The study of modern phonetics in China dates from 1924 when Liu Fu published the book *Sisheng shiyan lu* 四聲實驗錄 (Record of Experiments on Four Tones) (1924) which first discovered that the basis of the tones is fundamental frequency. It symbolized the beginning of modern phonetics in China. In the following 60 years, the researches of modern phonetics developed with difficulties because of wars and political movements. The turning point came in 1979 when Professor William Shi-Yuan Wang gave a series of phonetics lectures at Peking University, just after China ended the called “great cultural revolution” and adopted the open door policy to the world in politics, economics and academic research.

In the following 36 years, modern phonetic study in China which has been developing greatly can be divided into two periods. The first period is from the 80s to the end of the last century. Phonetic study mainly focused on the acoustical analysis by which the fundamental frequency, formants and amplitude were mainly used because acoustical parameters were relatively easier to extract by a researcher with a non-technological background. A collection of papers entitled *Beijing Yuyin Shiyan Lu* 北京語音實驗錄 (Record of Experiments on Mandarin Speech) edited by Professors Lin Tao and Wang Lijia was published in 1985. The second period is from the beginning of this century to the present during which the phonetic study focused mainly on the physiological aspects because more and more instruments and methods were used such as laryngography (EGG), instruments for measuring air-pressure, air flow, aspiration, MRI, EMA and so on.

Some of the researches and the publication were supported by NNSFC grant (No.10&ZD125).

In this special issue, 10 papers on physiological aspects of phonetics are presented.

There are four papers on speech models, including three papers for geometrical models of Mandarin, and one paper for the physiological articulatory model. “An Articulatory Model of Standard Chinese Using MRI and X-ray” has explored the articulatory mechanism of speech production in Standard Chinese and developed a geometrical articulatory model, both in visual and acoustic modalities, based on the data of MRI images and an X-ray movie. “A Two-Dimension Lip Model for Mandarin Chinese”, established a two dimensions lip model with inner and outer lip contours which has well defined the linguistic term ‘lip rounding’ and was used to generate audio-visual stimuli for the speech perception experiment of the McGurk Effect. “A Dynamic Glottal Model through High-speed Imaging” has introduced a dynamic glottal model based on high-speed imaging and the model controlled by dynamic glottal widths, lengths, F0, Open Quotient (OQ) and Speed Quotient (SQ) which can produce speech sources with different phonation types. “A Control Strategy of a Physiological Articulatory Model for Speech Production” has constructed a full three-dimensional physiological-articulatory model, including the tongue, jaw, hyoid bone and vocal tract wall, based on the continuum finite element method.

There are three papers on physiological phonetic studies that used electropalatography (EPG), instruments for air-pressure and flow, and aspiration. “Prosodic Boundaries Effect on Segment Articulation in Standard Chinese: An Articulatory and Acoustic Study” has investigated the EPG and the acoustic data of the prosodic boundaries effect on the domain-initial segments in Standard Chinese, aiming to examine the domain-initial strengthening in both spatial and temporal dimensions. “A Study on the Features of Chest and Abdominal Breathing between Reciting and Chanting Chinese Poetry” has studied the features of chest and abdominal breathing between reciting and chanting Chinese poems of different styles. “An Aerodynamic Study on Articulation of Mandarin Initials” has studied the aerodynamic features of Mandarin initial consonants with different articulatory places and manners and found that the parameters of expiratory airflow duration (EAD), peak air-pressure (PAP), peak expiratory airflow (PEA) and expiratory volume (EV) can all

be regarded as the distinctive features in consonants.

There are three papers concerned with phonation types of tones by using EGG signal and perception test. “Variations of Laryngeal Features in the Jianchuan Bai” has examined the tonal quality based on three parameters of F0, OQ and SQ from EGG signals and found that there are two non-modal phonation types, namely Harsh and Pressed. “The Role of Phonation Cues in Mandarin Tonal Perception” has investigated the role of phonation cues in perceiving Mandarin tones in isolated syllables and found that it is necessary to define language tones in a finer model by incorporating detailed phonation parameters. “The Creaky Voice and its Tonal Description Method” has studied the different phonation types in Chinese Dilu dialect through F0, OQ and SQ from EGG signals and proposed a tone transcription system for languages that have various phonation types.

Along with the development of speech science and neuroscience, more and more modern phonetic researches have adopted instruments and methods of neuroscience, such as ERP, fMRI and the perception method in studying the activities in the human brain. We hope another special issue of collected papers about speech neuroscience will be published in the near future. Finally, with this special issue, we commemorate the 90th anniversary of the establishment of the Phonetics and Music Laboratory in 1925 at Peking University.

Spring 2015

Jiangping Kong*
Guest Editor

* **Jiangping Kong** 孔江平 is a professor of Peking University at the Center for Chinese Linguistics, Department of Chinese Language and Literature, and Joint Research Center for Language and Human Complexity. Contact: Dept. of Chinese Language and Literature, Peking University, Beijing 100871, China; [jpkong@pku.edu.cn]

A DYNAMIC GLOTTAL MODEL THROUGH HIGH-SPEED IMAGING

Jiangping Kong*

Peking University

ABSTRACT

This paper is a study for an improved dynamic glottal model through high-speed imaging (HSI).¹ As is well known, speech production comprises three parts, namely speech source, speech resonance and lip radiation. Among these three parts, speech source is the most important one because it is the basis of speech. In research on speech production, acoustical models of speech source have been well established. But the physiological speech source, that is to say, the activity of the glottis is seldom researched, because the vibration of the vocal folds is difficult to observe and sample. A study on the glottal model was established many years ago (Kong 2007), and in that model, the static glottis was modeled by four quarters of ellipses in three modes namely normal mode, leakage mode and open mode. The dynamic glottal control function was modeled by an approximation of multiplication of sine and exponential. The problem of the dynamic glottal model is that the control parameters can't be well explained, though the glottis can be simulated. In this study, more high-speed images were sampled, the image processing was greatly improved and the dynamic glottal control function was modeled with parameters which were significant to speech perception.

SUBJECT KEYWORDS

High-speed imaging Vibration of vocal folds Dynamic glottal model

* **Jiangping Kong** 孔江平 is a professor of Peking University in the Center for Chinese Linguistics, Department of Chinese Language and Literature, and Joint Research Center for Language and Human Complexity. Address: Dept. of Chinese Language and Literature, Peking University, Beijing 100871, China; [jpkong@pku.edu.cn]

1. INTRODUCTION

Speech source is very important because it is one of the three parts in speech production, these are speech source, vocal tract resonance and lip radiation. From the viewpoint of the speech signal, at least three kinds of signals, i.e., sound pressure, airflow and glottal area function, can be used to model the dynamic glottis. The glottal excitation, that is to say, the glottal flow and sound pressure, have been studied and modeled acoustically by many researchers. But up to now, the activities of the glottis are seldom studied and modeled, because the vibration of the vocal folds is difficult to observe and sample. In recent years, more and more good quality high-speed images were sampled through high-speed video cameras. This study has focused on improving the glottal control function of the dynamic glottal model, and then the application of the model was discussed.

The classical models of speech source are the one-mass model and two-mass model (Flanagan and Landgraf 1968; Lucero 1993, 1996; Pelorson et al. 1994). These models have been established theoretically. Acoustical models based on speech sound have also been well studied. There are seven models which are considered important, namely: 1) Rosenberg (1971); 2) Hedelin (1984); 3) Fant (1979); 4) Fant (1982); 5) Ananthapadmanabha (1984); 6) Fant, Liljencrants, and Lin (1985) and 7) Ljungqvist and Fujisaki (1985). The brief definitions of these acoustical models are shown in Figure 1 to 3.

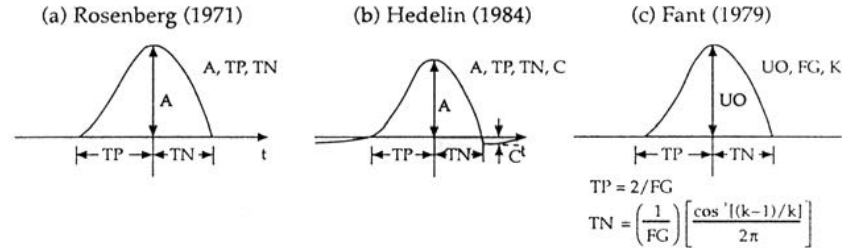


Figure 1 Acoustical models developed by (a) Rosenberg, (b) Hedelin, and (c) Fant

Figure 1 shows three models that are developed in pulse of glottal flow. The model of Rosenberg was proposed in 1971. The model of Hedelin was developed in 1984. The model of Fant was set up in 1979. The brief definitions are as shown in Figure 1.

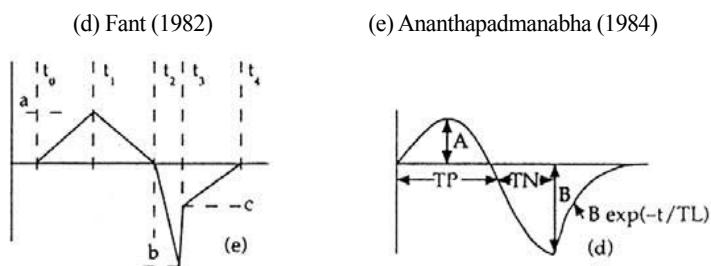


Figure 2 Models developed by (d) Fant and (e) Ananthapadmanabha.

Figure 2 shows the models developed in the differential form of glottal flow, that is to say, the sound pressure. Fant's model was developed in 1982, and the Ananthapadmanabha's model was developed in 1984. The brief definitions are shown in Figure 2.

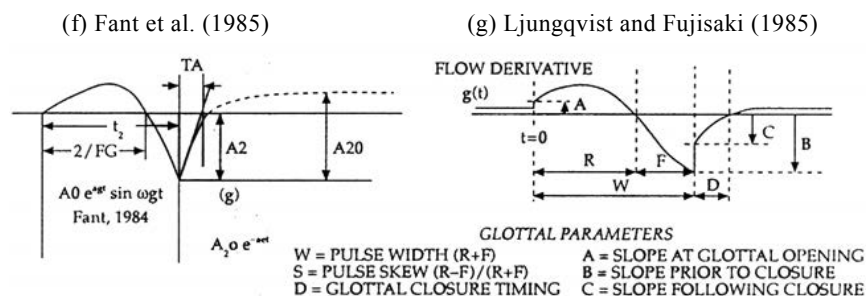


Figure 3 Models developed by (f) Fant et al and (g) Ljungqvist and Fujisaki

Figure 3 shows the models developed through the sound pressure. Fant's model was developed in 1985, and Ljungqvist and Fujisaki's model was also developed in 1985. The brief definitions are shown in Figure 3.

By comparing these models, we can find that the three models in Figure 1 are based on glottal flow and the others in Figures 2 and 3 are based on sound pressure. It is clear that the scholars first developed their models according to pulse of glottal flow with few parameters, and then found that the models were not effective and flexible for simulating different phonation types. Then the researchers established their models on sound pressure with relatively more parameters which were more flexible and effective. Among these models, the LF-model developed by

Fant et al (1985) is the most effective and flexible one.

The earlier model established by Fant used three parameters and it allowed the change of open quotient (Fant 1979a, 1979b, 1980). Then the LF-model reported by Fant et al in 1985 comprises F-model (Fant) and a modification by Liljencrant (Fant et al. 1985).

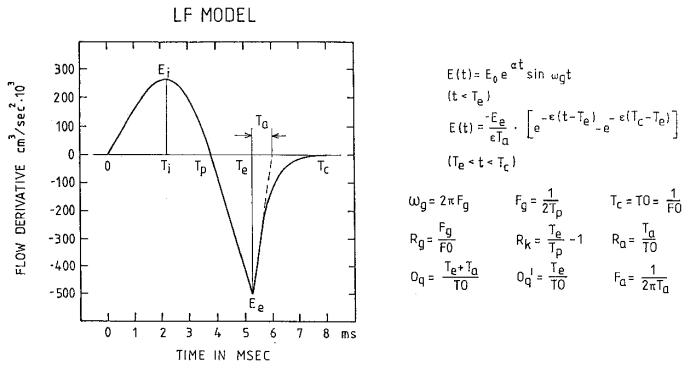


Figure 4 The basic definition of the LF-model.

The LF-model has two phases as shown in Figure 4. The first phase is from 0 to 'Te', and it is created by multiplication of sine and exponential. The second phase is from 'Te' to 'Tc', which is created by the function of exponential. The formulas in Figure 4 show the relationships among the parameters. With this model, a glottal pulse of sound pressure can be specified by four parameters, i.e., 'Tp, Te, Ta and Ee'. In addition, the parameter 'Tc' is equal to 'T0', which is equal to 1/F0. Up to now, the LF-model is one of the best models in simulating the different glottal source. The purpose of introducing the LF-model here is to provide a reference for a comparative study with the physiological glottal model as presented in this paper.

There are many methods to study the speech source, such as the acoustical method and the physiological method. From the viewpoint of signals, speech source can be studied through the acoustical signal, electroglottography (EGG) signal, air-pressure signal, high-speed image signal of the vocal folds, and so on. Among these signals, the signal of the high-speed digital image of vocal vibration is the one which can directly and reflect well the nature of vocal vibration and explain the relationship of

the movement of the vocal folds and the characteristics of the speech sound.

The earliest sample of vocal vibration was captured by Bell Telephone Laboratories in the 1930s, and from that time on, high speed motion pictures have been used to study the vibration of the vocal folds. At present, there are some systems which can be used to study the vibration of the vocal folds including the high-speed digital imaging system in the University of Tokyo, the system of kymography by Kay, the Weinberger Speedcam system, the Kodak Ektapro system, and so on. The development of high-speed digital image systems has provided a good foundation for studying the vibration of the vocal folds.

2. GLOTTAL DETECTING

The method and procedures of this research include the sampling of high-speed digital images, image rotation and cropping, image motion compensation, image contrast adjusting and parameter extracting.

2.1 Sampling of High-speed Images

In this study, samples of vocal fold vibration that are simultaneous with the signals of EGG and speech sound were captured by the high-speed imaging system produced by KAY with endoscope. The sampling rates are 2000 and 4500Hz. The samples include different phonation types, sustained vowels with low, middle and high pitches, samples of vowels with gliding voice from low to high and high to low and samples of the four basic Mandarin tones.

2.2 Image Rotation and Cropping

The samples were pre-processed by rotation and cropping. The images could be rotated automatically or manually, according to the concrete samples, because sometimes it was very difficult to rotate samples into the right position, especially for the disordered glottis. The samples captured by the Kay system were in the size of 128×256 pixels, and they could be cropped to the size that one wants. The image in the far left of Figure 5 was cropped to the size of 100×100 pixels and appears at the far right in Figure 5.

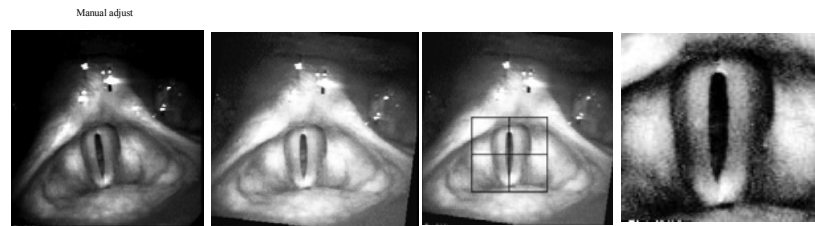


Figure 5 The first picture is an original image, the second has been rotated by the algorithm of bi-cubic, the third has been windowed, and the fourth has been cropped to the size of 100×100 pixels.

2.3 Image Motion Compensation (MC)

One of the adverse factors affecting the accuracy and validity of high-speed video (HSV) quantitative assessment is the motion of the endoscope lens relative to the larynx. Endoscopic motion makes it difficult to track the dynamic characteristics of the laryngeal anatomic structures, when we divide the glottis in to left and right parts, as shown in Figures 6 and 7.

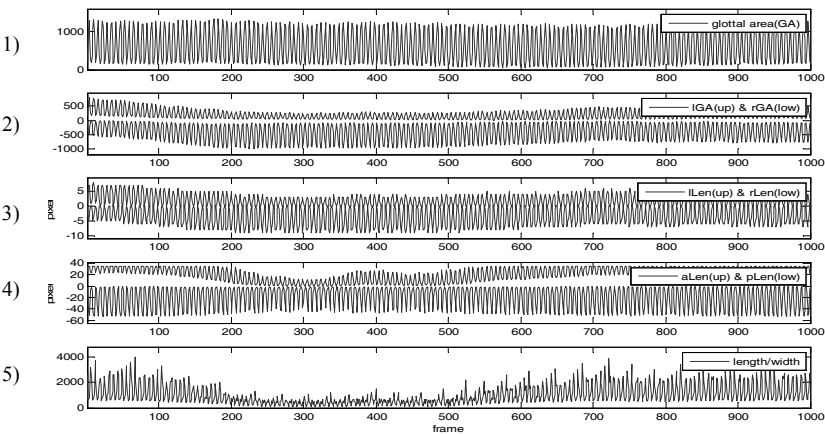


Figure 6 The glottal parameters before MC

From top to bottom: 1) the glottal area (GA); 2) the left and right GA; 3) the left and right glottal width; 4) the anterior and posterior glottal length; 5) the ratio of glottal length to glottal width.

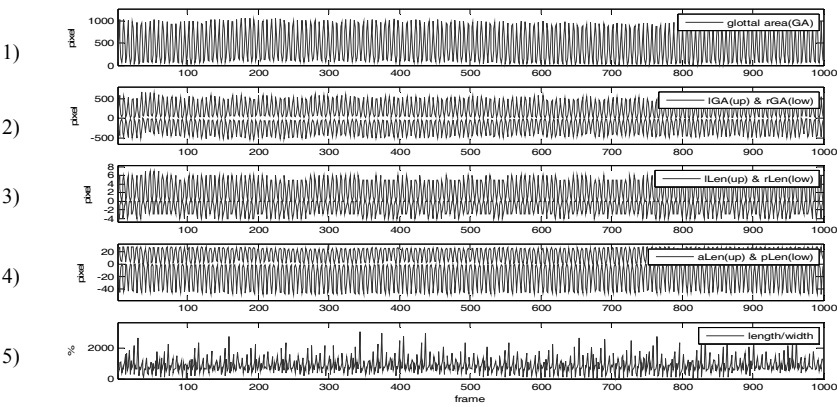


Figure 7 The glottal parameters after MC

Therefore, we have to perform “motion compensation” (MC), which means the process of detecting and removing the endoscopic motion from the HSV image. Here we mainly use the idea and method published by Dimitar D. Deliyski (2005). Figure 7 shows the result after MC. Comparing the parameters in Figures 6 and 7, we can see that results after MC are very good.

2.4 Image Contrast Adjusting

Another adverse factor affecting HSV quantitative assessment is the contrast adjustment of the HSV image when we binarize the images to get the shape of the glottis, which depends on different operators. This will influence the quantitative estimation of the glottis area and other parameters. Therefore, we automatically adjust the contrast of the HSV images with the method as shown in Figure 8.

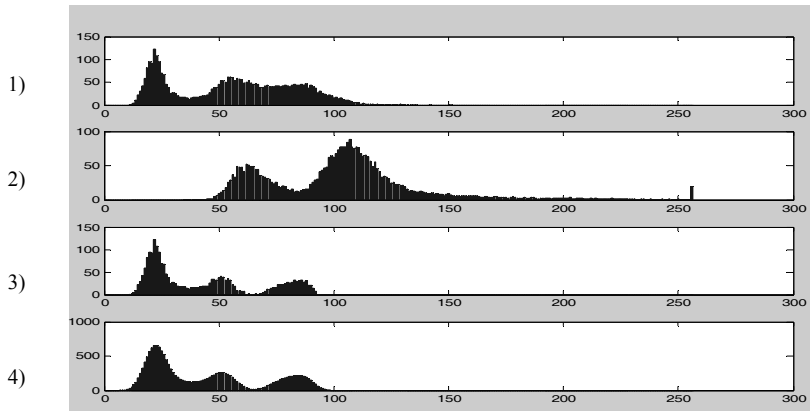


Figure 8 The method of contrast adjusting

From top to bottom: 1) The accumulated histogram when the glottis is fully opened in the first 100 frames of the video; 2) The accumulated histogram when the glottis is fully closed in the first 100 frames of the video; 3) The subtraction which gives us the histogram of the glottis area, where there is a peak in the low gray region, which reflects the gray value of the glottal region. 4) We smooth this histogram, and use the gray values at the left and right sides of the first peak to automatically adjust the contrast of all the frames in this HSV to get binarized images which show us the shape of the glottis.

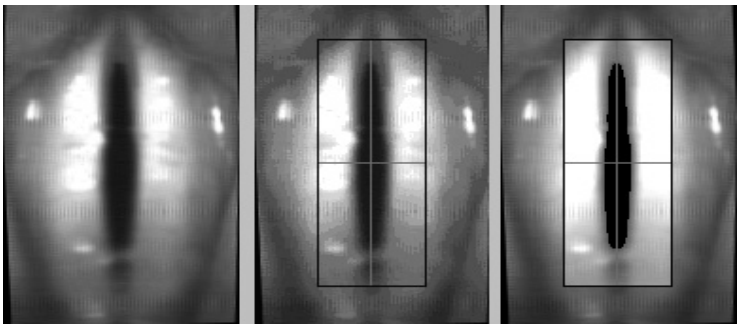


Figure 9 The left image is an original image, the middle is the image with a window and the right is the glottis detected image.

Figure 9 shows the result of contrast adjusting. The left image is an original image of the modal female voice. The middle image has a window which is used for limiting the area of the glottis. The right image shows the glottis automatically detected by the system.

2.5 Parameter Extracting

After the glottis has been detected and the area of the glottis has been obtained, definitions need to be given in order to extract parameters of the glottis. In Figure 10, two graphs are given to help in describing the glottis and setting up definitions.

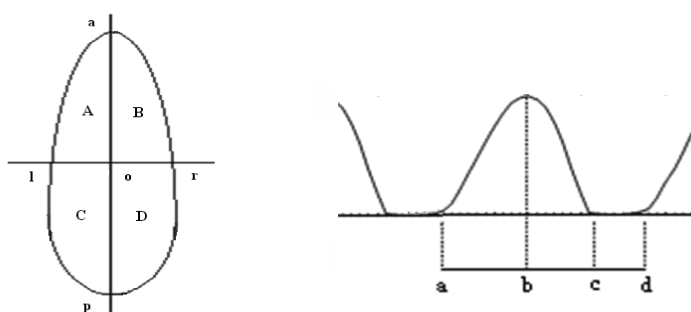


Figure 10 The basic definition of the glottis and glottal parameters, fundamental frequency (F0), open quotient (OQ) and speed quotient (SQ)

The left diagram in Figure 10 shows a glottis, in which ‘ABCD’ stands for the whole glottis, ‘AB’ stands for the anterior glottis, ‘CD’ stands for the posterior glottis, ‘AC’ stands for the left glottis, ‘BD’ stands for the right glottis. In addition, ‘o’ is the center of the glottis, ‘lo’ is the left width of the glottis, ‘ro’ is the right width, ‘ao’ is the anterior length, and ‘po’ is the posterior length. The right diagram in Figure 10 illustrates a period of glottal area function, in which ‘a’ stands for glottal opening instance, ‘b’ stands for the local maximum of the glottal area, ‘c’ stands for glottal close instance, and ‘d’ stands for the next glottal opening instance.

The following are definitions of F0, OQ and SQ: 1) the fundamental frequency is defined as $1/'ad'$ (Hz); 2) The open quotient is defined as the ratio of ‘ac’ over ‘ad’; 3) The speed quotient is defined as the ratio of ‘ab’ over ‘bc’. See the right portion of Figure 10.

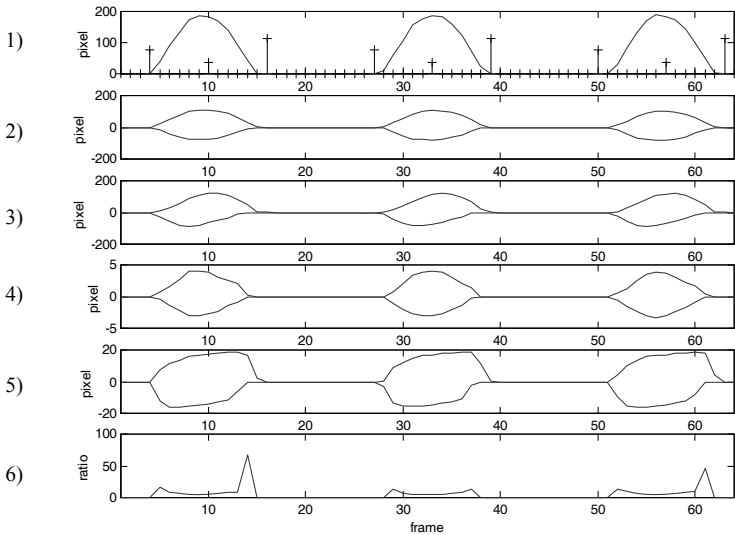


Figure 11 The 13 parameters extracted by the system

Figure 11 shows the parameters extracted by our system. From top to bottom, 1) the dynamic glottal area with glottal opening instance, glottal close instance and the local maximum; 2) left and right glottal areas; 3) anterior and posterior areas; 4) left and right widths; 5) anterior and posterior lengths; 6) ratio of length over width. In this study, these parameters were used for dynamic glottal modeling.

3. MODELING OF DYNAMIC GLOTTIS

In the dynamic glottal model (Kong 2007), the static glottis was modeled by four quarters of ellipses in three modes, namely, normal mode, leakage mode and open mode. The dynamic glottal control function was modeled by an approximation of multiplication of parabola and sinusoid. The problem of the dynamic glottal model is that the control parameters can't be well explained, though the glottis can be well simulated. In this study, the static glottis was also modeled by four quarters of ellipses and the improvement of the model was focused on the dynamic glottal control function.

3.1 Model of Static Glottis

In this study, the static glottis was also modeled by four quarters of ellipses, and the normal mode of static glottal model was used to explain the new dynamic glottal control function, (see Figure 12).

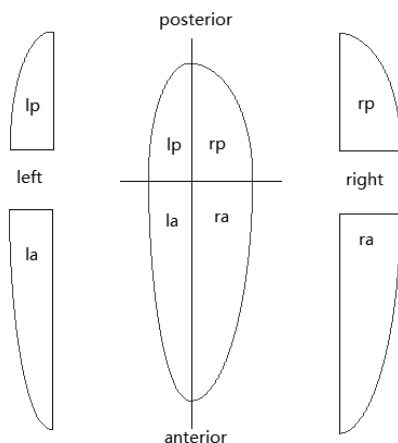


Figure 12 The static glottis was modeled by four quarters of ellipses.

In Figure 12, we can see the static glottis was modeled by four quarters of ellipses, which are ‘left-posterior’ (lp) ellipse, ‘right-posterior ellipse’ (rp), ‘right-anterior ellipse’ (ra) and ‘left-anterior ellipse’ (la). The four quarters of ellipses were calculated by the two elliptical semi-major axes and two elliptical semi-minor axes, respectively.

3.2 Glottal Properties of Dynamic Glottis

The left and right dynamic widths and the anterior and posterior dynamic lengths in one glottal period are regarded as dynamic glottal control functions. In the dynamic model, they are the contours of the two elliptical semi-major axes and two elliptical semi-minor axes in one period, which are used to drive the dynamic glottal model and synthesize a glottal pulse. According to the parameters extracted from the high-speed images of different phonations, the dynamic glottal control function can be classified into six basic types, which are approximations to different parts of the sinusoid as shown in Figure 13.

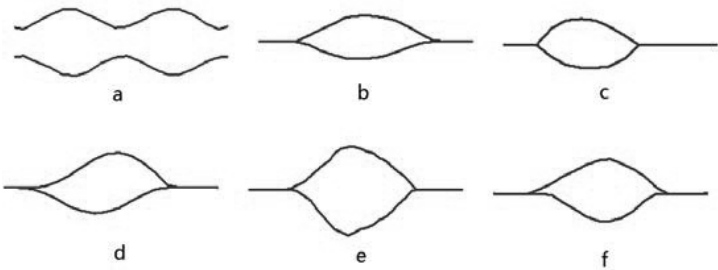


Figure 13 The basic types of dynamic glottal control function

There are six images in Figure 13. Image ‘a’ displays two dynamic glottal control functions which are approximations to two sinusoids; image ‘b’ displays two dynamic glottal control functions which are approximations to a part of the sinusoid from 300° of the first sinusoid to 240° of the second sinusoid; image ‘c’ displays two dynamic glottal control functions which are approximations to a part of the sinusoid from 0° to 180°; image ‘d’ displays two dynamic glottal control functions whose local maximum are not the same, and which lead to the different SQs; image ‘e’ displays two dynamic glottal control functions which are approximations to a part of the sinusoid from 270° of the first sinusoid to 180° of the second sinusoid; image ‘f’ displays two dynamic glottal control functions whose lengths are not the same and which lead to the different OQs.

3.3 Modeling of Dynamic Glottis in Open Phase

In order to model the dynamic glottal control function more exactly, four parts of functions were chosen from two periods of the sinusoid (see Figure 14). The angles of these two sinusoids are from 0° to 720°. The chosen part of the sinusoid in this study was defined by the angle of the first period of the sinusoid and the angle of the second period of the sinusoid, respectively.

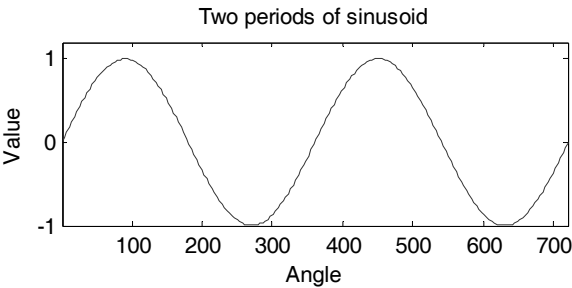


Figure 14 Two periods of the sinusoid are displayed in this figure. The x axis is angle, which is from 0° to 720° and the value is from -1 to 1.

In these two periods of sinusoid, only the part between 270° of the first period and 270° of the second period is used for modeling the dynamic control function. According to the properties of the real glottis, four typical parts of the sinusoid were chosen to model the dynamic glottal control function.

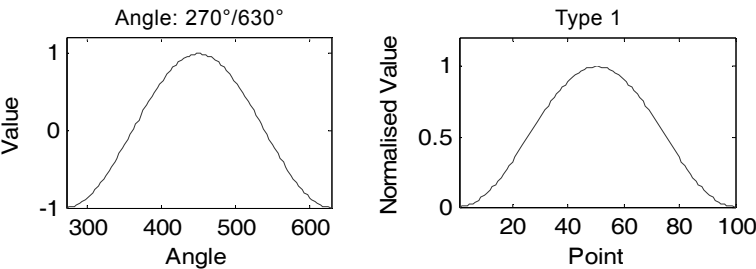


Figure 15 The left image shows a part of the sinusoid chosen from 270° to 630° of the two periods of sinusoids and the right image shows the normalized period of the left image.

In Figure 15, the left image displays a part of the sinusoid from 270° to 630° of the two periods of the sinusoid, whose value is between -1 and 1. The values are normalized from 0 to 1 for y axis and 100 points for the x axis in the right image. It will then be used as a part of the dynamic glottal control function. It is called ‘type 1’.

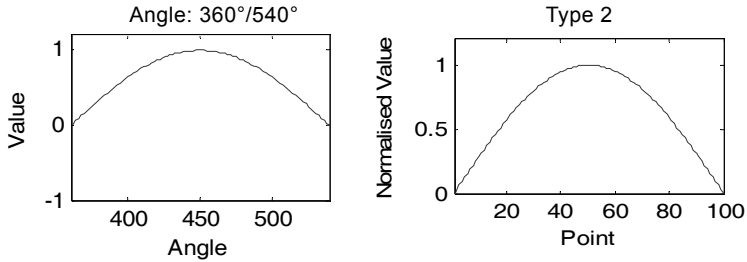


Figure 16 The left image shows a part of the sinusoid chosen from 360° to 540° of the two periods of the sinusoid, and the right image shows the normalized period of the left image.

In Figure 16, the left image displays a part of the sinusoid from 360° to 540° of the two periods of the sinusoid, whose value is from 0 to 1. The values are normalized from 0 to 1 for the y axis and 100 points for the x axis in the right image. It will be used as a part of the dynamic glottal control function. It is called ‘type 2’.

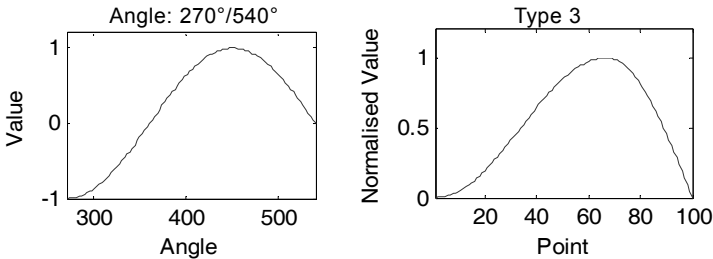


Figure 17 The left image shows a part of the sinusoid chosen from 270° to 540° of two periods of the sinusoid and the right image shows the normalized period of the left image.

In Figure 17, the left image displays a part of the sinusoid chosen from 270° to 540° of the two periods of the sinusoid, whose value is between -1 and 1. The right image shows the normalized period in the left image whose value is from 0 to 1 for the y axis and 100 points for the x axis. It will be used as a part of dynamic glottal control function. It is called ‘type 3’.

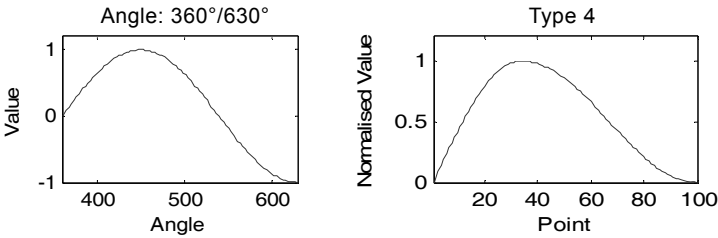


Figure 18 The left image shows a part of the sinusoid chosen from 360° to 630° of two periods of the sinusoid and the right image shows the normalized period of the left image.

In Figure 18, the left image displays a part of the sinusoid chosen from 360° to 630° of the two periods of the sinusoid, whose value is between -1 and 1. The right image shows the normalized period of the left image whose value is 0 to 1 for the y axis and 100 points for the x axis. It will then be used as a part of the dynamic glottal control function. It is called ‘type 4’.

3.4 Modeling of Dynamic Glottal Control Function

Based on the four types discussed above, the parameters of F0, OQ and SQ were added to produce a whole dynamic glottal control function. In order to explain them easily, the angles chosen from the two periods of the sinusoid will be set and explained separately. See Figure 19.

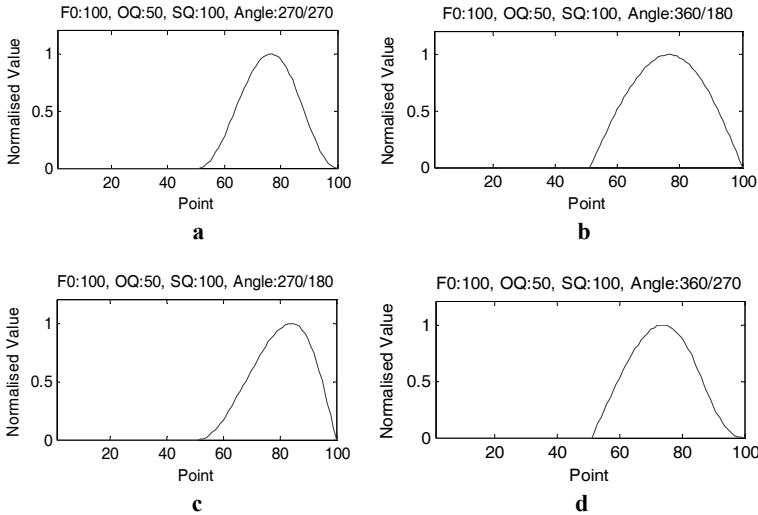


Figure 19 Four plots which show the four typical types of dynamic glottal function. Plot ‘a’ shows type1, plot ‘b’ shows type 2, plot ‘c’ shows type 3, and plot ‘d’ shows type 4.

In Figure 19, plot ‘a’ displays the dynamic glottal control function of type 1 whose F0 is 100 Hz (the sampling rate is 10k), OQ is 50% and SQ is 100%, and the pulse in open phase is chosen from 270° of the first period of the sinusoid to 270° of the second period of the sinusoid; plot ‘b’ displays the dynamic glottal control function of type 2 whose F0 is 100 Hz, OQ is 50% and SQ is 100% and the pulse in open phase is chosen from 360° of the first period of the sinusoid to 180° of the second period of the sinusoid; plot ‘c’ displays the dynamic glottal control function of type 3 whose F0 is 100 Hz, OQ is 50% and SQ is 100%, and the pulse in open phase is chosen from 270° of the first period of the sinusoid to 180° of the second period of the sinusoid; plot ‘d’ displays the dynamic glottal control function of type 4 whose F0 is 100 Hz, OQ is 50% and SQ is 100%, and the pulse in open phase is chosen from 360° of the first period of the sinusoid to 270° of the second period of the sinusoid.

4. MODAL VOICE SYNTHESIS

According to the model of the static glottis and the four types of glottal control functions, different pulses of phonation types can be synthesized by parameters of four F0s, four OQs, four SQs, two angles of sinusoid, two widths of the glottis and two lengths of the glottis.

4.1 Modal Voice Synthesis

Modal voice is a kind of phonation type which is most commonly used in speech. Usually, the F0 is around 70 and 300, the SQ is around 100 and 300% and the OQ is around 50 and 60%. See Table 1.

Table 1 The synthesis parameters of the modal voice

Semi axis	F0	OQ	SQ	Angle 1	Angle 2	width/length
Left	100	50	300	360	180	1.3
Right	100	50	300	360	180	1
Anterior	100	50	300	360	180	8
Posterior	100	50	300	360	180	4

In Table 1, the basic parameters for voice synthesis are listed. ‘Left’ stands for left width; ‘Right’ stands for right width; ‘Anterior’ stands for anterior length and ‘Posterior’ stands for posterior length. For

synthesizing a modal voice, the parameters of 100 (Hz), 50% , 300%, 360° and 180° for F0, OQ, SQ, angle 1 and angle 2 are set. The left width is 1.3mm, the right width is 1mm, the anterior length is 8mm and the posterior length is 4mm. The dynamic parameters and the synthesized dynamic glottal area are displayed as plot ‘a’ in Figure 20.

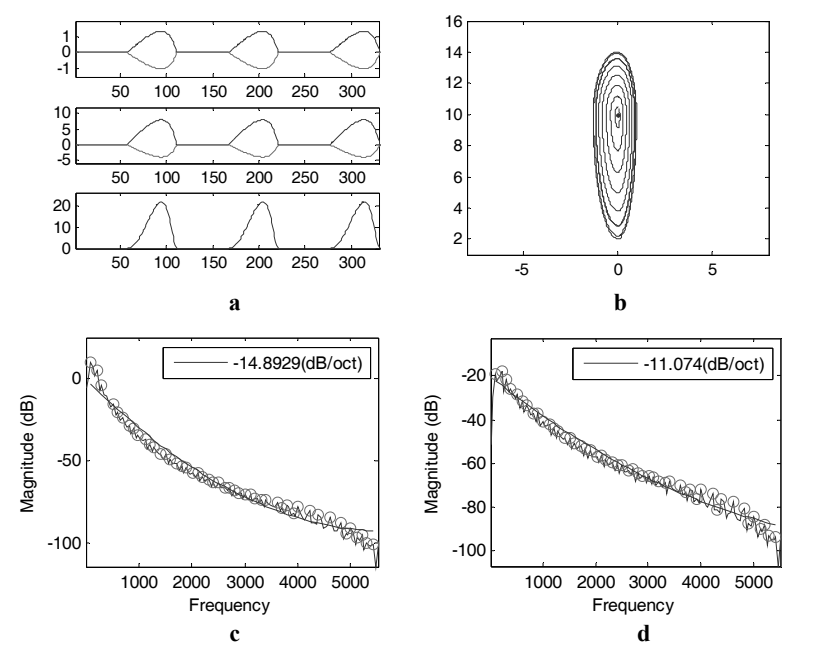


Figure 20 Parameters, synthesized glottal area and the spectra.

There are four plots in Figure 20. Plot ‘a’ displays the left width, right width, anterior length, posterior length and glottal area from top to bottom. Plot ‘b’ displays the dynamic glottises overlapped together. Plot ‘c’ displays the spectrum of the glottal area function, which is -14.8929 dB/oct. Plot ‘d’ displays the spectrum of the glottal area in differential form, which is -11.074dB/oct.

4.2 Modal Voice Synthesis with Different SQ

In the vibration of the human vocal folds, the left and right vocal folds often do not abduct and adduct at the same time in one vocal period. The basic parameters for synthesizing such a voice are listed in Table 2.

Table 2 The basic parameters of the modal voice with different SQs

Semi axis	F0	OQ	SQ	Angle 1	Angle 2	width/length
Left	100	50	300	360	180	1.5
Right	100	50	75	360	180	1.5
Anterior	100	50	300	360	180	8
Posterior	100	50	75	360	180	4

Table 2 lists the basic parameters of a modal voice. The SQs of the left and right widths and anterior and posterior lengths are not the same. The SQs for the dynamic glottal control function of the left widths and anterior length are 300%, and the SQs for the dynamic glottal control function of the right width and the posterior lengths are 75%. The angles for choosing the parts from the sinusoids are 360° and 180°. The widths of the left and right glottis are 1.5 mm, and the lengths of the anterior and posterior glottis are 8 mm and 4 mm, respectively.

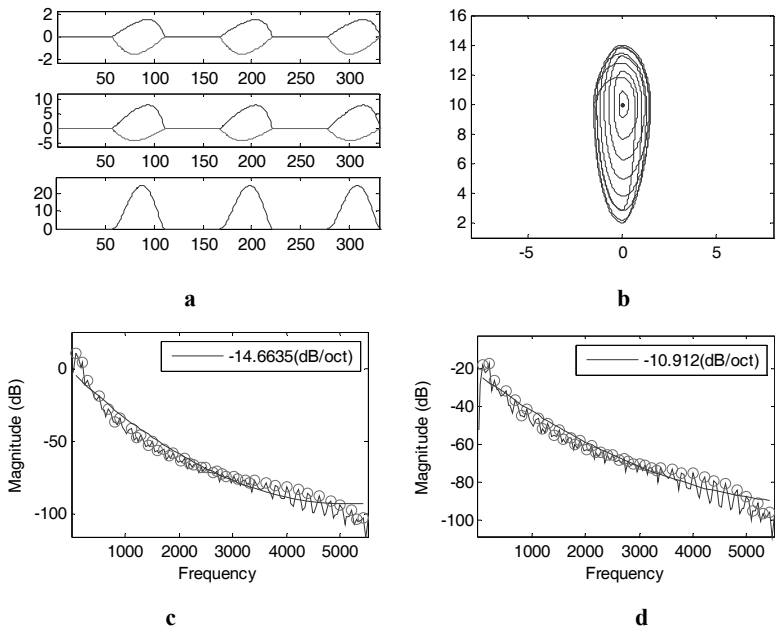


Figure 21 a.The synthesized parameters, b.the dynamic glottal areas, c.the two spectra of the glottal area, and d. the glottal area in differential form

There are four plots in Figure 21. Plot ‘a’ displays the left width, right width, anterior length, posterior length and glottal area from top to bottom. Plot ‘b’ displays the dynamic glottis overlapped together. Plot ‘c’ displays the spectrum of the glottal area function, which is -14.6635 dB/oct. Plot ‘d’ displays the spectrum of the glottal area in differential form, which is -10.912 dB/oct. From the parameters and synthesized glottal areas, we can find that the acoustical models mentioned above can’t synthesize such glottal pulses, which is more flexible than those synthesized by the acoustical models.

4.3 Falsetto Synthesis

Falsetto is a kind of voice with a very high pitch. Falsetto is not usually used in normal spoken language but is often used in oral performance, such as singing and opera. In Chinese oral cultures, such as Kunqu and Peking opera, falsetto is often used by Dan (young female) performers. In Table 3 the parameters for synthesizing a falsetto are listed.

Table 3 Falsetto synthesis parameters

Semi axis	F0	OQ	SQ	Angle 1	Angle 2	width/length
Left	400	100	100	315°	225°	0.7
Right	400	100	100	315°	225°	0.7
Anterior	400	100	100	315°	225°	7
Posterior	400	100	100	315°	225°	5

In Table 3, we can see that the F0 is 400 Hz, which is very high for a male speaker. The OQ is 100% which is the largest OQ in the vibration of the vocal folds and the SQ is 100% which is very small. The small SQ indicates the small power in high frequency. The part of the sinusoid was chosen from 315° of the first period of sinusoid to 225° of the second sinusoid. The widths of the left and right glottis are 0.7 mm and the lengths of the anterior and posterior glottis are 7 mm and 5 mm which mean that the glottis is very narrow and long. The synthesized dynamic control function, the glottal area function, the dynamic glottis, and the spectra of the glottal area function, and the differential form of the glottal area function are shown in Figure 22.

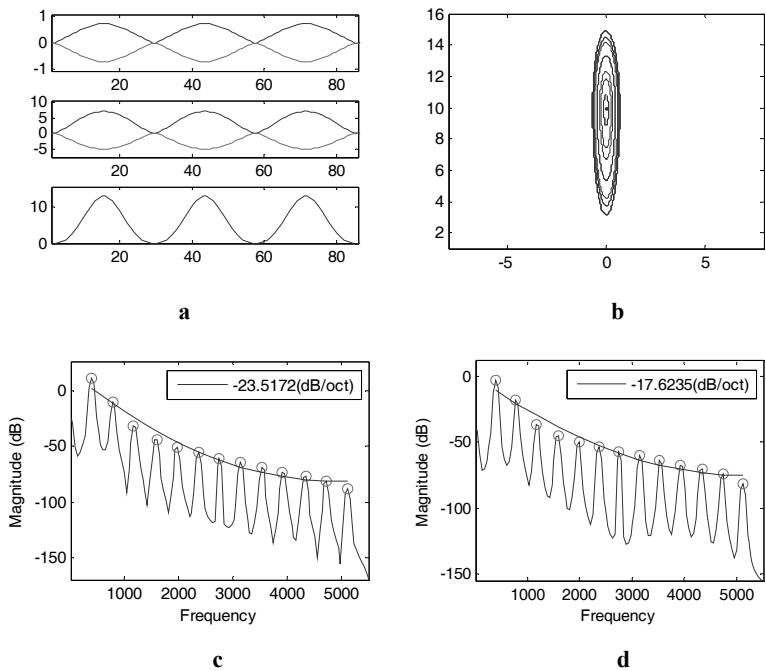


Figure 22 a. The synthesized dynamic glottal control function, b. the glottal area function, c. the dynamic glottis, and d. the spectrum of glottal area and the differential form of glottal area.

In Figure 22, we can see the synthesized glottal area function resembles the sinusoid very much, and the glottis looks narrow and long. The spectrum of glottal area function is -23.5172 dB/oct, which means that the power in high frequency is small, and the spectrum of differential glottal area function is -17.6235 dB/oct which means that the power is still small in the speech production of the human voice.

4.4 Vocal Fry Synthesis

Vocal fry is a kind of phonation with very low pitch and large power in high frequency. Vocal fry sometimes appears in the middle of the low tone (tone 3) in Mandarin. This phonation is very close to that of creaky voice which usually has irregular periods. The basic parameters for synthesizing a fry voice are listed in Table 4.

Table 4 Vocal fry synthesis parameters

Semi axis	F0	OQ	SQ	Angle 1	Angle 2	width/length
Left	40	15	300	360	180	1.1
Right	40	15	100	360	180	1.1
Anterior	40	15	300	360	180	1.3
Posterior	40	15	300	360	180	1.2

In Table 4, the F0 is 40 Hz which is very low for both male and female speakers. The OQ is 15%, which is very small and the SQ is 300% for left width, anterior length and posterior length, and 100% for right width of glottis. The angle 1 is 360° of the first period of the sinusoid and the angle 2 is 180° of the second period of the sinusoid. The left and right widths are 1.1mm, and the left and right lengths are 1.3 mm and 1.2 mm. The synthesized parameters and spectra are shown below in Figure 23.

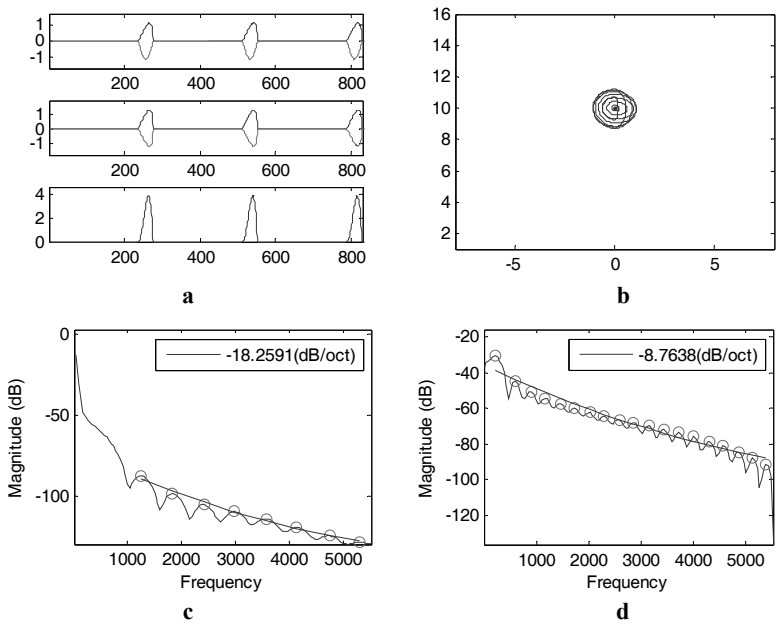


Figure 23 This figure displays a. the synthesized dynamic glottal control function, b. the glottal area function, c. the dynamic glottis, and d. the spectrum of glottal area and the differential form of glottal area of a vocal fry.

In Figure 23 (c and d), we can see the synthesized glottal area function looks like a saw waveform, and the glottis looks round and small (b). The spectrum of the glottal area function is -18.2591 dB/oct which indicates that the power in high frequency is not high, and the spectrum of differential glottal area function is -8.7638 dB/oct which means that the power is very large in the speech production of the human voice.

4.5 Synthesis of Diplophonia

Diplophonia is a kind of voice with different fundamental frequencies of left and right vocal folds. It is not a normal kind of voice in human speech but often appears in disordered voice. Sometimes people would produce diplophonia for singing performance. The basic parameters for synthesizing diplophonia are listed in Table 5. The synthesized parameters and spectra are shown in Figure 24.

Table 5 Basic parameters for synthesizing diplophonia

Semi axis	F0	OQ	SQ	Angle 1	Angle 2	width/length
Left	200	100	100	270	270	1.5
Right	180	100	100	270	270	1.5
Anterior	200	100	100	270	270	6
Posterior	180	100	100	270	270	4

In Table 5 the F0 of left glottis is 200 Hz and the F0 of right glottis is 180 Hz. These two F0s are not the same but differ by 20 Hz. The F0 of the anterior glottis is 200 Hz and the F0 of the posterior glottis is 180 Hz. These two F0s also have 20 Hz difference. The OQ is 100% which is very large, and SQ is 100% which is very small. The angle 1 is 360° of the first period of the sinusoid and the angle 2 is 180° of the second period of the sinusoid. The left and right widths are 1.5 mm, and the anterior and posterior lengths are 6 mm and 4mm. The synthesized parameters and spectra are shown in Figure 24.

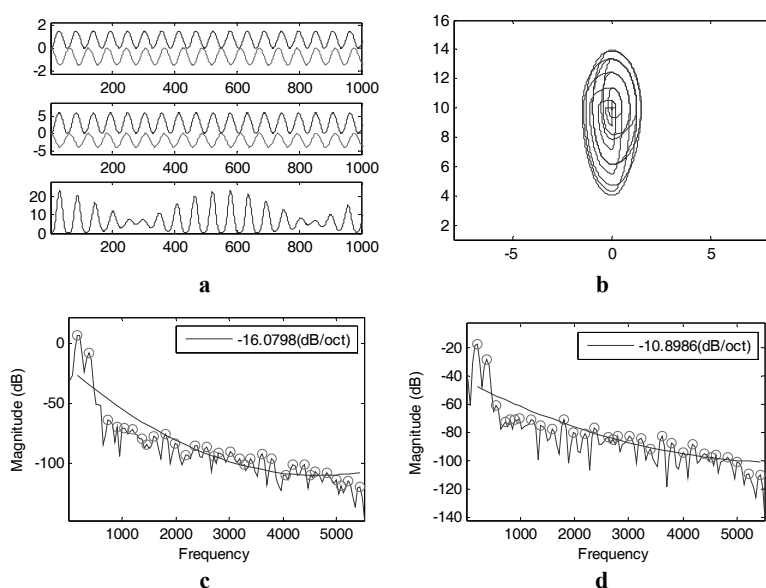


Figure 24 a. The synthesized dynamic glottal control function, b. the glottal area function, c. the dynamic glottis, and d. the spectrum of glottal area and the differential form of glottal area of a Diplophonia.

Figure 24 displays the dynamic glottal control functions, the dynamic glottis and the spectra of glottal area functions. In plot 'a'; we can see the F_0 s of the left and right widths of glottis are not the same and a super period which covers around 10 vocal periods. In plot 'b', we can see that the glottis is not symmetrical in vibration. The spectrum of glottal area function is -16.0798 dB/oct, and the spectrum of the differential glottal area function is -10.8986 dB/oct which indicates that the power is not small in the high frequency, though the SQs are small. This sample can't be synthesized by any acoustical models; this tells us that the glottal model is at a deeper level and so is more effective and flexible in source production, while the acoustical model is at the surface level.

5. CONCLUDING REMARKS

As is well known, voice models can be studied from the viewpoints of speech science, phonetics, speech engineering and so on, and developed for different purposes; an acoustical model may be used in speech synthesis; the physiological model may be used to imitate the physiological activity of vocal folds for medical purpose; a phonetic model which may be used to

study the linguistic significance of phonation types; and the singing model may be used in synthesizing special songs or in teaching singing. In this research, the dynamic glottal control functions were studied and improved on the basis of physiological model developed by Kong (2001) and published in 2007. The improvement in this study includes three kinds of basic parameters, namely, 1) the basic parameters of F0, OQ and SQ which are the most common parameters in the study of speech science and phonetics, and whose properties have close relationship with the human perception and linguistic significance; 2) the basic parameters of four types of the dynamic glottal control functions, which are chosen from parts of the sinusoid defined by two angles; 3) the basic parameters of the glottal size which are the widths of the left and right glottis, and the lengths of the anterior and posterior glottis. With this improved physiological model, different kinds of phonation types can be simulated and further studied for many purposes and in many fields.

Although the model can now be easily used to synthesize many different phonation types, many things can still be further studied and improved. First of all, since the sampling rate of high-speed image system is not high enough, the parameters extracted from the voice sample with high pitch are not very accurate, especially for the parameter of SQ and F0, because the position of peak in one period of the glottal area function is not stable. Secondly, the motion compensation and automatic contrast adjusting in the image processing system can also be improved. Thirdly, for the sake of sampling rate, it is still difficult to study the spectrum of the glottal area function and the relationship between the glottal area function and the other signals. We believe that along with the development of high-speed image system, good samples with a high sampling rate, maybe three-dimensional samples, can be captured for modeling a three-dimensional dynamic glottal model to simulate the vibration of vocal folds and synthesize different phonations with more accurate dynamic parameters.

NOTES

1. This research is funded by the National Natural Sciences Foundation of China (No: 61073085). We would like to give our thanks to Prof. Edwin Yiu in the University of Hong Kong for the high-speed image sampling, all the subjects, and Wang Gaowu for the program improvement of the high-speed image system.

REFERENCES

- ANANTHAPADMANABHA, T.V. 1984. Acoustic analysis of voice source dynamics. *Speech Transmission Laboratory – Quarterly Progress and Status Report* (STL-QPSR) 25(2-3):1-24, <http://www.speech.kth.se/qpsr>, accessed Feb.5, 2015.
- DIMITAR, D. Deliyski. 2005. Endoscope motion compensation for laryngeal high-speed video endoscopy. *Journal of Voice* 19(3):485-496
- FANT, G. 1979a. Glottal source and excitation analysis. *STL-QPSR*, 20(1):85-107, <http://www.speech.kth.se/qpsr>, accessed Feb. 5, 2015.
- _____. 1979b. Voice source analysis – a progress report. *STL-QPSR*, 20(3-4):31-53, <http://www.speech.kth.se/qpsr>, accessed Feb. 5, 2015.
- _____. 1980. Voice source dynamics. *STL-QPSR*, 21(2-3):17-37, <http://www.speech.kth.se/qpsr>, accessed Feb. 5, 2015.
- _____. 1982. The voice source, acoustic modeling. *STL-QPSR*, 23(4):28-48, <http://www.speech.kth.se/qpsr>, accessed Feb. 5, 2015.
- FANT G., J. Liljencrants, and Q. Lin. 1985. A four parameter model of glottal flow. *STL-QPSR*, 26(4):1-13, <http://www.speech.kth.se/qpsr>, accessed Feb. 5, 2015.
- FANT, G. and Q. Lin. 1988. Frequency domain interpretation and derivation of glottal flow parameters. *STL-QPSR*, 29(2-3):1-21, <http://www.speech.kth.se/qpsr>, accessed Feb. 5, 2015.
- FLANAGAN, J.L. and L.L. Landgraf. 1968. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics* 16(1):57-64
- HEDELIN P. 1984. A glottal LPC-vocoder. Paper presented at IEEE International Conference on Acoustics, Speech, and Signal Processing, San Diego. Proceedings 9:21-24.
- KONG, Jiangping. 2001. Study on dynamic glottis: Though high-speed digital imaging. PhD. diss, City University of Hong Kong.
- _____. 2007. *Laryngeal Dynamics and Physiological Models*, Peking University Press.
- LUCERO, J. C. 1993. Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region. *Journal of the Acoustical Society of America* 94(6):3104-3111.

- _____. 1996. Chest- and falsetto-like oscillations in a two-mass model of the vocal folds. *Journal of the Acoustical Society of America* 100(5):3355-3359.
- LJUNGQVIST M. and H. Fujisaki. 1985. A comparative study of glottal waveform models. *Technical Report of the Institute of Electronics and Communications Engineers* (Japan). EA85-58:23-9.
- PELORSON, X., A. Hirochberg, R.R. van Hassel, A.P.J. Wijnands, and Y. Aurégan. 1994. Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *Journal of the Acoustical Society of America* 96(6):3416-3431.
- ROSENBERG, A.E. 1971. Effect of glottal pulse shape on the quality of natural vowels. *Journal of the Acoustical Society of America*, 49(2B):583-590.

基於高速數位成像的動態聲門模型

孔江平*

北京大學

提要

本文利用高速數位成像技術對動態聲門模型進行了研究。眾所周知，言語產生包括噪音聲源、聲道共鳴和唇輻射三個方面，其中噪音聲源尤其重要，因為噪音聲源是言語產生的基礎。在言語產生的研究中，聲學模型已經有了很深入的研究，但由於聲帶振動難於觀察和採集樣本，噪音的生理模型研究的很少。多年前作者建立了一個動態聲門模型（Kong 2007），在此模型中，靜態聲門是用四個四分之一橢圓來建模的，並有正常、漏氣和敞開四種模式。模型的動態聲門控制函數是通過正弦和拋物線的乘積來建模。雖然這種方式有效，但合成噪音的參數解釋性較差。在本項研究中，採集了更多更高品質的聲帶振動高速數位成像樣本和大大改進了數位影像處理的技術，最終模型的動態聲門控制函數所用的參數對噪音聲源的感知具有很好的解釋性。

主題詞

高速數位成像 聲帶振動 動態聲門模型

* 孔江平為北京大學教授，執教研究于中文系，中國語言學研究中心，語言與人類複雜系統聯合研究中心；地址：中國北京，郵編 100871，北京大學中文系；[jpkong@pku.edu.cn]

Vocal Attack Time of Different Pitch Levels and Vowels in Mandarin

*Ruifeng Zhang, †,‡R. J. Baken, and *Jiangping Kong, *Beijing and ‡Hong Kong SAR, China, and †Valhalla, New York

Summary: The purpose of this study was to investigate how vocal attack time (VAT) varies when young adults articulate the three vertex vowels in Mandarin Chinese at five linguistically unconstrained pitch levels. Sound pressure and electroglottographic signals were recorded simultaneously from 53 male and 53 female subjects saying sustained /A/, /i/, and /u/ at five equally spaced pitch heights, each being higher than the preceding one. Then analyses of means, variance, and correlation were performed to explore the relationships of VAT/pitch levels and VAT/vowels. Findings were As mean STs (semitone) increase linearly from levels 1 to 5, mean VATs decrease nonlinearly in a big group of subjects but increase nonlinearly in a small group of them. Based on the body-cover model of F_0 control, data here lead to the guess that different people incline to use different strategies in increasing pitch height. When males, females, and males plus females are considered as a whole, average STs and VATs tend to be positively correlated among the three vertex vowels.

Key Words: Vocal attack time–Pitch levels–Vertex vowels–Semitone.

INTRODUCTION

Vocal attack time

Vocal attack time (VAT) is the time lag between the rise of the simultaneously recorded sound pressure (SP) and electroglottographic (EGG) signals, measured at the onset of phonation.¹ When airflow goes through the glottis during the initiation of a vowel or a voiced consonant, the vocal folds oscillate with very small amplitudes before their first contact is achieved and stabilized. Therefore, the SP signal, which is a record of SP emitted from the mouth, begins its growth of amplitude well before the vocal folds touch each other, but the EGG signal, as a record of vocal-fold contact area, has nearly no amplitude until the vocal fold contact occurs, and only after that does its magnitude show up and grow. This is the occasion when VAT values are positive. However, in other cases when the initiation of EGG signals leads that of SP signals, namely, when vocal fold contact precedes the appearance of SP signals such as in a hard glottal attack, VAT is negative. If the two sorts of signals rise at the same point of time, VAT equals zero. Consequently, VAT can be taken to be the duration from the start of vocal cord oscillation to the instant for the first vocal cord contact and provides a useful index to indicate the prephonation laryngeal adjustment.

The effectiveness of VAT measurement was experimentally verified by Orlikoff et al.¹ using five vocally normal subjects. EGG and SP signals of different phonation types were recorded synchronously with high-speed video endoscopy, from which a digital kymogram (DKG) was generated. The DKG attack duration data obtained by hand were then compared with the VAT

measures extracted by using computer programs. The strong and direct relationship between the VAT- and DKG-measured data proved VAT to be a valid and convenient measure of vocal attack. In 2012, a figure of merit (FOM), which assesses a critical assumption of vocal startup on which the VAT measure is based and therefore represents integrity of the derived measure, was proposed by Roark et al.² for the VAT measurement of sustained /a/. SP and EGG signals from 102 tokens were visually inspected to empirically derive a criterion level of FOM less than 0.75 to indicate when the assumption underlying the measurement had failed and the VAT value obtained should be disregarded. The example of using VAT for nonlinguistic research was the measurement by Roark et al.³ acquiring normative data of VAT in healthy young adults. They collected SP and EGG signals from 55 males and 57 females performing multiple tokens of three tasks (sustained /a/, “always,” and “hallways”) at comfortable pitch and loudness. The average VATs were significantly shorter for females than for males and mean VAT was 1.98 milliseconds in the screened sample of normal young speakers. The use of VAT in linguistic research was exemplified by the measurements done by Ma et al.⁴ examining the association between VAT and tone in Cantonese speakers.

Pitch levels and vowels

It is well known that pitch increases along with the acceleration of vocal fold oscillation, resulting from step-by-step augmentation of the vocal folds' tension. The VAT study of three phonation types by Orlikoff et al.¹ seems to suggest that tenser vocal folds tend to be associated with smaller VAT values. Therefore, how VAT varies with increasing pitch in Mandarin Chinese appears to be an attractive research subject that has never been touched upon. We required the subjects to produce vowels at five different linguistically unconstrained pitch levels out of two considerations. For the purpose of devising tone letters, Chao⁵ divided the pitch range of a person into four equal parts with five points numbered 1, 2, 3, 4, and 5, corresponding to low, half-low, medium, half-high, and high, respectively. It has also been found by subsequent linguistic researchers that no language uses over five pitch levels to distinguish its tones.⁶ On the other hand, many subjects felt it natural and easy to

Accepted for publication July 15, 2014.

This research was funded by the National Natural Sciences Foundation of China and the grant number was 61073085.

From the *Department of Chinese Language and Literature, Research Center for Chinese Linguistics and Joint Center for Language and Human Complexity, Peking University, Beijing, China; †Department of Speech-Language Pathology, New York Medical College, Valhalla, New York; and the ‡Hong Kong University, Hong Kong SAR, China.

Address correspondence and reprint requests to Jiangping Kong, Department of Chinese Language and Literature, Peking University, Beijing 100871, China. E-mail: kongjp@gmail.com

Journal of Voice, Vol. 29, No. 5, pp. 542-547

0892-1997/\$36.00

© 2015 The Voice Foundation

<http://dx.doi.org/10.1016/j.jvoice.2014.07.010>

TABLE 1.
Maximum, Minimum, Mean, and Standard Deviation of ST and VAT Across Pitch Levels in Groups A: Males, B: Females, and C: Total

Groups	A: Males (n = 1458)				B: Females (n = 1369)				C: Total (n = 2827)			
Pitch Levels	Max	Min	Mean	SD	Max	Min	Mean	SD	Max	Min	Mean	SD
1												
Pitch (ST)	18.87	3.09	11.46	3.09	26.40	16.45	20.75	1.93	26.40	3.09	15.96	5.32
VAT (ms)	39.27	-21.11	2.69	8.64	53.13	-46.24	3.33	12.26	53.13	-46.24	3.00	10.54
2												
Pitch (ST)	21.37	5.19	14.25	3.20	29.37	17.97	23.46	1.94	29.37	5.19	18.74	5.32
VAT (ms)	22.83	-22.70	0.59	7.35	27.73	-36.74	-0.27	8.82	27.73	-36.74	0.17	8.10
3												
Pitch (ST)	23.03	7.91	16.15	3.32	31.44	19.55	25.22	2.10	31.44	7.91	20.53	5.33
VAT (ms)	20.45	-21.72	0.39	6.56	31.93	-29.37	0.07	8.18	31.93	-29.37	0.24	7.38
4												
Pitch (ST)	23.40	8.73	17.70	3.24	34.07	21.02	26.94	2.49	34.07	8.73	22.21	5.45
VAT (ms)	20.11	-21.22	0.59	7.00	26.44	-23.61	0.95	7.69	26.44	-23.61	0.77	7.34
5												
Pitch (ST)	26.06	11.01	19.11	3.26	35.31	23.26	28.66	2.36	35.31	11.01	23.68	5.57
VAT (ms)	15.85	-25.67	0.27	7.36	27.57	-56.26	-1.20	10.98	27.57	-56.26	-0.43	9.29

space five pitch levels equally within their voice range, but difficult to manage six or more levels that way. So the present study, by focusing on five sustained pitch heights that are not linguistically distinctive, is intended to be conducive to future work on language tones. The three vertex vowels /A/. (Since /A/ in Mandarin is the lowest central vowel and different from the front /a/ and back /a/ on the IPA (International Phonetic Association) chart, Chinese linguists prefer to represent it with a capital letter.) /i/ and /u/ in Mandarin Chinese were chosen to be produced at five pitch levels because they occupy the utmost points on the vowel chart and represent the entire scope of tongue movement during articulation. All in all, the purpose of this study was to explore how VAT varies when young people produce the three vertex vowels in Mandarin Chinese at five linguistically unconstrained pitch heights.

METHOD

Subjects and instrumentation

Fifty-three females (aged 18–22 years) and 53 males (aged 18–22 years), all of whom were college students, participated in the research. They spoke standard Mandarin Chinese for daily communication, had no voice or hearing problems, and were all in good health at the time of recording. The recording was accomplished in the sound-treated booth at the Language Laboratory of the Chinese Department, Beijing University, where the background noise was below 25dBA. The *Adobe Audition 2.0* on the computer (Lenovo, x220i, Lenovo, Beijing, China) was set at the stereo interface with a sampling rate of 44 100 Hz and a resolution of 16 bits for each channel. The electroglottograph (Model 6103) used for collecting EGG signals, and the microphone and sound card (Creative Labs Model No.sb1095) used to gain SP signals were synchronously connected to the computer through a sound console (Behringer XENYX502, MUSIC Group Macao Commercial Offshore

Limited, Zhongshan, Guangdong, China). With their lips about 10 cm away from the microphone, the subjects were asked to say sustained /A/, /i/, and /u/ at five pitch heights, each of them being higher than the preceding one. All pitch levels were repeated twice and 30 tokens ($3 \times 5 \times 2$) were obtained from each speaker.

Parameter extraction

F₀, VAT, and FOM measures were extracted largely automatically from the speech samples using the software developed by Roark et al,³ which processed signals in four stages: signal verification, signal segmentation, F₀-based frequency filtering and signal modeling, and extraction of measures. From the 3180 samples (1590 for males and 1590 for females), 3165 values (1590 for males and 1575 for females) were obtained for each of the three parameters, with 15 female speech recordings unable to be evaluated by the software, possibly due to the poor quality of their EGG signals.

Data preprocessing

Because our research required the subjects to pronounce each of the three vowels with increasing pitch heights, the F₀ values they produced for each vowel should theoretically increase with the level shift from 1 to 5. Consequently, a correlation analysis was first done that discarded 210 ineffective speech samples (140 for females and 70 for males) whose pitch values had a negative correlation with the pitch level numbers. The 2955 measures left were then divided into 10 groups: male-level 1, male-level 2, male-level 3, male-level 4, male-level 5, female-level 1, female-level 2, female-level 3, female-level 4, and female-level 5. Each group was processed separately in the same way: because vocal fold vibrations normally do not go beyond 500 Hz, measures that were beyond ± 3 standard deviation (SD) from the mean F₀ were first removed from each group; and second, according to the observation that there

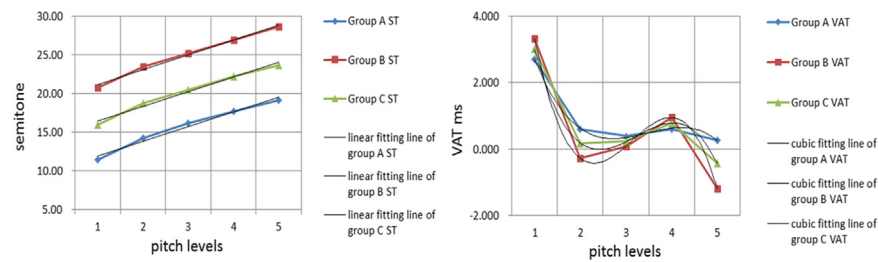


FIGURE 1. Average pitch in semitones and mean VAT in milliseconds as a function of pitch levels for groups A: males ($n = 1458$), B: females ($n = 1369$), and C: total ($n = 2827$).

were more outliers among VAT values, measures that were beyond ± 2 SD from the mean VAT were eliminated. Among the 2827 measures that remained eventually, F_0 ranged from 77.3 Hz to 497.06 Hz, the mean (SD) being 220.35 Hz (75.26 Hz), and VAT from -56.26 ms to 53.13 ms, the mean (SD) being 0.75 ms (8.69 ms). For the *Excel* (Excel 2010, Microsoft) and *SPSS* (SPSS13.0, SPSS, Inc) analyses in the following, all F_0 values were converted to semitones (ST) re 64.66 Hz. This reference level was chosen not only because it was close to the minimum pitch value 77.3 Hz but also because Liu⁷ had been using it in his groundbreaking research of Mandarin tones.

RESULTS

VAT and pitch levels

Among the 1458 male speech samples, pitch ranges from 3.09 ST to 26.06 ST (mean = 15.72 ST; SD = 4.19) and VAT from -25.67 ms to 39.27 ms, (mean = 0.91 ms; SD = 7.46). Among the 1369 female speech samples, pitch ranges from 16.51 ST to 35.31 ST (mean = 24.98 ST; SD = 3.49) and VAT from -56.26 ms to 53.13 ms (mean = 0.59 ms; SD = 9.84 ms). One-way analyses of variance and *post hoc* tests have shown that, at significance level $P = 0.01$ for both males and females, ST values are significantly different between any two of the five pitch levels ($P < 0.01$ for all), exceeding a preselected $\alpha = .01$, whereas for VAT, significant differences are only seen between pitch level 1 and each of these levels: 2, 3, 4, and 5 ($P < 0.01$ for all). According to a correlation analysis, ST has a significant negative correlation with VAT among all the 2827 speech samples ($n = 2827$, $r = -0.077$, $P < 0.01$). When male and female measures are calculated separately, the significant negative cor-

relation still holds for both but with a higher strength of correlation for females than for males ($n = 1458$, $r = -0.084$, $P < 0.01$ for males; $n = 1369$, $r = -0.115$, $P < 0.01$ for females).

In Table 1 are listed the maximum, minimum, mean and SD of ST and VAT across pitch levels calculated according to this grouping: A: males (1458 tokens); B: females (1369 tokens); and C: total (2827 tokens). Here, vowel considerations and individual differences are temporarily left aside (Table 1). As pitch levels shift from 1 to 5, all three groups show a linear increase of pitch, but a nonlinear and nonmonotonic decrease of VAT: unlike ST means, each mean value of VAT from levels 2 to 5 is not always larger than the one that follows; however, in all these cases, the average VAT at level 1 is always the largest among the five pitch levels and is much larger than the mean VATs at levels 2, 3, 4, and 5, making all five going on a downward trend. Mean STs and VATs as a function of pitch levels for the three groups are indicated in Figure 1, from which VAT variations, compared with ST changes, can be seen more intuitively: from level 1 to levels 2 and 3, there is a deep declination, and then a clear upturn at level 4, followed by a steep dip at level 5 to the minimum mean VAT value.

Now consider how individuals uttered five pitch levels differently. Because each two repetitions of a pitch level were produced consecutively (eg, /A/ level 1 $\rightarrow$ /A/ level 1 $\rightarrow$ /A/ level 2 $\rightarrow$ /A/ level 2 $\rightarrow$ /A/ level 3 $\rightarrow$ /A/ level 3 $\rightarrow$ /A/ level 4 $\rightarrow$ /A/ level 4 $\rightarrow$ /A/ level 5 $\rightarrow$ /A/ level 5, and so forth), and the pitch values of them were nearly the same, they can and have to be averaged if we want to see intersubject differences associated with ST-VAT correlation while ignoring the slight differences between tokens repeated by the same person. The findings after that done are displayed in Figure 2. Among the 103 subjects (52

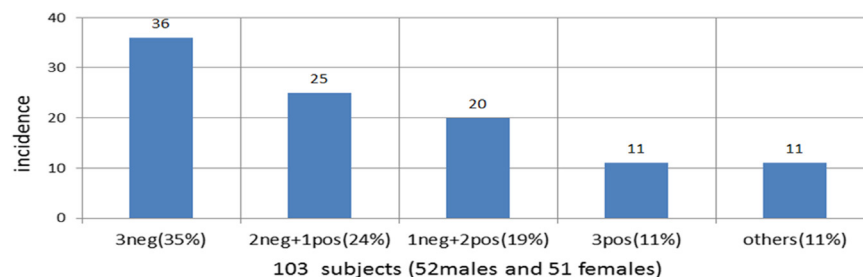


FIGURE 2. The 103 subjects are divided into four categories depending on how many of their three vowels were uttered with negative VAT-ST correlation coefficients and how many with positive ones. 3neg means all three vowels were produced with negative coefficients and 3pos means all three vowels with positive ones. 2neg + 1pos and 1neg + 2pos can be understood by analogy.

TABLE 2.
Means and Standard Deviations of ST and VAT Across Five Pitch Levels in the Two Different Categories of Subjects

Pitch Level	36 Subjects Who Uttered All 3 Vowels at 5 Pitch Levels With Negative VAT-ST Correlation Coefficients				11 Subjects Who Uttered All 3 Vowels at 5 Pitch Levels With Positive VAT-ST Correlation Coefficients			
	Mean ST	SD of ST	Mean VAT	SD of VAT	Mean ST	SD of ST	Mean VAT	SD of VAT
1	16.57	5.04	6.29	9.57	16.55	3.63	−1.55	6.03
2	19.05	4.93	0.68	6.63	19.54	3.66	−1.55	5.66
3	20.93	5.02	−0.71	7.33	21.19	3.63	1.29	4.60
4	22.57	5.23	−0.74	7.28	22.43	3.50	3.01	4.05
5	24.25	5.52	−2.93	10.29	24.42	3.39	3.32	4.56

males and 51 females) whose data were retained for analyses, 36 (35%) produced all /A/, /i/, and /u/ with negative VAT-ST correlation coefficients, 25 (24%) uttered two of the three vowels with negative VAT-ST correlation coefficients and one with positive ones, 20 (19%) pronounced one vowel with VAT-ST negatively correlated and two with them positively correlated; 11 (11%) articulated all three vowels with positive VAT-ST correlation coefficients. The remaining 11 subjects (11%) cannot enter any of these categories because the measures of one or two of their vowels were culled out during data preprocessing. In summary, many subjects produced increasing pitch heights with VAT declining, but a small number of them did it with VAT increasing, and of course, there were also subjects belonging to the intermediate type.

In Table 2 are listed the means and SD of ST and VAT across five pitch levels of the 36 subjects who uttered all three vowels with negative VAT-ST correlation coefficients and the 11 who did all three with positive ones. Their mean STs and VATs as a function of pitch levels are indicated in Figure 3. The VAT-ST covariation of the former can be seen in Figure 3A: as pitch levels shift linearly from 1 to 5, the mean VATs drop gradually except that there is a slight upturn on pitch level 4. But that of the latter shows a quite different picture: as pitch levels go linearly from 1 to 5, average VATs increase gradually except for the small turns at pitch levels 2 and 4 (Figure 3B).

VAT and vowels

To have a picture of VAT variation among different vowels while leaving pitch level considerations aside, mean VATs and STs of /A/, /i/, and /u/ are calculated for groups A: males

(1458 tokens), B: females (1369 tokens), and C: total (2827 tokens). The resultant means of the three groups are listed in Table 3 and displayed in Figure 4, from which it is seen that the average STs of /A/, /i/, and /u/ in groups A, B, and C are all thus ordered: /u/ > /i/ > /A/ although they are only slightly different between one another. The mean VATs of these three groups are all thus patterned also: /u/ > /i/ > /A/, with the average VATs of /u/ being the largest among the three vowels and much larger than those of the other two vowels. High vowels tend to have both larger pitch values and longer VATs than low vowels.

DISCUSSION

All the analyses concerning VAT and pitch levels point to the same finding. As is indicated in Figure 1, when all males, females, or males plus females are considered as a group, a nonlinear contravariant relationship can be seen between the average values of VAT and ST at five pitch levels, because the best trend line of the increasing pitch is linear while that of the decreasing VAT is cubic. When individuals are taken into consideration, a slightly different picture appears: Many people tend to produce the three vertex vowels (/A/, /i/, and /u/) with negative VAT-ST correlation coefficients, but a small number of subjects incline to produce them with positive ones (Figures 2 and 3). However, the nonlinear relation still holds in both cases because although Figure 3A presents a contravariant VAT-ST relationship but Figure 3B shows an orthokinetic one, the best fit lines of the increasing pitches in both graphs are linear while those of the decreasing or increasing VAT are still cubic. The author listened and compared the subjects

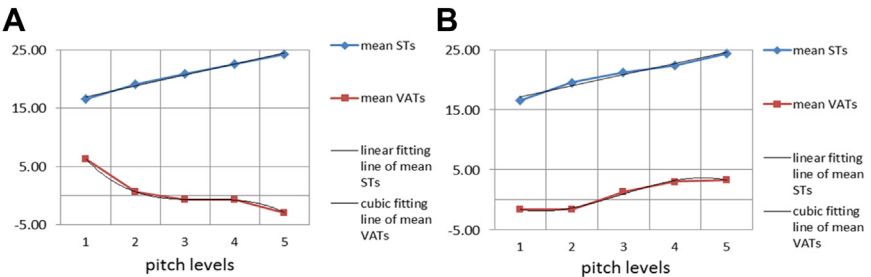


FIGURE 3. Mean STs and VATs across five pitch levels of (A) the 36 subjects who uttered all three vowels with negative VAT-ST correlation coefficients, (B) the 11 subjects who did all three with positive VAT-ST correlation coefficients.

TABLE 3.
Means of ST and VAT Across the Three Vowels in Groups A: Males, B: Females, and C: Total

Groups	A: Males		B: Females		C: Total	
Number	n = 1458		n = 1369		n = 2827	
Vowels	Mean ST	Mean VAT	Mean ST	Mean VAT	Mean ST	Mean VAT
/A/	15.41	0.07	24.87	-1.39	19.87	-0.62
/i/	15.79	0.87	24.97	-0.15	20.32	0.37
/u/	15.97	1.78	25.08	3.13	20.41	2.44

who pronounced five pitch heights of a vowel with a significantly negative VAT-ST correlation and those who did it with a significantly positive one and his impression was that the former sounded much more laborious than the latter at pitch level five.

Watson et al⁸ reported that adjusted mean VAT for their high frequency condition was smaller than the adjusted mean VAT for their mid and low frequency conditions and that VAT appears to be sensitive to increases in vocal fold tension in normal speakers. The present findings not only accord with their statements but also suggests different strategies people tend to use in increasing pitch height. The body cover model of F_0 control proposed by Titze⁹ concerns the activities of cricothyroid (CT) and thyroarytenoid (TA) muscles. The contraction of CT elongates the vocal folds but that of TA tends to shorten them. In combination, the two muscles are responsible for most of the length change that can be achieved. TA muscle activity is also used to regulate the effective depth of vocal fold vibration, which reduces rapidly as pitch increases. At low to intermediate F_0 and loud productions, namely, in modal voice conditions, CT and TA activities are both relatively low and a significant portion of the vocal fold body is in vibration while the mucosa and ligament remain somewhat lax. A rise in F_0 is generally obtained by increased TA activity, as long as CT activity is not near its maximum. As pitch boosts from high to falsetto or a high F_0 is to be achieved (as in high-pitched singing), CT activity gradually becomes dominant, whereas the TA action is decreased. Only the surface of the vocal fold vibrates, with a stiff ligament and a loose mucosa in combination.⁹ Some guesses can be made based on this model of pitch control and the data reported here. Many people are not good at using falsetto, and, when required to utter vowels at five increasingly higher pitch levels, tend to start at the very low point and go step by step through modal voice register with the vocal fold tension being increased gradually. But quite a few speakers, untrained

but skillful in using voices, prefer to start at a relatively higher point and gradually get somewhere around falsetto where the vocal folds, on the contrary, turn somewhat slack. This might be the reasons why the former displayed a negative VAT-ST correlation but the latter a positive one.

It was reported by Zhang¹⁰ and Dong¹¹ that the three Chinese vertex vowels, while pronounced comfortably, display their intrinsic F_0 in such a pattern: /u/ > /i/ > /A/. The results here support their finding. Figure 4 indicates that both mean STs and mean VATs of /A/, /i/, and /u/ in men, women, and all subjects combined are all ordered: /u/ > /i/ > /A/, suggesting that VAT of the three vowels normally tends to be positively correlated with their intrinsic pitch. But what causes such a relationship needs to be further explored.

CONCLUSION

The purpose of this study was to investigate how VAT varies when young adults articulate the three vertex vowels in Mandarin Chinese at five linguistically unconstrained pitch levels. In a large group of young adults, it was found that pitch and VAT change in opposite directions, but in some individuals, the two vary in the same direction. But in all cases, pitch and VAT tend to present a nonlinear relationship. The possibility is that people, out of habit or by reason of laryngeal physiology, tend to manipulate their vocal folds in two different ways when they increase pitch from low to high levels. Those who utter vowels at increasing pitch levels with positive VAT-ST correlation coefficients may have more potential in using falsetto, which of course needs to be further investigated. After this research, our study on VAT of the lexical tones in Mandarin Chinese is now under way.

Acknowledgments

The authors thank all the voice experts for their gentle participation of the investigation.

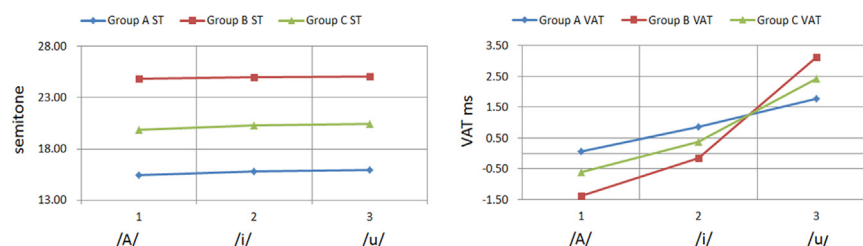


FIGURE 4. Average pitch in semitones and mean VAT in milliseconds as a function of vowels for groups A: males, B: females, and C: total.

REFERENCES

1. Orlikoff RF, Deliyski DD, Baken RJ, Watson BC. Validation of a glottographic measure of vocal attack. *J Voice*. 2009;23:164–168.
2. Roark RM, Watson BC, Baken RJ. A figure of merit for vocal attack time measurement. *J Voice*. 2012;26:8–11.
3. Roark RM, Watson BC, Baken RJ, Brown DJ, Thomas JM. Measures of vocal attack time for healthy young adults. *J Voice*. 2012;26:12–17.
4. Ma EP-M, Baken RJ, Roark RM, Li P-M. Effect of tones on vocal attack time in Cantonese speakers. *J Voice*. 2012;26:670.e1–670.e6.
5. Chao YR. A system of “tone-letters”. *方言*. 1980;2:81–83.
6. Maddieson I. *Universals of tone. Universals of human language*. Vol 2. Stanford, California, USA: Stanford University Press; 1978:338.
7. 刘复. 乙二声调推算尺. *史语所集刊*. 1934;4本4分:355-361.
8. Watson BC, Baken RJ, Roark RM, Reid S, Ribeiro M, Tsai W. Effect of fundamental frequency at voice onset on vocal attack time. *J Vitaminol*. 2013;27:273–277.
9. Titze IR. *Principles of Voice Production*. 2nd ed. Iowa City, IA, USA: National Center for Voice and Speech; 2000:211–242.
10. 张家驊. 元音的内在基频与讲话方式对共振峰的影响. *声学学报*. 1989;14:401–406.
11. 董倩倩. 汉语普通话元音音高再探. *文教资料*. 2010:27–28.

A STUDY ON THE FEATURES OF
CHEST AND ABDOMINAL BREATHING BETWEEN
RECITING AND CHANTING CHINESE POETRY

Feng Yang*

Jinan University, Guangzhou

ABSTRACT

This research studies the features of chest and abdominal breathing between reciting and chanting Chinese poems of different styles.¹ Eight participants were recruited to recite and chant 85 modern style poems (近體詩) and 39 Song poems (宋詞). The chest and abdominal breathing signals as well as speech signal were recorded simultaneously. Programs for breathing analysis have been written to extract parameters, such as breathing reset amplitude, time of inhalation phase, and slope of exhalation phase. The results show that the poem chanting has a larger depth of breathing and amount of breath, and more frequent exhalations compared with the poem reciting. In both poem reciting and chanting, the pause and declination of chest breathing is closely correlated with the prosodic boundaries. The major function of chest breathing in speaking is to keep the chest extended and provide enough breath for articulation and chest resonance. The function of abdominal breathing is to provide stable sub-glottal pressure through contraction of abdominal muscle and diaphragm, controlling continuous airflow to produce continuous sound.

SUBJECT KEYWORDS

Chest and abdominal breathing Poem reciting Poem chanting Speech production

* **Feng Yang** 楊鋒 is a lecturer in the College of Chinese Language and Culture, Center for Ecology of Asia-Pacific Languages and Cultures of Jinan University. Corresponding contact: College of Chinese Language and Culture, Jinan University, Guangzhou 510610, China; [yangzihuai@163.com]

1. INTRODUCTION

Breathing, as the driving force of speech, is part of speech production (Wang 2008). It plays an important role in human speech activities. The coordination of the muscular movements of the chest, diaphragm, and rectus abdominis has been proved to be closely related with the prosodic organization of human speech (Hixon 1987). One of the important functions of the breathing system is that it produces sub-glottal pressure for continuous sound flow, and this pressure is changing at the prosodic boundaries (Ohala 1990). The breathing pattern in speaking can be analyzed into smaller units, which were considered as the basic unit of intonation production and perception (Lieberman 1967). Ladefoged and Loeb (2002) believe that the production of stressed syllables is the result of the excessive respiratory power that is initiated by the internal intercostals.

Wu and Lin (1989) carried out research on Chinese with signals of airflow and pressure for the first time in China. Tan et al (2008) and Shi et al (2010) have studied abdominal breathing reset of Mandarin reciting of various writing styles, and made statistical analysis of different types of breathing reset. Yang et al (2013) explored mechanisms of chest and abdominal breathing.

The present research compares breathing features of reciting and chanting of Chinese traditional poems through analyzing simultaneous signals of chest breathing, abdominal breathing and speech sound. The effects of chest breathing and abdominal breathing and their interrelationship are also investigated.

2. METHOD

2.1 Data Sampling

Eighty-five five-syllable or seven-syllable modern style poems and 39 Song poems performed by eight professional scholars were sampled in three styles, namely Standard Chinese, dialect and chanting in dialect. All the poems were recorded in the laboratory of Peking University. The major equipment for recording was the 16-channel Powerlab PL3516. This experiment collected signals of 4 channels using Chart 5. Channel 1 is speech signal collected through microphone and console mixer; Channel 2 is phonation signal collected through Electrolottography; Channel 3 is

pectoral breathing signal collected by Breathing strip, MLT1132 sensor; Channel 4 is abdominal breathing signal also collected through Breathing strip, MLT1132 sensor. All sample rates are 44kHz. The breathing strip used for measuring chest breathing is fastened 10cm below the armpit, while breathing strip used for measuring abdominal breathing is fastened above the navel. Devices used in recording are: microphone, Sony ECM-44B; console mixer, Behringer XENYX502; Electroglottography, Kay 6103; breathing strips, ADInstrument MLT1132.

2.2 Parameter Extraction

In order to study simultaneously the signals of speech, phonation, chest breathing, and abdominal breathing, a data analysis platform was built in Matlab. As shown in Fig 1.

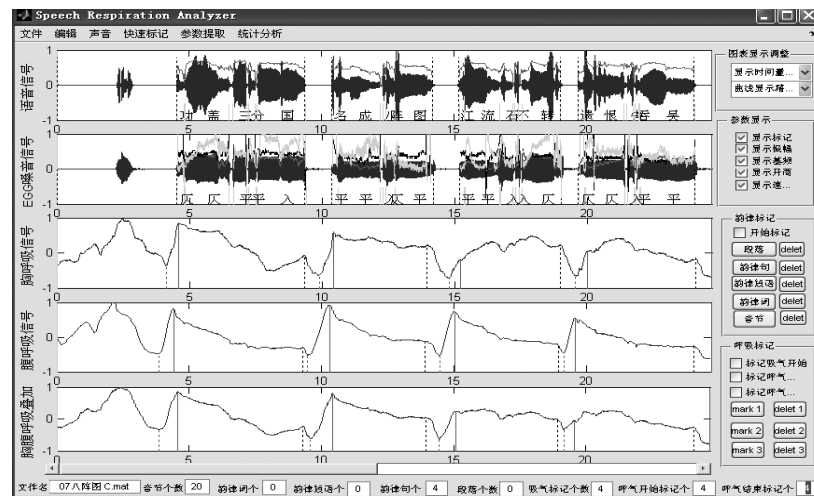


Figure 1 The interface of the data analyzing platform

In the interface of the data analyzing platform, there are five sub-windows in which five signals are displayed. The five signals, from top to bottom, are: 1) speech signal; 2) EGG signal; 3) chest breathing signal; 4) abdominal breathing; and 5) the sum of chest and abdominal breathing. On the right of the interface, there are four function plots and in each of them, there are some functional buttons for text labeling and signal processing.

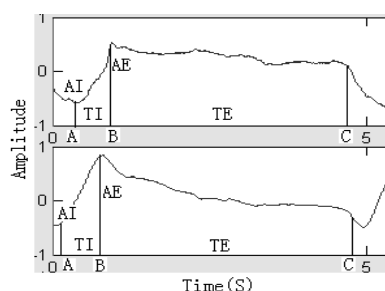


Figure 2 Definition of breathing parameters (from top to bottom: chest breathing, abdominal breathing)

The definition of breathing parameters is shown in Figure 2. The horizontal axis is time measured in seconds, while the vertical axis is the amplitude of breathing. Above is the signal of chest breathing and below is the abdominal breathing. The rising curve indicates inhalation and the falling curve exhalation. A, B, and C are inhalation onset time, breathing onset time, and exhalation end time respectively. The duration of inhalation $TI=B-A$, duration of exhalation $TE=C-B$. AI = inhalation reset amplitude, AE = exhalation reset amplitude. Other parameters, like the slope of inhalation period, slope of exhalation period, area of inhalation phase, and area of breathing phase, can be measured with those parameters mentioned above.

First, the signal is low-pass filtered to eliminate high-frequency interference. Then, the signal is normalized to facilitate inter-person comparison. The formula of normalization is:

$$\text{BreathSignal} = \text{OriginalSignal} / \max(\text{abs}(\text{OriginalSignal})) * 0.8$$

The prosodic boundaries of each layer are marked manually. The inhalation onset time, breathing onset time, and exhalation end time of each period of chest and abdominal breathing are marked automatically. Other parameters are calculated by the analyzing platform. The parameters are saved as .mat format to produce breathing parameters in Microsoft Excel for statistical analysis.

3. RESULTS

3.1 The Difference of the Respiratory Patterns between Reciting and Chanting Modern-style Poems

Chanting poems is a traditional way to read out poems with prolonged sounds and characteristic melody. It draws on the feature of reciting and singing. When chanting modern-style, chanters apply a respiratory pattern that has a much larger respiratory magnitude and breathing volume and the protracted exhalation phase in each respiratory cycle.

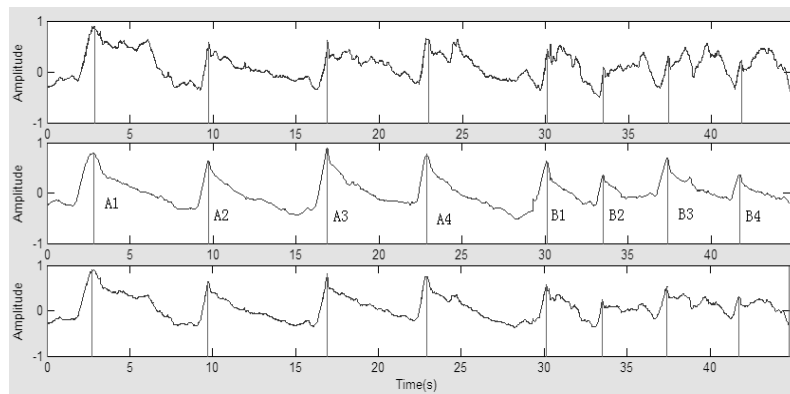


Figure 3 Comparison of chanting and reciting in chest and abdominal breathing (top: chest breathing, bottom: abdominal breathing, left: chanting, right: reciting)

Figure 3 shows examples of chest and abdominal breathing cycles in chanting and reciting a seven-syllable modern-style poem entitled “Jin Ling Yi” (Port Jinling). The three signals from top to bottom are chest breathing, abdominal breathing, and the superposition of two breathing signals. The exhalation reset amplitude from A1 to A4 represents the respiratory pattern of poem chanting, and that from B1 to B4 of poem reciting. A direct observation shows that the respiratory pattern of chanting involves larger exhalation reset amplitude compared with that of reciting the poem. In addition, an alternating pattern emerges in the breathing signal of reciting poems with the smaller exhalation reset amplitude immediately following the larger one.

Table 1 Comparison of parameters of exhalation phase between chanting and reciting of modern-style poems

		amplitude of exhalation reset	duration of exhalation(s)	slope of exhalation phase	area of exhalation phase
chest breathing	reciting	0.29	4.07	-0.10	6.36
	chanting	0.13	1.65	0.07	0.94
abdominal breathing	reciting	0.69	4.26	-0.18	5.85
	chanting	0.42	1.82	-0.27	1.54

Table 1 shows the average exhalation parameters of the chest and abdominal breathing in two modes of performance for the modern-style poems. It can be seen that the chanting has larger exhalation reset amplitude for both chest and abdominal breathing. The duration of the exhalation phase is three times longer than that of the inhalation phase. The smaller exhalation slope in chanting indicates that the abdominal muscular contraction tends to be slowly escalated to produce sustainable breath. The area of the exhalation for chest and abdominal breathing in chanting is about nine times and five times larger respectively, than those in reciting. In summary, the four exhalation parameters indicate that more breathing volume is applied in chanting compared with reciting.

Table 2 Comparison of parameters of inhalation phase between chanting and reciting of modern-style poems

		amplitude of inhalation reset	duration of inhalation(s)	slope of inhalation phase	area of inhalation phase
chest breathing	reciting	-0.54	0.46	1.80	0.99
	chanting	-0.31	0.33	1.54	0.43
abdominal breathing	reciting	-0.45	0.59	2.34	1.08
	chanting	-0.27	0.47	1.75	0.48

Table 2 shows the average inhalation breathing parameters of the chest and abdominal breathing in chanting and reciting modern-style poems. It is clearly seen that chanting consumes more breath volume. The inhalation reset amplitudes of the chest and abdominal breathing in chanting are -0.54 and -0.45, which are lower than those in reciting (-0.31 and -0.27). This result indicates that chanters take a deeper breath in each respiratory cycle. Meanwhile, the chanting has a longer inhalation

duration and steeper inhalation slope, and the inhalation area is two times larger than that in reciting. The above results indicate that more breath volume is applied in chanting compared with reciting.

3.2 The Difference of the Respiratory Patterns between Reciting and Chanting *ci*

Similar respiratory pattern is applied in chanting *ci* (or classical Chinese poetry from Song Dynasty) with larger inhalation amplitude and breath volume, and longer exhalation phase.

Table 3 shows the average chest and abdominal breathing parameter of the chanting and reciting 39 of poems. It can be seen that the exhalation reset amplitude is larger in chanting compared with reciting. The exhalation duration and area of exhalation phase in chanting are two times larger than those in reciting. The exhalation curve of abdominal breathing in chanting declines much more slowly than that in reciting which yields a larger slope, and this is indicative of gradual and continual contraction of the abdominal muscles. To summarize, more breathing volume is consumed in chanting, and it is released gradually through the continual contraction of the abdominal muscles.

Table 3 Comparison of parameters of exhalation phase between chanting and reciting of Song poems

		amplitude of exhalation reset	duration of exhalation(s)	slope of exhalation phase	area of exhalation phase
chest breathing	reciting	0.21	5.16	-0.06	17.43
	chanting	0.07	3.35	-0.01	7.03
abdominal breathing	reciting	0.58	3.85	-0.19	18.99
	chanting	0.37	1.95	-0.25	7.72

Table 4 shows the average inhalation parameters of the chest and abdominal breathing in chanting and reciting *ci*. Like those in chanting modern-style poems, a deeper breath is taken in chanting than in reciting poems, which indicates more breathing volume is taken into the lungs. The inhalation reset amplitudes of the chest and abdominal breathing reach -0.56 and -0.41, respectively, lower than their counterparts in reciting (-0.40 and -0.14). Other parameters show that chanting poems involves longer inhalation duration and steeper inhalation slope. The inhalation phase area in chanting is two times larger than that in reciting.

Table 4 Comparison of parameters of inhalation phase between chanting and reciting of Song poems

		amplitude of inhalation reset	duration of inhalation(s)	slope of inhalation phase	area of inhalation phase
chest breathing	reciting	-0.56	0.88	1.89	6.83
	chanting	-0.40	0.55	1.25	3.61
abdominal breathing	reciting	-0.41	0.72	1.74	2.67
	chanting	-0.14	0.65	1.10	1.12

3.3 The Effect between Chest and Abdominal Breathing and their Relationship

In speechless natural breathing, abdominal breathing is the major means. The lengths of inhalation phase and exhalation phase are almost the same. The upper column of Fig. 4 indicates chest breathing, while the column below indicates abdominal breathing. In speechless natural breathing, the amplitude of chest breathing is comparatively small, while the abdominal breathing is large. Both are obviously periodical with the inhalation and exhalation phase almost equal.

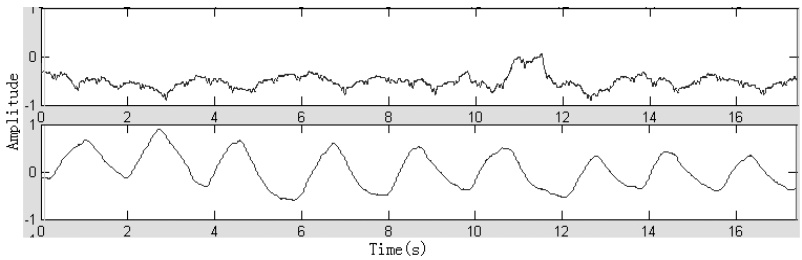


Figure 4 The chest and abdominal breathing in speechless natural breathing

When one is speaking, the major effect of chest breathing is to maintain the extended state of the chest, providing enough breath for articulation and chest resonance. The major effect of abdominal breathing is to provide stable sub-glottal pressure through contraction of abdominal muscle and diaphragm, controlling continuous airflow to produce continuous sound. The reset time of abdominal breathing is earlier than the onset time of chest breathing and sound.

The chanting of 80 poems was selected from all samples of the eight chanters and evaluated by 20 college students. The result shows that the samples with good evaluation have a feature in common: the patterns of chest and abdominal breathing are similar. As shown in Fig. 5, the onset times of sound and chest breathing (dotted line) overlap, that is, the chest is in the extended state when chanting. The times of inhalation and exhalation reset of abdominal breathing (solid line) are earlier than chest breathing. Sound begins after the exhalation phase of abdominal breathing the curve of which falls slightly. This shows the continuous and tender contraction of abdominal muscle and diaphragm.

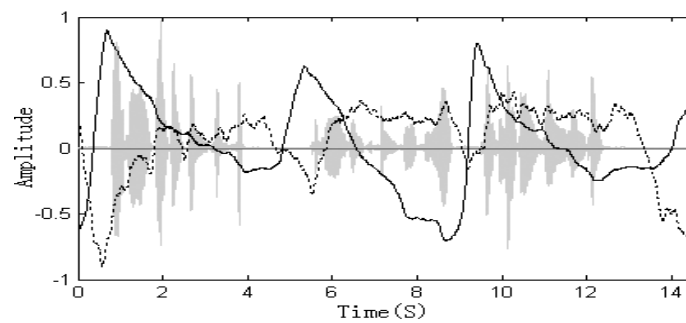


Figure 5 The relationship among chest breathing, abdominal breathing and sound

The major effect of the chest is to provide enough breath for articulation and chest resonance. The pause or declination of chest breathing is related to prosodic boundaries. Fig. 6 shows that chest breathing is related to prosodic boundaries. There are four seven-syllable poem lines in the figure, corresponding to four periods of chest breathing. A, B, C, and D are boundaries between prosodic phrases in one sentence, where chest breathing pauses. Small fluctuations at other positions correspond to boundaries between prosodic words.

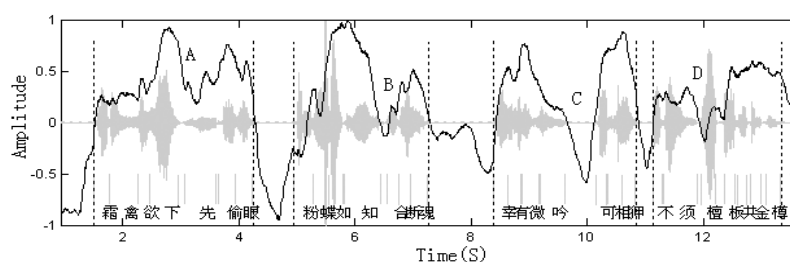


Figure 6 Chest breathing and prosodic boundaries

4. CONCLUSION AND PROSPECT

This research makes use of chest and abdominal breathing signals associated with reciting and chanting of ancient Chinese poems, in order to study the features of chest and abdominal breathing in different styles of poems and in reciting and chanting, respectively, and the effect of chest and abdominal breathing and their interrelationship. Computer programs for breathing analysis have been written to extract parameters, such as breathing reset amplitude, time of inhalation phase, and slope of exhalation phase. The results show that:

(1) In chanting poems both the amplitude of breathing reset and the amount of breath are larger than in reciting. That is, the depth of breathing, the amount of breath, and the time of exhalation increase.

(2) The times of inhalation and exhalation reset of abdominal breathing are earlier than chest breathing. The pause or declination of chest breathing is related to prosodic boundaries.

(3) The major effect of chest breathing when speaking is to keep the extended state of the chest to provide enough breath for articulation and chest resonance.

(4) The major effect of abdominal breathing is to provide stable sub-glottal pressure through contraction of abdominal muscles and diaphragm, controlling continuous airflow to produce continuous sound.

For future study, the barometric pressure and airflow are required to be measured to calculate more exactly the amount of airflow, sub-glottal pressure, and the volume of the lungs. Images of the lungs

obtained by spiral CT or MRI are also needed to model three-dimensional dynamics of the lungs, which play an important role in exploring the breathing mechanism of speech production.

NOTES

1. This study is supported by funds from National Social Sciences Funds [10&ZD125], Ministry of Education of China [13YJC740119], China Postdoctoral Science Funds [2014M552277], and Guangzhou Social Science Foundation [13Q03]. Prof. Kong Jiangping read the draft and offered his insights. Prof. Li Yinghao read the whole paper and provided the author with detailed comments and suggestions. Part of the results of this paper was orally presented at 12th National Conference on Man-Machine Speech Communication, 2013, Guiyang, China.

REFERENCES

- HIXON, T. J., 1987. Respiratory function in speech. In *Respiratory Function in Speech and Song*, ed. T. J. Hixon, 1-54. Boston: College-Hill Press.
- LADEFOGED, P. and G. Loeb. 2002. Preliminary studies on respiratory activity in speech. *UCLA Working Papers in Phonetics* 101: 50-60.
- LIEBERMAN, P. 1967. *Intonation, Perception, and Language*. Cambridge, Mass.: The MIT Press.
- OHALA, J. J. 1990. Respiratory activity in speech. In *Speech Production and Speech Modelling*, ed. W. J. Hardcastle, A. Marchal, 23-53. Dordrecht: Kluwer Academic Publishers.
- SHI, Feng, Jinyu Zhang, Xuejun Bai, and Zhaohong Zhu. 2010. Intonation and respiration: A preliminary analysis. *Journal of Chinese Linguistics* 38(2):323-335.
- TAN, Jingjing 譚晶晶, Yonghong Li 李永宏, and Jiangping Kong 孔江平. 2008. Hanyu Putonghua butong wenti langdu shi de huxi chongzhi texing 漢語普通話不同文體朗讀時的呼吸重置特徵 (Breathing-reset when reading literature in Mandarin). *Tsinghua daxue xuebao* (ziran kexue ban) 清華大學學報(自然科學版)

48(S1):613-620.

WANG, W. S-Y 王士元. 2008. Hongguan yuyinxue 宏觀語音學 (Macro phonetics). In *Zhongguo yuyin xuebao* (di 1 ji) 中國語音學報, 第 1 輯, 1-9. Beijing 北京: Shangwu Yinshuguan 商務印書館.

WU, Zongji 吳宗濟 and Maocan Lin 林茂燦. 1989. *Shiyan yuyinxue gangyao* 實驗語音學綱要 (Compendium of experimental phonetics) Beijing: Gaodeng Jiaoyu Chubanshe 高等教育出版社.

YANG, Feng 楊鋒, Xingquan Hou 侯興泉, and Jiangping Kong 孔江平. 2013. Yanyu chansheng de xiongfù hūxī jīzhì 言語產生的胸腹呼吸機制 (Chest and stomach breathing control for speech production). *Tsinghua daxue xuebao* (ziran kexue ban) 清華大學學報 (自然科學版) 53(6):852-855.

古詩詞朗讀與吟誦的呼吸特徵研究

楊鋒*

暨南大學, 廣州

提要

為探討胸腹呼吸在不同文體與不同誦讀方式中的特徵，以及胸腹呼吸的作用和二者的關係，本文同步錄製了 8 位吟誦人的 85 首近體詩和 39 篇詞的語音、胸呼吸和腹呼吸三路信號。使用自主編寫的呼吸分析程式，提取了呼吸重置幅度、吸氣相時間、呼氣相斜率等參數。實驗結果表明，古詩詞吟誦比朗讀中呼吸重置幅度大，即呼吸深度大，氣息量增大，呼氣時間長，氣息更為平緩。言語狀態下腹呼吸重置時間早於胸呼吸和語音起始時間，胸呼吸間斷或下傾與韻律邊界相關。言語狀態下胸呼吸主要作用是在發音時胸腔保持擴張狀態至發音結束，為發音提供足夠的氣息，同時提供胸腔共鳴；腹呼吸主要作用是，通過腹肌和膈肌的穩健收縮，以保證穩定的聲門下壓，控制氣流持續釋放，以獲得連續的語音。

主題詞

胸腹呼吸 朗讀 吟誦 言語產生

* 楊鋒執教于暨南大學華文學院及暨南大學亞太語言文化生態研究中心 通訊：中國廣州暨南大學華文學院，郵編 510610；[yangzihuai@163.com]

AN ARTICULATORY MODEL OF
STANDARD CHINESE USING MRI AND X-RAY MOVIE

Gaowu Wang^{*}, Jiangping Kong[†]

Beijing Normal University

Peking University

ABSTRACT

To better understand speech production from the phonological inputs to articulatory movements and then to acoustic outputs, it is important to establish an elaborate articulatory model of the vocal tract. This paper has explored the articulatory mechanism of speech production in Standard Chinese and developed a geometric articulatory model in both the visual and acoustic modalities. This model was based on the data of MRI images and X-ray movie, with the former providing detailed volumetric information of the vocal tract, and the latter the dynamic information of articulation. In this model, the seven articulators have been studied and modeled, respectively, including the hard palate, pharynx, jaw, lips, velum, tongue, and larynx. The tongue is modeled as two parts: tongue tip and tongue body, thus reducing the necessary number of parameters. The relation between larynx height and fundamental frequency in regard to the four tones is also modeled. Both of them have contributed new ideas to the articulatory modeling of Standard Chinese. This model can serve as a research tool for linguists, phoneticians, and speech engineers, and can be used in parameter speech synthesis, virtual speaker, and visual assistant speech training of Standard Chinese.

^{*} Gaowu Wang 汪高武 conducts research at Dept. of Chinese Language and Literature, Peking University and teaches at Beijing Normal University.

[†] Jiangping Kong 孔江平 is a professor of Peking University at Center for Chinese Linguistic, Department of Chinese Language and Literature, and Joint Research Center for Language and Human Complexity. He is the corresponding author of the article: Dept. of Chinese Language and Literature, Peking University, Beijing 100871, China; [jpkong@pku.edu.cn]

SUBJECT KEYWORDS

Speech production Articulatory model Vocal tract **MRI X-ray** Movie

1. INTRODUCTION

Speech is the important tool for humankind's thinking and communication, therefore the study of speech sounds, known as phonetics, has fundamentally theoretical and practical value. With the development of advanced instruments and scientific methods, traditional phonetics can overcome its limitations of "by mouth and ear" (Luo 1956) to develop into modern experimental phonetics, thus bridging the gap between the traditional language culture and modern scientific study. In modern phonetics, speech production has become the frontier, since it involves interdisciplinary cooperation, including linguistics, phonetics, psychology, acoustics, computer engineering, and other fields.

To better understand speech production from the phonological inputs to articulatory movements and then to acoustic outputs, it is important to establish an elaborate articulatory model of the vocal tract. At present, this topic has attracted broad interests internationally, based on the speech materials of many languages, including English, French, German, Japanese, Swedish, and others, but few of Chinese (Wang et al. 2008). And there are few advanced studies to establish an articulatory model embracing the specific properties of Chinese.

This paper can serve as an exploratory research in the field of speech production modeling in Standard Chinese, as it has established an articulatory physiological geometric model of Standard Chinese. In order to establish the model, a database of articulatory movements of Standard Chinese was established, including MRI images and X-ray movie, for which the former provides detailed volumetric information of the vocal tract and the latter the dynamic information of articulation.

2. MRI AND X-RAY DATABASE

2.1. MRI Experiment

Magnetic Resonance Imaging (MRI) allows a tomographic view of body tissues in any plane within the human body, and gets the 3D shape of the vocal tract without known risks for the subject, therefore, it has been increasingly applied in speech research over the past 20 years (Baer et al. 1991; Demolin, Metens, and

Soquet 1996;Story, Titze, and Hoffman 1996;Dang and Honda 1997;Badin et al. 1998;Stone et al. 2000). The articulatory data collected using MRI are valuable in understanding and accurately modeling the vocal tract, particularly the pharynx, the behavior of which during speech is traditionally hard to capture. Also, the volumetric production data are very important in articulatory synthesis, in which scientists have been involved for several decades.

At present, MRI studies have been carried out on several languages, e.g. English, French, Japanese, Swedish, and others. However, few MRI studies have been conducted on Standard Chinese. This study has started to establish an MRI database of Standard Chinese for phonetic studies, including articulatory modeling. A very brief introduction of this database follows.

Subjects: There are six male and two female subjects in this database. The subjects have no speech or voice problems. They are native Mandarin speakers or have passed the Putonghua Proficiency Test (PPT) and achieved Grade One, Level B. All the subjects have been trained to ensure articulatory stability for obtaining clear MRI images.

Speech materials: This MRI database covers the nine single vowels in Standard Chinese /a o e i u ü(i)e (s)i (sh)i/, or in IPA / a o ㄛ i u y ㄩ ㄣ/, respectively.

The nine vowels are produced in Chinese single syllable words, “啊喔厠衣烏淤椰思詩”, respectively, while the stable posterior parts in “椰思詩” were segmented and treated as independent vowels in the following speech analysis. All the characters are pronounced in the first tone (high flat tone) to ensure the stability of the sustained vowels. Several sustained consonants including /s ㄙ ㄘ ㄑ ㄒ ㄓ ㄔ ㄕ/ are also covered. In addition, 75 consonant variants, such as /b(a) b(i) b(u) b(o) ... s (a) s(i) s(u)/, produced by one male subject are also imaged in the mid-sagittal plane.

MRI equipment and scan specifications: The MRI data were acquired with the Shimadzu-Marconi ECLIPSE 1.5T PowerDrive 250 installed at the Brain Activity Imaging Center, Advanced Telecommunications Research Institute (ATR-BAIC), in Kyoto, Japan. The MRI machine parameters were as follows: a 3.4 [ms] echo time (TE), a 2200 [ms] relaxation time (TR), 44~51 sagittal slice planes, a 1.5 [mm] slice thickness, a 1.5 [mm] slice interval, a 256*256 [mm] field of view (FOV), and a 512*512 pixel image size.

Reconstruction of volumetric data: Based on the three dimensional

information provided by MRI images, the volumetric shape of articulators can be obtained. Along the center line of the vocal tract from the glottis to the lips, the cross sections of the vocal tract can be sliced, and then the vocal tract air way can be reconstructed. An example of vocal tract configurations for the vowel /a/ in Standard Chinese is shown in Figure 1.

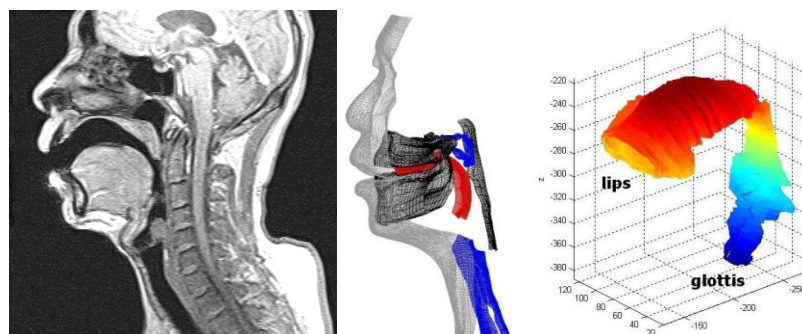


Figure 1. MRI image (left), 3D volume of articulators (middle), and vocal tract airway (right) of vowel /a/

2.2. X-ray Movie

Research on speech using X-ray has a long history. According to Dart's review (Dart 1987), as early as 1907, Barth and Grunmach had used still X-ray to study speech. In 1942, Chiba combined X-ray photography, palatography, and laryngoscopy to measure vocal tract shape in pioneering research (Chiba and Kajiyama 1942). Thereafter, Fant established the theory of speech production based on the extensive analysis of X-ray data. Also, some important studies (Heinz and Stevens 1964; Perkell 1969) have been successively carried out using X-ray data. Although other new technologies such as MRI, CT, and EMA were developed, nowadays the full sagittal view of the vocal tract articulators during running speech provided by cineradiography remains unsurpassed by more modern techniques. The data obtained from the cineradiography are still used in recent research (Beautemps, Badin, and Bailly 2001; Iskarous 2005).

In China, Zhou and Wu (1963) recorded X-ray photos for Standard Chinese vowels and consonants uttered by a female native speaker. Bao and Yang (1982) had measured the articulatory movements of five consecutive Standard Chinese vowels using the cineradiography. Also, some preliminary studies were

performed on the relationship between the cross-sectional area function of the vocal tract and formant frequencies of vowels (Bao 1983), and the synthesis of Standard Chinese single vowels by articulatory parameters (Zu 1983).

In this study, an X-ray video provided by Bao(1983, 1984) is the only available cineradiography corpus of Standard Chinese. In total 786 utterances of two male and two female subjects are shown to illustrate most of the sounds of Standard Chinese, covering 204 syllables, 36 “retroflexed” and 36 tonal syllables. This video has been segmented to syllables in avi files at 30 fps. The information of this database is detailed in Wang (2010).

A platform ‘VocalMarker’ in Matlab is programmed to trace the midsagittal articulatory movements. Figure 2 shows a part of the VocalMarker interface, in which the edges of articulators are extracted and the phonetic markers of speech sound are labeled.

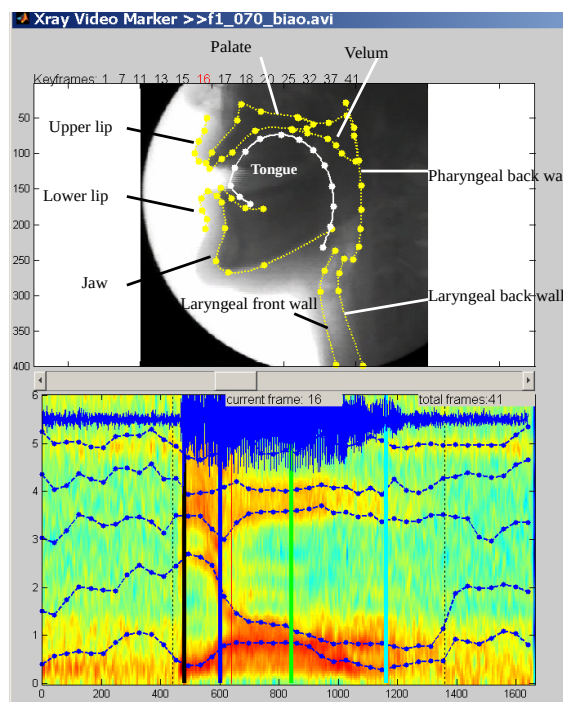


Figure 2. A part of the VocalMarker graphic interface: 1) upper: one frame in X-ray with the edges of articulators extracted; 2) lower: the speech sound wave and its spectrum are shown, the formants are measured and the phonetic markers are labeled. The two black dashed lines denote the period of sound; the four solid lines denote the

initial consonant, the head vowel, essential vowel, and tail vowel, respectively.

3. ARTICULATORY MODELING

Generally, speech production can be divided into two parts or steps: phonation (vibration of the vocal folds), and articulation (resonance of the supraglottal organs or vocal tract). According to source-filter theory in speech production (Fant 1970), the acoustic process of speech production can thus be modeled as two parts: the source (phonation) and the filter (articulation). As for the supraglottal part, the airway of the vocal tract can be regarded as a resonant tube filter, whose changing shape produces variant speech outputs. Up to the present many studies have attempted to model this articulatory process. Generally, the articulatory models can be divided into three categories: area function parametric model, geometrical articulatory model, and explicit biomechanical model.

Area function parametric models directly take the area function of the vocal tract as the parameters, regardless of the shapes of the individual articulator (Flanagan 1972;Stevens 1989). However, the weakness of these models lies in its merely focusing on the area function, thus hardly corresponding to the physiological structure and articulatory movement for further exploring.

Explicit biomechanical models explore the physiological and anatomy level, focusing on the internal muscles structure. The articulators are modeled on the basis of mechanical mechanisms (Perkell 1974) and can be modeled by the finite element method (FEM) (Wilhelms-Tricarico 1996;Dang and Honda 2004). At present, its weakness still lies in lacking the empirical data of muscle forces and requiring a high computational cost.

Geometrical articulatory models focus on the shape and movement of the articulators instead of area function of the vocal tract, thus having more directly physiological meaning and the potentially more precise modeling while with less parameters. There are usually two ways to establish geometrical articulatory models: pre-defined and data-mining. The former means that a set of parameters describing the shape and movement of articulators has been pre-defined and then the model is trained by collected data(Mermelstein 1973). The later means that some statistical analysis has been performed on the collected data to find some ‘factors’(Lindblom and Sundberg 1970). Usually the factor analysis is performed on the tongue since its contour is hard to pre-define by a few parameters or simple equations. These two

ways (pre-defined and data-mining) can actually be used in one model. For example, Maeda's model used pre-defined parameters and analyzed factors at the same time (Maeda 1982), which is one of the most successful models.

In this study, we will establish a geometrical articulatory model using the combined method of pre-defined and data mining, based on the database of MRI images and X-ray movie. It is assumed that the articulatory mechanism of Standard Chinese between the subjects in the X-ray movie and MRI data is consistent, therefore these two sets of data can be used together to establish the model.

Basically, the whole articulatory model comprises three parts: spatial modeling, temporal modeling, and acoustic output. In spatial modeling, the main task is to find a set of parameters which can describe the articulatory configuration using as few parameters as possible and have vivid articulatory meanings easily perceived by people at the same time. In temporal modeling, the main task is to analyze the dynamic pattern of articulatory movement. Since Standard Chinese is monosyllabic, our study will focus on the articulatory movement within the monosyllable. In acoustical modeling, the main task is to generate the speech sound signal from the dynamic articulatory configurations.

3.1 Spatial Modeling

The task of spatial analysis is to model the structure and shape of articulators. Usually, the articulatory movement is very complicated, therefore, based on the physiological structure, the vocal tract will be disassembled as different articulators: jaw, lips, velum, tongue, palate, and larynx. In this study, an articulatory model of Standard Chinese is established, with its 12 parameters shown in Figure 3. In the following sections, we will analyze the articulators and their parameters individually.

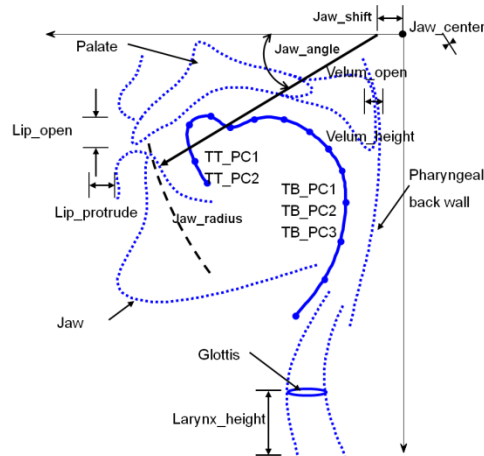


Figure 3. The articulatory model of Standard Chinese using 12 parameters

Palate and Pharynx

In most articulatory models, the hard palate and laryngeal back wall are usually set as a fixed curve or straight line because they are attached to fixed bones. Although X-ray movie showed that the hard palate will be raised a little during some low-vowels and nasals, it is believed that these movements are not necessary, as we can fix the subject's head just like we have done in MRI experiments. Hence, the hard palate and the laryngeal back wall can be set as fixed position. As shown in Figure 3, the hard palate curve and the laryngeal back wall are set in fixed position.

Jaw

In articulatory movement, the jaw as a rigid body decides the openness of the mouth and influences the shape of the lips. Although intelligible utterance can still be produced when the jaw bone is clinched, the movement of the jaw in normal speech has been studied in many researches, as well as in this paper. In general, the jaw may be assumed to execute rigid rotation about a point fixed with respect to the stationary structures (Mermelstein 1973). In our study, Figure 4 shows the trajectory of the lower incisor tip (the highest point of the jaw) accumulated in 10,376 frames of the X-ray movie. The movement of the jaw cannot be described only by an arc with a fixed center and radius, as showed by the dashed arc line in Figure 4.

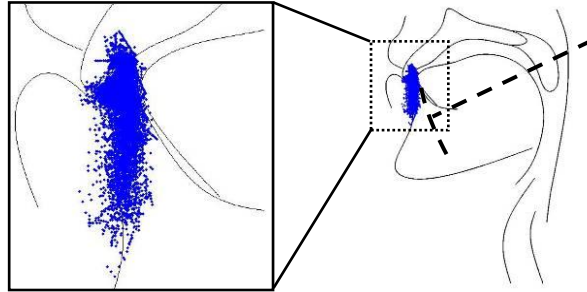


Figure 4. The trajectory of the lower incisor tip accumulated in 10,376 frames of X-ray movie

An arc equation is not enough to describe the movement of the jaw, therefore, we add a horizontal component vector in order to make a good accordance to the real trajectory of jaw movement (physiologically the jaw bone can move horizontally). As shown in Figure 3, the movement of the jaw is decomposed into two components: the arc decided by the parameter 'Jaw_angle' and the horizontal movement decided by the 'Jaw_shift'.

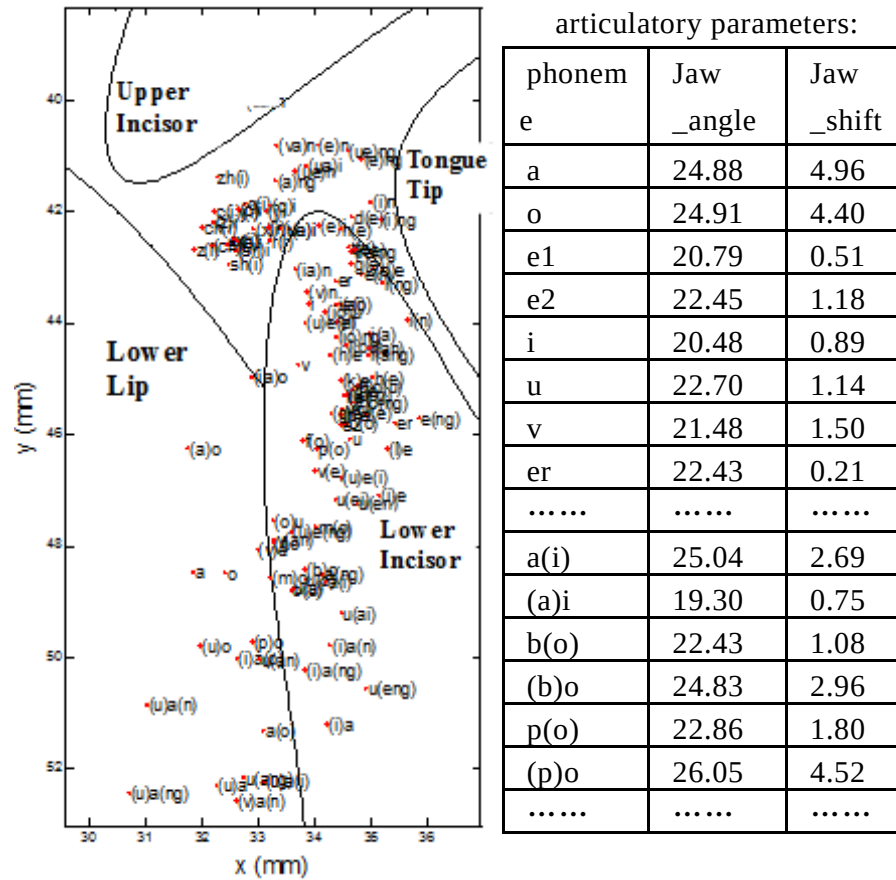


Figure 5. The distribution of the lower incisor tip in different phonemes (subject 01, X-ray movie) and its articulatory parameters

Figure 5 shows the distribution of the lower incisor tip in different phonemes of the specific subject in X-ray movie, and their corresponding ‘Jaw_angle’ and ‘Jaw_shift’ can be calculated by the following equation:

$$Jaw\_angle = \arcsin(y - Jaw\_center\_x) / Jaw\_radius / (\pi / 180)$$

$$Jaw\_shift = x - Jaw\_center\_x - Jaw\_radius \times \cos(Jaw\_angle),$$

where (Jaw_center_x, Jaw_center_y) are the location of the jaw pivot point, Jaw_radius (102.6mm in this case) is the radius of the jaw, (x,y) are the locations of the lower incisor tip. Therefore, using the two articulatory parameters ‘Jaw_angle’

and ‘Jaw_shift’, we can decide the jaw position in different phonemes.

Lips

In articulatory movement, the lips decide the production of the rounded or unrounded vowels and labial consonants. It is assumed here that the lower lip is attached to the jaw, therefore in the midsagittal plane the shapes of the lips can be described by openness and protrusion (the movement component caused by the jaw being removed). Accordingly, we can set two articulatory parameters : Lip_protrude, the extent of the lip protruding, whose value is in the range of [0 1], where 0 means the original shape, 1 means the lips are rounded and protruded; Lip_open, the distance between the upper and lower lips, whose value is in the range of [0 1], where 0 means the lips are closed, 1 means that the lips are opened.

Velum

In the field of phonetics, the distinction between the nasal and non-nasal is produced by the opening and closing of the nasal cavity which is controlled by raising the velum (soft palate). As observed in the X-ray movie and MRI images, the velum is not a rigid on-off switch, but rather an elastic body whose shape changes in its articulatory movements, especially when contacting the pharyngeal wall or for different vowels. This change influences the vocal tract shape, especially in the 2D articulatory model in the mid-sagittal plane. Therefore a detailed model of the velum is proposed in this paper for better simulation.

Since the target is to find the articulatory parameters of the velum, the cluster analysis (K-Means, Euclidean distance) has been performed on the shape of the velum in MRI images (subject 07, 10 vowels & 75 consonant variants). It is found that the best result comes out when the cluster number is 3. In Figure 6, the phonemes in Cluster 1: ma, mi, mu, mo, na, ni, nu, ne, are illustrated by the blue solid line ; Cluster 2 illustrated by red dotted line: a, ar, e, e2, pa, fa, da, di, du, ta, ti, tu, te, nv, la, ga, gu, ge, ka, ku, ke, ha, hu, he, ji, za, zi, ze, ca, ci, cu, ce, sa, si, su, se, zha, zhi, zhu, zhe, cha, chi, che, sha, shi, shu, she, ri, ru, re; Cluster 3 illustrated by green dashed line: er, i, i2, i3, o, u, v, ba, bi, bu, bo, pi, pu, po, fu, fo, de, li, lu, le, lv, ju, qi, qu, xi, xu, zu, chu. The phonemes in Cluster 1 are nasals; those in Clusters 2 and 3 are non-nasal; the difference between them is the ‘inner height of vowels’, with high vowels (such as i, v, u) tending to have a higher velum position.

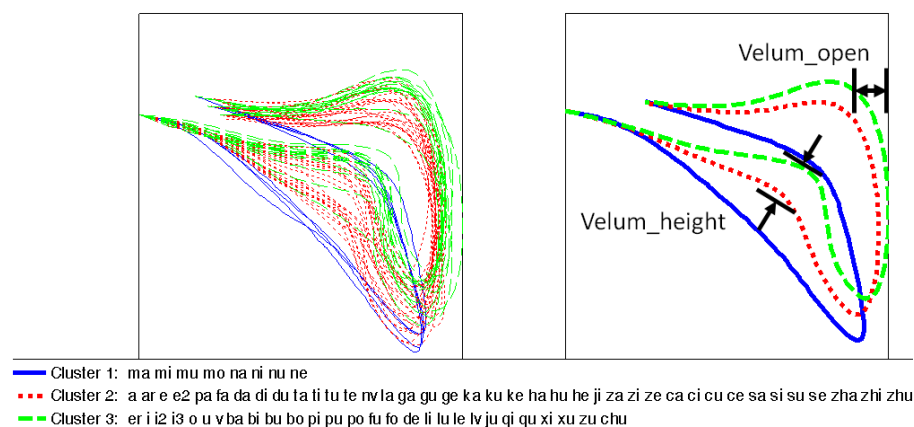


Figure 6. The clustering result of velum and the articulatory parameters

Just as with modeling of the lips, phoneticians do not have to care about the detailed shape of the velum, which is speaker-specific. Accordingly, we can set two articulatory parameters : Velum_open, whose value is in the range of [0 1], where 0 means the nonnasal and 1 means the nasals. Velum_heigh, whose value is in the range of [0 1], where 1 meansthe maxsium height of the velum during high vowels.

Tongue

Tongue is the most flexible and primary articulator, which consists totally of muscles and soft tissues. It determines the two main dimensions of vowels: the height and advancement of the tongue position. It also determines the articulatory mannar and position of all the consonants except bilabials and labiodentals. There is no such distinctive features as between the rounded vs unrounded, nasal vs nonnasal, therefore the shape and movement of the tongue can not be categorized or clustered. It remains big problem for how to describe the shape and movement of tongue using as few parameters as possible with articulatory meaning. In this paper, we propose a method to reduce the number of parameters by dividing the complicate tongue shape into two simpler parts : tongue body and tongue tip.

Principal Component Analysis (PCA) is a good mathematical tool to reduce the number of variables, while keeping as much information as possible and using as few orthogonal variables (called principal components). For instance, if we

perform PCA on the colors of nature, then three principal components: Red, Green, and Blue could be obtained. PCA has also been applied to the vocal tract, especially the tongue in many studies (Stone, Goldstein, and Zhang 1997;Beautemps, Badin, and Bailly 2001;Mokhtari et al. 2007). In this study, we performed the PCA on the tongue shape curve which is marked and generated by 14 key points ($x+y_i$). Figure 7 shows the result. Left image is for all the tongue shapes of one subject in MRI images, which shows the variability of the tongue shape. Right image is the screen plot of the percent variability explained by each PC (totally 14, but only 10 are plotted), which shows that the percentages are 61.36%, 25.98%, 6.97%, 3.20%, 1.01%, 0.47%, 0.32%, 0.23%, 0.14%, 0.10%, 0.08%, 0.07%, 0.04%, 0.03%, and the accumulated percentages are 61.36%, 87.34%, 94.31%, 97.51%, 98.52%, 98.99%, 99.31%, 99.54%, 99.68%, 99.78%, 99.86%, 99.93%, 99.97%, 100.00%. In other words, the first 2 PCs can explain the 87.34% variability of tongue shape with an average error 2.3mm, and the first 4 PCs are 97.51% and 1.0mm, which is the least requirement of error since the pixels in MRI image is 1.0mm/pixel.

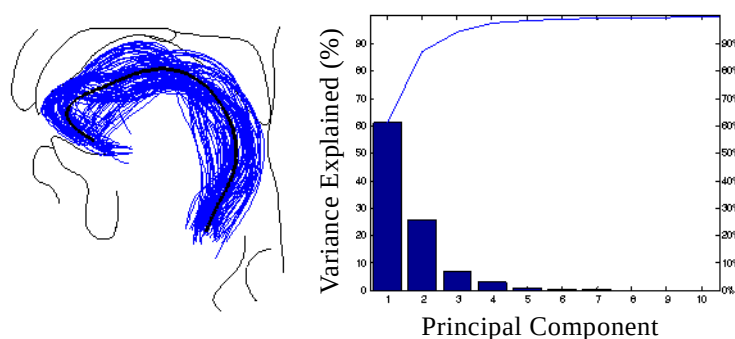


Figure 7. Variability of tongue shapes (left) and the percent variability explained by each PC (right)

Then what kind of tongue shape or movement do these PCs represent? Figure 8 shows the ‘articulatory movement’ presented by the first 4 PCs. The solid blue line denotes the ‘average/original tongue shape’, the dashed lines denote the result of adding ± 5 PC1 on the original tongue shape, and the dotted lines denote the result of adding ± 10 PC1. As can be seen from the figure, PC1 represents a kind of advancing movement of the tongue. Similarly, PC2 represents the advancing movement of the tongue tip with relatively static tongue body. PC3: advancing of

tongue root and retroflex of tongue tip; PC4: roughly a kind of arching of tongue body. However, these PCs' representations of articulatory movement are not yet so explicit.

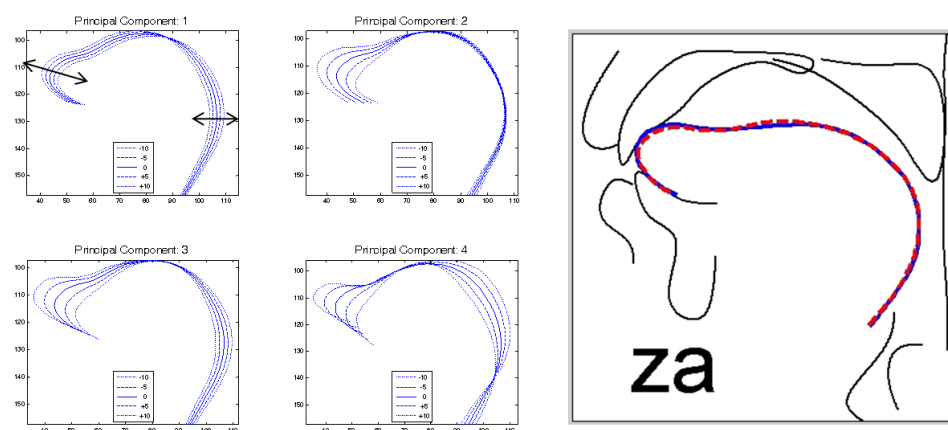


Figure 8. The articulatory movement represented by the first 4 PCs (left) and an example of unqualified tongue shape reconstruction (right)

If the first four PCs are used to reconstruct the tongue shape, the average error will be within 1 mm. However, for some phonemes having narrow constriction in the vocal tract, a higher precision will be required, otherwise the reconstruction error of tongue shape will lead to another configuration (with distinctive features, obstruction or un-obstruction, e.g. /z/ or /s/) of vocal tract undertaking different phonemic meanings. Just like the examples in Figure 10, the reconstructed tongue shape (red dashed line) of /z(a)/ (obstruction) is actually another distinctive configuration /s(a)/ (un-obstruction) different from the original tongue shape (blue solid line). This precision is very crucial from the viewpoint of linguistics, which is accordingly the principle requirement of mathematics and engineering analysis and modeling. Therefore more PCs are required to reconstruct the tongue shape precisely enough.

At the same time, we want to describe the tongue shape using as few parameters as possible. As observed in the X-ray movie and MRI images, the tongue body and tongue tip are relatively independent in movement. Therefore we propose a method for performing PCA on the tongue body and tongue tip, respectively, especially for Mandarin which has apical vowels. As shown in Figure 9, this method

can reduce the number of required parameters from six to five, and is more precise in the sensitive tongue tip region.

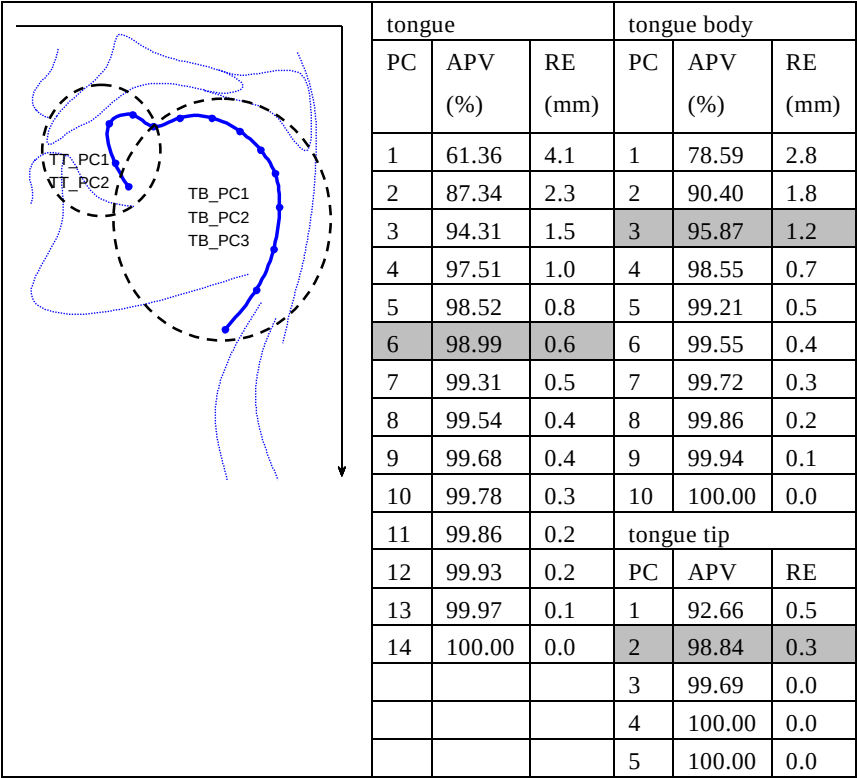


Figure 9. Tongue vs. <tongue body + tongue tip>, the tongue analyzed as two parts (left) for the benefit of reducing the number of parameters (right), where APV means accumulated percent variability, RE means reconstruction error.

As shown in Figure 10, the articulatory movements represented by the first three PCs of tongue body are: tongue body lifting; tongue body advancing; tongue body arching, accordingly, the first 3 PCs can be named as TB_elevation, TB_advancing, and TB_arching. An example of reconstruction result of tongue body is also shown in Figure 10, in which the distinctive feature (obstruction or un-obstruction) between /j(i)/ and /x(i)/ is represented adequately.

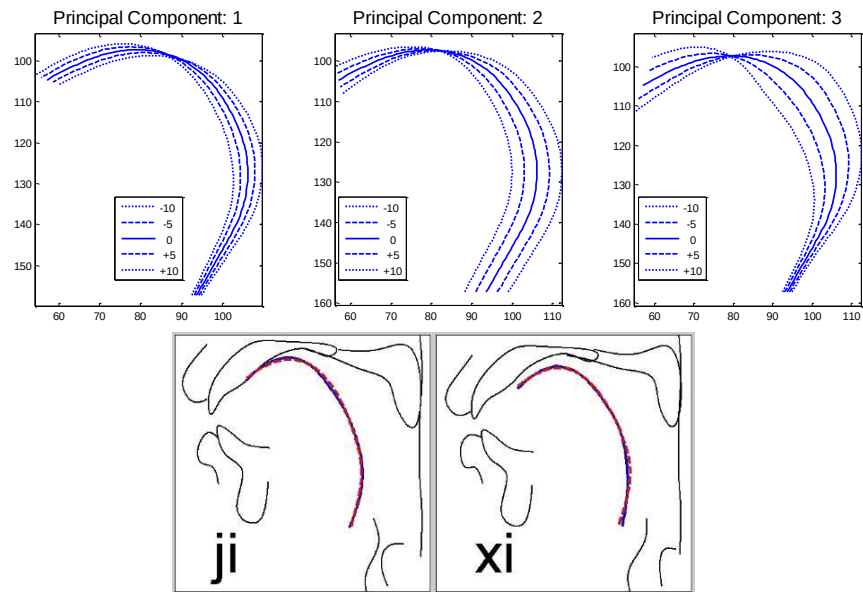


Figure 10. The articulatory movement represented by the first three PCs of tongue body and an example of reconstruction

As shown in Figure 11, the articulatory movement represented by the first two PCs of the tongue tip are tongue tip raising and tongue tip protruding, accordingly being named as TB_elevation, TT_raising and TT_protruding. An example of reconstruction result of tongue tip is also shown in Figure 11, in which the distinctive feature (obstruction or un-obstruction) between /z(a)/ and /s(a)/ is adequately represented.

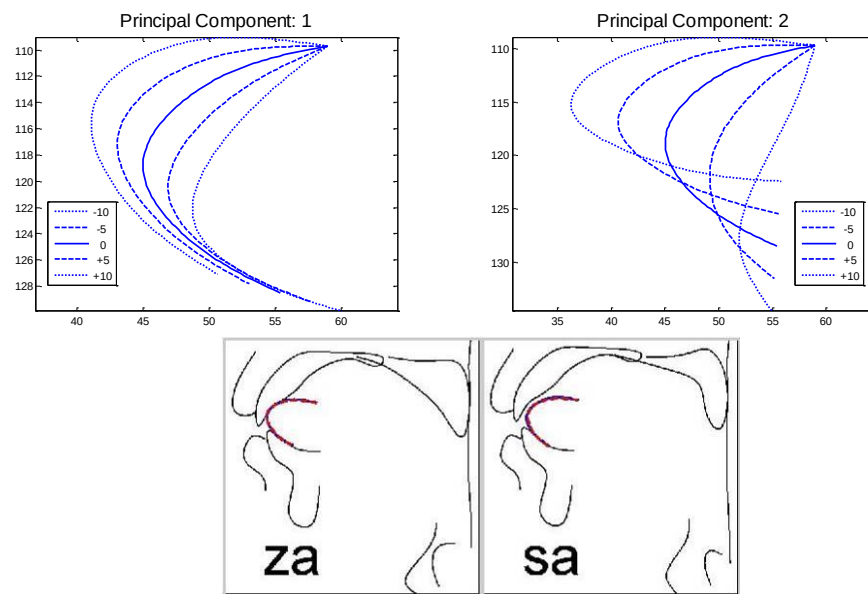


Figure 11. The articulatory movement represented by the first two PCs of tongue tip and an example of reconstruction

Larynx

Standard Chinese is a tonal language. When the fundamental frequency is increasing or decreasing, the larynx height is also changing, which will lead to the change of the shape and length of the vocal tract and the source-filter interaction. The relation between the fundamental frequency in the four tones and the larynx height will be analyzed in this study, which has not yet been performed in other articulatory models.

In the corpus of the X-ray movie, there are totally 36 syllables withba1/ba2/ba3/ba4 as a segment and 9 segments (ba, bi, bu, lao, liao, qian, qiang, zhai, chuai). Figure 12 shows the relation between the fundamental frequency and the larynx height for the 4 tones (36 syllables). The red stars denote the F0 which varies from 100 Hz to 250 Hz, the blue points denote the Larynx Height, varying from -7 mm to 7 mm (the LH is 0 when the subject is in natural rest). As we can see, for tone 1, the larynx raises when F0 remains flat; for tone 2, the larynx raises as F0 increase, and this is expressed with the equation: $\text{Larynx Height} = 0.61 \cdot \text{F0} - 135.97$; for tone 4,

larynx descends as F0 decreases, as expressed with the equation: Larynx Height = 0.31*F0 - 44.26; for tone 3, larynx descends first and then raises as F0 decreases and then increases, the same equations for tone 2 and tone 4 apply. Thus, we can set the parameter Larynx_height to describe the articulatory movement of the larynx for the 4 tones in Standard Chinese, as shown in Figure 3.

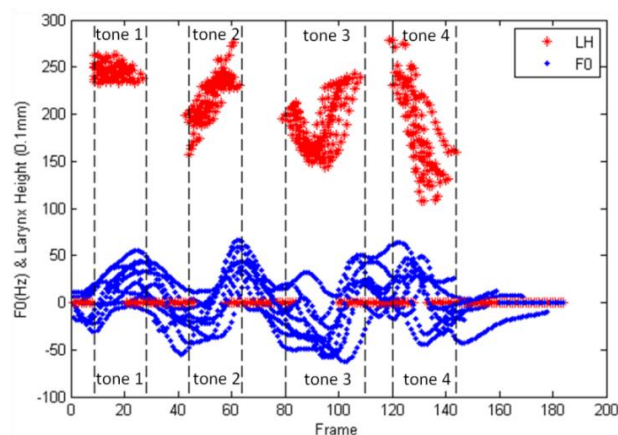


Figure 12. The relation between the fundamental frequency and the larynx height for the four tones

3.2 Temporal Modeling

Standard Chinese is a monosyllable language, and our study will focus on the articulatory movement within the monosyllable. The syllable structure of Chinese is: initial + head vowel + (essential vowel) + tail vowel, for which the essential vowel is indispensable. The movement pattern of each articulator may be different, which can be explored further. Here we take the tongue as an example, and the distance between one point on the tongue surface and gum ridge as an example of the articulatory movement pattern in the monosyllable. In Figure 13 the distance measured in the X-ray movie of 217 monosyllables is shown, which is roughly conic with the essential vowel target time as the symmetric axis. Thus a trajectory equation can be derived by quadratic fitting of these trajectory data.

$$\frac{D - D\_sheng, yunt, yunw}{D\_yunf - D\_sheng, yunt, yunw} = 1 - \left(\frac{T - T\_sheng, yunt, yunw}{T\_yunf - T\_sheng, yunt, yunw} \right)^2$$

where T is the time, D is the amplitude of the articulatory movement on T ; T_{sheng} , $yunt$, $yunw$ is the target time of initial, head vowel, or tail vowel, while D_{sheng} , $yunt$, $yunw$ is the amplitude of the articulatory movement of the time; T_{yunf} is the target time of the essential vowel, while T_{yunf} is the amplitude of the time. The period of the vowels can be adjusted, for example, according to Wu and Lin (1989), the best duration ratio of vowels in the falling diphthongs (ai/ei/ao/ou) is 6:4; for rising diphthongs (ia/ie/ua/uo/ve) 4:6; for triphthongs (iao/iou/uai/uei) 4:4:2.

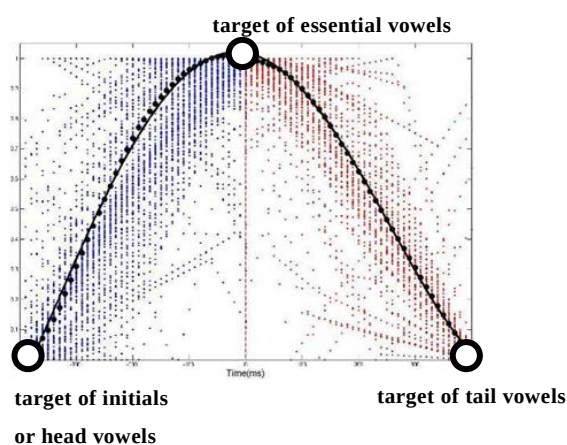


Figure 13. An example of the articulatory movement pattern inside a monosyllable

3.3 Acoustical Output

This articulatory model will generate not only visual output (the articulators' shape in mid-sagittal plane), but also acoustic output (the simultaneous speech sound). According to the source-filter theory of speech production (Fant 1970), speech is produced by the vibration of the vocal folds as a source (input) and the resonance of vocal tract as a filter. And according to transmission line model, the vocal tract is analogous to a cascade circuit, and the resonant properties (transmission function) can be calculated from the cross-sectional area of the vocal tract. Our articulatory model is 2D in mid-sagittal plane based on the X-ray movie, therefore we have to find how to calculate cross-sectional areas from the midsagittal width of the vocal tract. There are already many equations (Beautemps, Badin, and Laboissiere 1995; Baer et al. 1991; Sundberg et al. 1987), and the most famous $\alpha \beta$

-model (Heinz and Stevens 1964;Sundberg et al. 1987;Dang and Honda 2002) is used in this study. A set of α β parameters was estimated from the 3D data of MRI and applied to calculate the cross-sectional area from the mid-sagittal width of the vocal tract in the X-ray movie (Wang, Dang, and Kong 2008).

From the area function of the vocal tract, the frequencies and widths of formants of the speech signal (the transmission function of the vocal tract) can be calculated (Sondhi 1973 and1974;Dang and Honda, 1997;Flanagan, 1972).According to the comparison between the calculated formants and the measured formants from synchronous speech recording, the average error could be decreased to 7.4% (Wang et al., 2008).

3.4 Articulatory Model of Standard Chinese

Finally, based on the above analysis and modeling, an articulatory model of Standard Chinese has been established. Figure 14 demonstrates the model output. By a set of 12 articulatory parameters, each articulator is set and the vocal tract configuration in the midsagittal plane is generated (as the blue lines indicate). The center line of the vocal tract (as indicated by the red line) is generated and the mid-sagittal width of the vocal tract is measured. According to the α β -model, the area function of the vocal tract is obtained, as shown in window a. The transmission properties of the vocal tract are obtained, as shown in window b. The waveform in the time domain and the spectrum in the frequency domain of speech signal output are shown in window c and d, using the LF-model as the source signal.

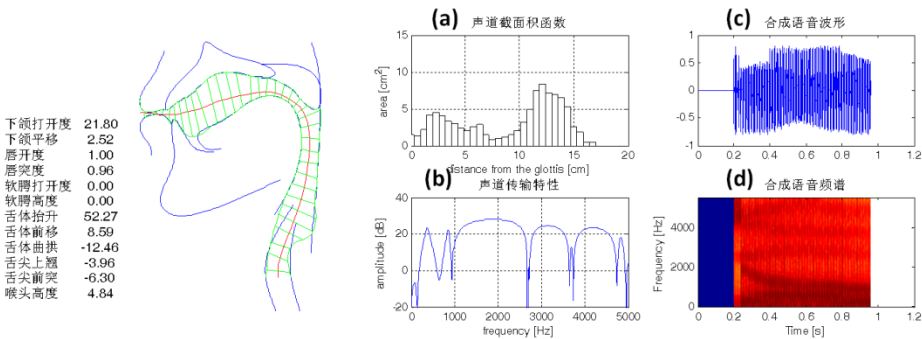


Figure 14. A demonstration of the articulatory model output (a frame for the synthesized syllable /iao/)

4. CONCLUSION

In this study, a preliminary articulatory model of Standard Chinese has been established using a set of 12 parameters, based on the corpus of the X-ray movie and MRI images. The tongue is modeled as two parts: tongue tip and tongue body, thus reducing the necessary number of parameters. The relation between the larynx height and fundamental frequency in the four tones is also modeled. Both of them have contributed new ideas to the articulatory modeling of Standard Chinese.

However, there is still much room for improving this model. In spatial modeling, the articulatory space is not fully explored since the modeling of some articulators is based only on the data of one subject. In addition, the mechanism of articulators' conflicts has not been studied. In the temporal modeling, the trajectory of articulators has been over-simplified as a conic curve, while actually for different monosyllables it might be different. In addition, the movement pattern of each articulator may be different. Both of them need to be explored further.

This model, although having some language-specific characteristics based on the MRI and X-ray database of Standard Chinese, is still a general (or non-language-specific) articulatory model. It is a bridge from the phonological inputs to articulatory movements and then to acoustic outputs. A well-established articulatory model will help to better understand speech production, and serve as a research tool for linguists, phoneticians, and speech engineers.

NOTES

1. This work was supported in part by NNSFC grant (No.61073085) and NSSFC grant (No.10&ZD125).

REFERENCES

- BADIN, P., G. Bailly, G. Raybaudi, and C. Segebrath. 1998. A three-dimensional linear articulatory model based on MRI data. In Proceedings of the 3rd ESCA/COCOSDA, International Workshop on Speech Synthesis 249-254.
- BAER, T., J. C. Gore, L. C. Gracco, and P. W. Nye. 1991. "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels", *Journal of the Acoustical Society of America* 90 (2):799-828.

- BAO, Huaiqiao 鮑懷翹. 1983. Shengdao mianji hanshu he gongzhenfeng pinlv de chubu baogao 聲道面積函數和共振峰頻率的初步報告 (On the relationship between cross-sectional area function of vocal tract and formant frequencies of vowel: A preliminary report). Yuyin yanjiu baogao 語音研究報告 1983-1984, 1-17. Zhongguo Shehui Kexueyuan Yuyanxue Yuyin Yanjiushi 中國社會科學院語言所語音研究室.
- BAO, Huaiqiao 鮑懷翹, and L. L. Yang 楊力立. 1982. Yanyin lianxu fayin de X guang dianying yanjiu 元音連續發音的X光電影研究——兼論元音分類的標準 (Study on cineradiography of consecutive vowels). In Yuyin yanjiu baogao 語音研究報告 1982, 1-18. Zhongguo Shehui Kexueyuan Yuyanxue Yuyin Yanjiushi 中國社會科學院語言研究所.
- BEAUTEMPS, D., P. Badin, and Rafael Laboissiere. 1995. "Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data." *Speech Communication* 16:27-47.
- BEAUTEMPS, D., P. Badin, and Gerard Bailly. 2001. "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling." *Journal of the Acoustical Society of America* 109 (5):2165-2180.
- CHIBA, T., and M. Kajiyama. 1942. *The Vowel: Its Nature and Structure*. Tokyo: Phonetic Society of Japan.
- DANG, Jianwu, and Kiyoshi Honda. 1997. "Acoustic characteristics of the piriform fossa in models and humans." *Journal of the Acoustical Society of America* 101 (1):456-465.
- _____. 2002. "Estimation of vocal tract shapes from speech sounds with a physiological articulatory model." *Journal of Phonetics* 30 (3):511-532.
- _____. 2004. "Construction and control of a physiological articulatory model." *Journal of the Acoustical Society of America* 115 (2):853-870.
- DART, S. N. 1987. "A bibliography of X-ray studies of speech." *UCLA Working Papers in Phonetics* 66:1-97.
- DEMOLIN, D., T. Metens, and A. Soquet. 1996. Three-dimensional measurement of the vocal tract by MRI. Paper presented at the 4th ICSLP, at Philadelphia.
- FANT, G. 1970. *Acoustic Theory of Speech Production*, 2nd ed. Mouton: The Hague.
- FLANAGAN, J. 1972. *Speech Analysis Synthesis and Perception*. New York:

Spinger.

- HEINZ, J. M., and K. N. Stevens. 1964. "On the derivation of area functions and acoustic spectra from Cineradiographic films of speech." *Journal of the Acoustical Society of America* 36 (1):1037-1038.
- ISKAROUS, K. 2005. "Patterns of tongue movement." *Journal of Phonetics* no. 33 (4):363-381.
- LINDBLOM, B., and J. Sundberg. 1970. "Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement." *Journal of the Acoustical Society of America* 48 (1):120-120.
- LUO, Changpei 羅常培 1956. *Hanyu yinyunxue daolun* 漢語音韻學導論 (Introduction to Chinese phonology). Beijing 北京: Zhonghua Shuju 中華書局
- MAEDA, S. 1982. "A digital simulation method of the vocal-tract system." *Speech Communication* 1 (3-4):199-229.
- MERMELSTEIN, P. 1973. "Articulatory model for the study of speech production." *Journal of the Acoustical Society of America* 53 (4):1070-1082.
- MOKHTARI, P., T. Kitamura, H. Takemoto, and K. Honda. 2007. "Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients." *Journal of Phonetics* 35 (1):20-39.
- PERKELL, J. S. 1969. *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*, M.I.T. press Research Mono. No. 53. Cambridge, MA.: MIT press.
- _____. 1974. "Quasiphysiological Tongue Model." *Journal of the Acoustical Society of America* 55:S78-S78.
- SONDHI, M. M. 1973. "Experimental determination of area function of lossy, dynamically varying vocal-tract." *Journal of the Acoustical Society of America* 53 (1):294-294.
- _____. 1974. "Model for wave-propagation in a lossy vocal-tract." *Journal of the Acoustical Society of America* 55 (5):1070-1075.
- STEVENS, K. N. 1989. "On the quantal nature of speech." *Journal of Phonetics* 17 (1-2):3-45.
- STONE, M., D. Dick, A. S. Douglas, E. P. Davis, and C. Ozturk. 2000. Modelling the internal tongue using principal strains. In *Proceedings of the 5th Seminar on Speech Production: Models and Data* 133-136. (Seminar at Bavaria,

Germany)

- STONE, M., M. H. Goldstein, and Y. Q. Zhang. 1997. "Principal component analysis of cross sections of tongue shapes in vowel production." *Speech Communication* 22 (2-3):173-184.
- STORY, B. H., I. R. Titze, and E. A. Hoffman. 1996. "Vocal tract area functions from magnetic resonance imaging." *Journal of the Acoustical Society of America* 100 (1):537-554.
- SUNDBERG, J., C. Johansson, H. Wilbrand, and C. Ytterbergh. 1987. "From sagittal distance to area - a study of transverse, vocal-tract cross-sectional area." *Phonetica* 44 (2):76-90.
- WANG, Gaowu. 2010. An articulatory model of vocal tract in Mandarin. Ph.D. Diss., Peking University (Dept. of Chinese Linguistics and Literature).
- WANG, Gaowu, J. W. Dang, and J. P. Kong. 2008. Estimation of vocal tract area function for Mandarin vowel sequences using MRI. Paper presented at INTERSPEECH 2008, Sep. 22-26, at Brisbane, Australia.
- WANG, Gaowu, T. Kitamura, X.G. Lu, J. W. Dang, and J. P. Kong. 2008. "MRI-based study of morphological and acoustical properties of Mandarin sustained steady vowels." *Journal of Signal Processing* 12 (4):311-314.
- WILHELMS-TRICARICO, R. 1996. "A biomechanical and physiologically-based vocal tract model and its control." *Journal of Phonetics* 24 (1):23-38.
- WU, Zongji 吳宗濟, and Maocan LIN 林茂燦. 1989. *Compendium of Experimental Phonetics* 實驗語音學綱要 Higher Education Press 高等教育出版社.
- ZHOU, Dianfu 周殿福, and Zongji WU 吳宗濟. 1963. *The Articulation Album of Mandari* 普通話發音圖譜. Beijing 北京: The commercial Press 商務印書館.
- ZU, Yiqing 祖漪清. 1983. A primary study of synthesizing vowels with articulatory parameters 聲道面積函數和共振峰頻率的初步報告 In *Report of Phonetic Research, Phonetics lab, CASS* 《語音研究報告 1983-1984》, 中國社會科學院語言研究所.
- WU, Zongji 吳宗濟, and Maocan LIN 林茂燦. 1989. *Shiyan yuyinxue gangyao* 實驗語音學綱要 (Compendium of Experimental Phonetics) Beijing: Gaodeng Jiaoyu Chubanshe 高等教育出版社.
- ZHOU, Dianfu 周殿福, and Zongji WU 吳宗濟. 1963. *Putonghua fayin tupu* 普通話發音圖譜 (The Articulation Album of Mandari). Beijing 北京: Shangwu Yishuguan 商務印書館.

ZU, Yiqing 祖漪清. 1983. Shengdao mianji hanshu he gongzhengeng pinlv de chubu baogao 聲道面積函數和共振峰頻率的初步報告(A primary study of synthesizing vowels with articulatory parameters) In Yuyin yanjiu baogao 語音研究報告 1983-1984, Zhongguo Shehui Kexueyuan Yuyanxue Yuyin Yanjiushi 中國社會科學院語言研究所語音研究室.

用磁共振成像和 X 光聲道資料建立漢語普通話調音模型

汪高武* 孔江平†

北京師範大學 北京大學

北京大學

提要

爲了更好地理解言語的產生過程，即如何從音位輸入到調音器官的動作、再到聲學輸出，需要建立詳盡的聲道調音模型。本文是對漢語普通話言語產生中調音機制的探索與研究，建立了一個具有視覺和聲學輸出的幾何調音模型。該調音模型的數據源自於聲道的磁共振圖像和 X 光錄影，前者主要提供聲道的立體形狀，後者提供調音的動態過程。在這個模型中，聲道被分解為七個調音部位進行研究：硬齶、喉腔後壁、下頷、雙唇、軟齶、舌頭和喉管。其中創新性地，舌頭又被分為舌體（相對簡單）和舌尖（更為靈活）兩部份，從而簡化所需要的參數；另外，普通話四聲中基頻高低與喉管上下高度的關係也加入到模型中。該模型可以作為研究工具服務於語言學、語音學和言語工程，並可用於語音參數合成、虛擬說話人、普通話輔助教學等領域。

關鍵詞

言語產生 調音模型 聲道 磁共振成像 X 光

* 汪高武執教于北京師範大學并治學于北京大學中文系語言學

† 孔江平爲北京大學教授執教研究于中文系，中國語言學研究中心，語言與人類複雜系統聯合研究中心，并爲本文聯繫作者：中國北京，郵編 100871，北京大學中文系；[jpkong@pku.edu.cn]

PROSODIC BOUNDARIES EFFECT ON
SEGMENT ARTICULATION IN STANDARD CHINESE:
AN ARTICULATORY AND ACOUSTIC STUDY

Yinghao Li*

Yanbian University, Yanji (China)

ABSTRACT

This paper presents an electropalatographic (EPG) and acoustic study of prosodic boundaries effect on the domain-initial segments in Standard Chinese.¹ Two speech sounds, namely, the voiceless unaspirated alveolar stop /t/ and the high front vowel /i/, were studied to examine the domain-initial strengthening in both spatial and temporal dimensions. The articulatory and acoustic parameters of the speech sounds were compared in initial positions of five prosodic constituents in Standard Chinese, namely, a Syllable, a Foot, an Immediate Phrase, an Intonational Phrase, and an Utterance. The results show that: (1) the production of the domain-initial consonantal gesture was prosodically encoded. The linguopalatal contact and the seal duration varied as a function of the prosodic boundary strength. The linguopalatal contact was dependent on the seal duration in a nonlinear fashion. Of the acoustic properties of the domain-initial stop, the total voiceless interval and voicing during closure were found to be reliable acoustic correlates that mark the hierarchical structure of the prosody. (2) At the release moment of the domain-initial stop, no consistent pattern was found to support the domain-initial strengthening. The linguopalatal contact of the vowel immediately following the domain-initial consonant did not show a clear trend of domain-initial strengthening; however, the phonatory features of vowels were indicative of pitch reset at major prosodic boundaries. These

* **Yinghao Li** 李英浩 is associate professor in the Foreign Languages College of Yanbian University. Address: Foreign Languages College, Yanbian University, Yanji, Jilin Province 133002, China; [youngho_lee@aliyun.com]

indicate that the domain-initial strengthening is restricted on the segment immediately following the boundary. In conclusion Standard Chinese strengthens the phonetic features of the domain-initial segments as a function of boundary strength, which serves as an important way to mark prosodic structure in Standard Chinese.

SUBJECT KEYWORDS

Prosodic boundaries Segment articulation Electropalatography
Standard Chinese

1. INTRODUCTION

The production of phonological units in utterances is subject to the conditioning of the prosodic structure. One way this structure information is encoded is how a phonological unit is produced at edges of prosodic constituents. In recent years, an increasing number of studies have observed that the segments at prosodic domain-initial position are produced more strongly than at domain-medial position. The study of this boundary-induced articulatory variation for individual speech segments has become an important topic called the domain-initial strengthening. The domain-initial strengthening refers to the greater magnitude of phonetic realization in the articulatory or acoustic dimension of a phonological unit at the initial position of prosodic constituents. The well-attested assertion that “the stronger the position, the stronger the articulation” (Cho and Keating 2001, 156) has been tested in various languages, namely, English (Fougeron and Keating 1997; Byrd and Saltzman 1998; Byrd 2000; Cho 2001; Cho and Keating 2009; Keating et al 2003), French (Fougeron 2001), Dutch (Cho and McQueen 2005), Korean (Cho and Keating 2001), German (Bombien et al 2010; Kuzla and Ernestus 2011), Taiwan Hokkien (Hsu and Jun 1998; Hayashi, Hsu, and Keating 1999; Keating, et al 2003), and Standard Chinese (Cao and Zheng 2006; Li and Kong 2011).

In previous studies the prosodic shaping of features of a segment has demonstrated a gradient variation in the articulatory and/or acoustic dimension as a function of the boundary strength. In the electropalatographic (EPG) studies on consonants in various languages, the peak linguopalatal contact and the articulatory seal duration were

progressively greater at stronger prosodic boundaries (Fougeron and Keating 1997; Fougeron 2001; Cho and Keating 2001, 2009; Keating et al 2003; Li and Kong 2011). In the acoustic domain Jun (1993) found that the voice onset time (VOT) for Korean aspirated stop was progressively shorter as the boundary strength becomes weaker. The reverse was the case for the prosodic positional effect on the VOT of fortis plosives in German (Kuzla and Ernestus 2011). The gradient variations as a function of domain strength has been considered to be related to the duration-dependent undershoot model as postulated by Lindblom (1990). That is to say, the time consumes the production of a segment determines how much the phonetic target is realized. The segment is probably hypoarticulated at weaker position because insufficient time is consumed. The linguistic motivation for this articulatory adjustment is hypothesized to be attributed to enhanced syntagmatic contrasts from the neighbouring segments through magnifying the associated features of the segments (Fougeron and Keating 1997; Hsu and Jun 1998; Cho 2005).

The prosodic strengthening effect is also temporally constrained. The scope for the domain-initial strengthening effect tends to be manifested at the first post-boundary segment, whereas the following segment in the same syllable is seldom affected (Fougeron and Keating 1997; Cho 2005; Byrd, Krivokapic, and Lee 2006). In an electropalatographic study Fougeron and Keating (1997) found that there was no consistent and reliable pattern of the linguopalatal contact for non-initial /o/ in the domain-initial syllables compared with domain-initial consonant /n/. Cho (2005) examined boundary-induced articulatory and acoustic variation of post-boundary vowels in domain-initial CV syllable in English. No consistent tongue fronting or raising was found for the high front vowel /i/, nor was tongue lowering and backing found for low back vowel /a/. Byrd, Krivokapic, and Lee (2006) investigated the temporal scope of prosodic boundary effect on the segment articulation. They found that the articulatory gesture of the post-boundary consonant was significantly strengthened with a longer duration and larger articulatory displacement. The following consonants were temporally shortened, accompanied by smaller articulatory displacement. However, some studies have found evidence of the manifestation of the domain-initial strengthening effect on non-initial

vowel in post-boundary CV syllable (Farnetani and Vayra 1996; Cho and Keating 2009; Kim and Cho 2011). In their electropalatographic study on prosodic effect on the production of CV syllables, Farnetani and Vayra (1996) found that the initial syllable was altogether strengthened and the vowel in the domain-initial syllable has a more open vocal tract, regardless of lexical accent condition. In the electropalatographic and acoustic study on the boundary effect on English segment production, Cho and Keating (2009) found mixed results for the locality hypothesis. They found that the articulatory parameters for V in post-boundary CV syllables were insensitive to the boundary strength; however, the vowel amplitude did show strengthening at utterance-initial position compared with that for lower boundaries. In the electromagnetic articulograph (EMA) study on the domain-initial strengthening effect, Kim and Cho (2011) found that the vowel following /h, p^h/ at the initial position of prosodic boundaries showed gradient tongue movement magnitude as affected by the boundary strength, which was rather similar with the case for the domain-initial vowel in the syllable. The above findings indicate that the domain-initial strengthening effect might interact with other confounding factors in affecting the post-boundary initial syllables. In a simulation study, Byrd and Saltzman (2003) showed that speech production mechanism controlling articulator(s) gets increasingly longer time when approaching a prosodic boundary, and this gestural actualization time becomes shorter when the boundary recedes. That is to say, the articulator(s) can be orchestrated to fully realize the phonetic target if provided with sufficient gestural preparatory time, and this timing mechanism speeds up immediately after the boundary.

If the domain-initial strengthening only affects the first segment after the prosodic boundaries, it indicates that the vowel in the post-boundary syllable is subject to the shaping of another prosodic mechanism instead of the domain-initial strengthening. In a tonal language such as Chinese, syllables are specified with lexical tones, which contrast meanings. In previous studies the syllable tones of Mandarin was proved to be hierarchically determined, which means that the tone specification is encoded according to the boundary strength (Tseng et al. 2005). Nevertheless, this hierarchically structured tonal specification is the manifestation of f₀ reset, which is another prosodic

means to mark the prosodic boundaries in Standard Chinese. In recent studies it was found that voice quality may also be affected by the prosodic structure. Electroglottographic (EGG) studies have shown that voice quality of the post-boundary vowel became progressively breathier at higher prosodic boundaries. For the gradient variation of voice quality measures, it was assumed to be related to f_0 reset at major boundaries (Garellek 2014).

Previous studies have shown that the domain-initial strengthening effect was dependent on the segmental identity and language. Regarding the segmental identity, it was found that the alveolar fricative is more resistant to articulatory variation as a function of prosodic positions. For example, in the electropalatographic study on the domain-initial strengthening effect on the production of the French alveolar fricative, Fougeron (2001) found that the linguopalatal contact pattern for the alveolar fricative was less subject to the prosodically-conditioned articulatory variations as evidenced in the production of alveolar stops. A similar result was also found in Korean in that the domain-initial strengthening effect on the articulatory magnitude for the phonological units was not as salient for fricatives as for stops (Kim 2001). This segment-specific response to the prosodic conditioning is explained away as the aerodynamic requirements for producing fricatives which results in rather rigid tongue gestures. In terms of the language-specific manifestation for the domain-initial strengthening effect, Cho and Keating (2001, 2009) found that the VOT for Korean lax and aspirated alveolar stops /t, t^h/ and English /t/ was increased in higher prosodic boundaries. However, the VOT variations as incurred by the prosodic positions are in the inverse direction for /t/ in Dutch and German, with shorter VOT appearing at stronger boundaries (Cho and McQueen 2005; Kuzla and Ernestus 2011). These results concerning the positional-dependent VOT difference involve the different laryngeal gestures across languages. The prosodic signature on segment articulation may be represented by different phonetic dimensions. In the articulatory study on the Tamil language, the duration and timing relationship in consonant clusters was affected by prosodic structure, but no such effects were found for the consonant articulatory magnitude (Byrd et al 2000). In short, the segment- and language-specific prosodic signature on the segment articulation is

attributed to the fine-grained articulatory control, on the one hand, and cross-language differences in segment articulation, on the other.

The current research investigated the effect of prosodic structure on the production of individual speech segments at prosodic domain-initial positions in Standard Chinese. By prosodic structure, we mean that a spoken utterance is hierarchically organized with the higher prosodic domain being decomposed into immediately lower constituents. Following the Strict Layer Hypothesis (Selkirk 1984), the utterance is shaped by the hierarchical prosodic structure in that the higher prosodic domain directly dominates one or more immediately lower prosodic domains, and a given prosodic domain must be contained by an immediately higher prosodic domain. The prosodic model for Standard Chinese used in the current paper follows Li 李 (2002) and Lin (2002) with some minor modifications. Based on the autosegmental-metrical theory, Li 李 (2002) and Lin (2002) assert that the speech utterance of Standard Chinese is hierarchically organized in that the larger prosodic units are composed of several immediately lower prosodic constituents (domains). These constituents include syllable, foot, prosodic word, minor phrase, major phrase, and utterance, which were claimed to be distinguished by the pitch contour and break. For the two prosodic phrases, we use the terms intermediate phrase and intonational phrase for the comparison with the results in other languages. The domain of foot is normally syntactically defined, which constitutes the basis for the higher prosodic domain-prosodic word, and the latter comprises a foot and/or a following unassigned monosyllable (Li 李 2002, 526). In the current paper, the bi-syllabic foot domain is to be studied because it is the basic unit of the metrical organization in Standard Chinese (Wang 2008). The domain of intermediate phrase comprises one or several prosodic words (in this paper we use foot as the immediately subordinate component for intermediate phrase), and is characterized by a phrasal accent and noticeable pause (either silent or filled pause). For intonational phrase, the prosodic cues include longer domain-ending pause and noticeable pitch contour resetting. The utterance, constituted by one or several major phrases, is cued by a sentence accent, substantial durational compression of domain-ending syllable and longest pause (Wang 2008, 257).

Figure 1 shows the prosodic hierarchy used in the current paper.

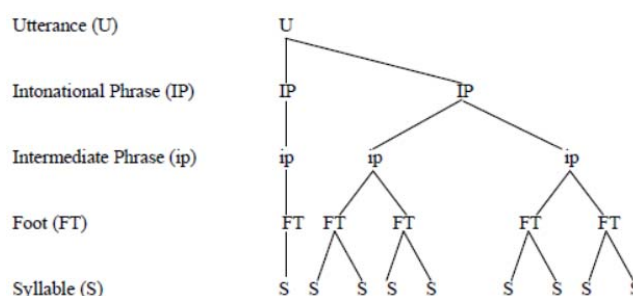


Figure 1 The hierarchy of prosodic structure of Standard Chinese

Extensive research on the prosodic structure in Standard Chinese has shown that the f0 reset, the syllable durational pattern, as well as the pause inside utterances are the three main acoustic correlates cueing the prosodic structure (Li 李 2002; Wang, Yang, and Chen 2004; Hu, Xu, and Huang 2002; Lin 2002). However, the function of the domain-initial strengthening in marking the Chinese prosodic structure has received little attention. In her acoustic study on the segmental lengthening, Cao (2005, 165) suggested:

‘The domain-initial segmental lengthening, likewise, has the function to mark the boundary strength, and this function can not be underestimated. It is more direct and reliable to indicate the prosodic boundary strength. ...meanwhile, the post-boundary consonantal duration is positively correlated with the perceived boundary strength, and it is progressively increased as a function of the boundary strength.’

Keating et al (2003, 161) hypothesized that a lexical tone language such as Taiwan Hokkien might have a more salient domain-initial strengthening effect than English because “it should have less recourse to pitch to mark domain edges”. Although their result did not support the hypothesis, the cumulative effect that was found did show the universality of the prosodic conditioning of domain-initial segments. Wang (2008) argued that the consonant initial inside a foot domain, the basic metrical

template for rhythmic organization in Standard Chinese, was reduced somehow compared with one that heads a foot domain; however, no gestural reduction was found on prosodic constituent initial position above the foot domain. This argument directly supports the tenets of the domain-initial strengthening in that the consonant gesture tends to be undershot in prosodically weak position. However, it is worth investigating whether gradient articulatory gesture exists at the initial position of prosodic boundaries of different strengths.

In the electropalatographic study, Cao and Zheng (2006) found that the linguopalatal contact for the phrase-initial consonant was greater than that for the phrase-medial consonant in Standard Chinese. Li and Kong (2011) investigated the articulatory strengthening phenomenon for domain-initial alveolar stop /t/ in Standard Chinese, and found a cumulative increase of linguopalatal contact as well as alveolar seal duration from syllable to utterance boundaries based on one female speaker's electropalatographic data. In this paper we will extend our previous study by investigating the same consonant, and the tautosyllabic high front vowel /i/ uttered by two speakers. The sole research question to be addressed in the current paper is how the prosodic position affects the production of consonantal and vocalic gestures in CV syllables in Standard Chinese. Two aspects of the domain-initial strengthening are to be covered: (1) Does the prosody affect the articulatory and acoustic properties for the domain-initial segment in the cumulative pattern? (2) Does the domain-initial strengthening affect all segments in the post-boundary syllable?

The universality of the domain-initial strengthening effect leads us to predict that the segment at domain-initial position is strengthened and that a cumulative scale exists, which is indicative of the linguistic encoding of the segmental production rather than an intrinsic biomechanical process. As a matter of fact, previous studies with one speaker showed a strong tendency of this prosodically conditioned segment articulation in that the linguopalatal contact for unaspirated /t/ was positively correlated with the prosodic hierarchy. The second prediction is concerned with the temporal scope of the domain-initial strengthening effect. It is hypothesized that the domain-initial strengthening effect scope is limited in the domain-initial segment, and the following vocalic segment is not subject to the domain-initial effect.

The domain-initial strengthening quickly fades away toward the end of the initial segment production, as predicted by Byrd and Saltzman (2003).

2. METHOD

2.1 Electropalatography (EPG)

The tongue-palate contact for lingual consonants captured by electropalatography is a reliable indicator of articulatory magnitude. Thus, more linguopalatal contact indicates more oral constriction and greater articulatory magnitude. WinEPG Electropalatography produced by Articulate Instruments was used to capture the tongue-palate contact signal. WinEPG uses custom-made pseudo-palates of thin acrylic base. The pseudo-palate covers the palate from root of upper teeth to the anterior portion of the soft palate. The layout of the 62 electrodes is designed to be relative to the anatomical landmarks in order to carry out inter-subject comparison. When the tongue contacts an electrode, the circuit is completed and the signal is recorded by the WinEPG system.

2.2 Speech Stimuli

The test segments were the voiceless unaspirated alveolar stop /t/ and the high front vowel /i/. The stop was placed at domain-initial positions of five prosodic domains, namely, syllable (SYL), foot (FT), intermediate phrase (ip), intonational phrase (IP), and utterance (U); and the stop was placed in two symmetrical vocalic environments, low vowel /a/ and high front vowel /i/. The high front vowel /i/ in syllable /ti/ was the other test segment for investigating the temporal scope of the domain-initial strengthening. The syllable was preceded by the five prosodic domains immediately before the initial stop, and was followed by syllables starting with the alveolar stop /t/. For the tonal specification of the test syllables, the low tone (T3) was avoided as much as possible because it would possibly affect the linguopalatal contact of /i/ (Hoole and Hu 2004). The falling tone was used in most cases for examining the vowel-initial f0 and voice quality of the vowels with varying boundary conditions.

Table 1 shows the samples of the stimuli of the sentence set for test consonant /t/. The utterance domain was elicited by a full period, and speakers were instructed to make a long pause after it. The comma was used to parse the intonational phrase domain, with the syllable number in

the two sentence components ranging from 6 to 12 syllables. The speakers were instructed to ignore the comma. The intermediate phrase domain was a noun phrase that comprised at least two feet. The foot domain was the second word inside an immediate phrase. The syllable domain was defined as the boundary occurring inside a foot. The five-level break index based on C-ToBI (Li 2002) was used to code the prosodic domains: the break index number 4 (B4) was used to transcribe the Utterance domain, and break index 0 (B0) the Syllable domain. The rest break numbers were used to transcribe the intermediate domains.

Table 1. Sample sentences for the test consonant /t/. (The underscored characters and syllables were the locations where the test consonant /t/ headed the prosodic domains. In the right column of each cell, the stimuli sentence was given first, and the phonetic transcription between slash brackets and with the literal translation in parenthesis following.)

U-initial	国家对高等教育的投入加大。<B4>大学教师的工资不断提高。 /kuo35 tɕia55 tɕei51 kau55 tən214 tɕiau51 y51 də0 tʰou35 zu51 pu35 tuan51 tɕia55 ta51. ta51 ɕye35 tɕiau51 ʂɿ55 də0 kuŋ55 tsɿ55 pu35 tuan51 tʰi35 kau55/ (The state increases its input in higher education. The salary for university teachers is increasing.)
IP-initial	这一地带昼夜温差大，<B3>大麦颗粒饱满。 /tɕɿ51 i35 tʰi51 tai51 tɕou51 ie51 uən55 tɕʰa55 ta51, ta51 mai51 kʰɿ55 li51 pau35 man214/ (The temperature difference is dramatic between day and night in this area, thus the barley seed appear plump.)
ip-initial	负债重重的加大<B2>大力削减学生的奖学金计划。 /fu51 tɕai51 tɕʰuŋ35 tɕʰuŋ35 də0 tɕia55 ta51 ta51 li51 ɕy55 tɕien214 ɕye35 ʂən55 də0 tɕiaŋ214 ɕye35 tɕin55 tɕi51 xua51/ (The heavily-indebted UC cut down the student scholarship program.)
FT-initial	华夏<B1>大地充满了勃勃的生机。 /xua35 ɕia 51 ta51 ti51 tɕʰuŋ55 man214 lə0 po35 po35 də0 ʂən55 tɕi55/ (The China land is full of vigor.)
SYL-initial	负债重重的加<B0>大大力削减学生的奖学金计划。 / fu51 tɕai51 tɕʰuŋ35 tɕʰuŋ35 də0 tɕia55 ta51 ta51 li51 ɕy55 tɕien214 ɕye35 ʂən55 də0 tɕiaŋ214 ɕye35 tɕin55 tɕi51 xua51/ (The heavily-indebted UC cut down the student scholarship program.)

2.3 Speakers and Procedures

Two university students (one male and one female) participated in the experiment. They had lived in a northern province or municipality in China before entering into university. After enrolling in the university they had worked as part-time announcers at the university TV station. At the time of the recording session they were 27 years old and were paid for their participation.

The recording was conducted in a sound-attenuated booth at Peking University. Before the recording session the participants were given 30~50 minutes to adapt to the reading task with a pseudo-palate installed in their mouths and to familiarize themselves with the sentences. In the recording session, the sentences were randomized blocked and each sentence was repeated three times for the female speaker and five times for the male speaker. The sentence blocks were presented on the computer screen which was positioned about one meter in front of them, and they were instructed to read the sentence list at a normal speech rate. No specific instructions on prosodic phrasing were given, except that they were required to make an intentional pause after an orthographic period in the sentence. The electropalatographic, electroglottographic, and the speech signal were simultaneously recorded into a computer (see Figure 2 for an example). The sampling rate for EPG signal was 100 Hz, and that for speech and EGG signal was 22 kHz. For the male speaker the respiratory signals for chest and stomach breathing were also recorded, which will not be analyzed in the present paper. After the recording the sentences that had unclear pronunciation of the test segments or signal problem were eliminated. In the end a total of 264 sentences were submitted for further analysis.

The articulatory and acoustic analysis was carried out on the Matlab program developed to process the electropalatographic and speech signals. First, the electropalatographic signal was temporally aligned with the acoustic signal by using the algorithm in Li and Pan (2012). Then each utterance was parsed from the recording item and annotated at PRAAT (Boersma 2001) for two tiers, the syllable tier and the break index tier. Figure 3 demonstrates the break index tier in this paper.

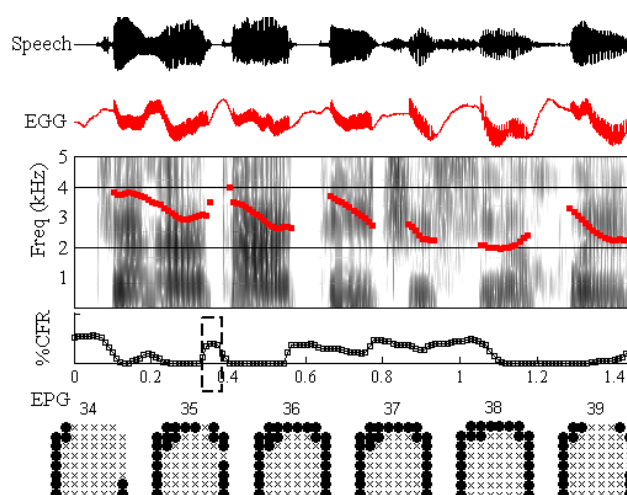


Figure 2 The simultaneously-recorded speech signal, EGG and EPG signals. (From top to bottom: speech signal, EGG signal, the spectrogram with the f0 contour plotted on it, one EPG contact measures, namely, % contact of the Front Region (defined later), and six consecutive EPG frames that show the alveolar closure gesture between 0.33 and 0.38 second. The filled dots indicate tongue-palate contact, and the “x” no contact.)

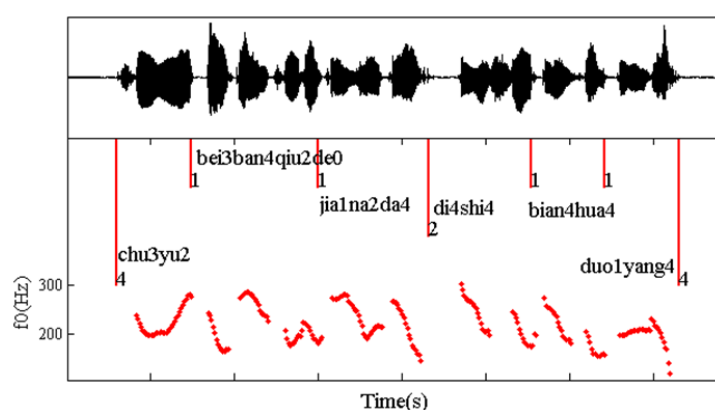


Figure 3 The break index tier of the utterance “处于北半球的加拿大地势变化多样 (Canada, located in the northern hemisphere, has diversified geographic features.)”. (The numerals on the right of the line indicate break index number. The phonetic annotation is in Chinese *pinyin*, and the numerals following each syllable stand for the tones, 1 for high level tone, 2 for rising tone, 3 for low tone, 4 for falling tone, and 0 for neutral tone.)

2.4 Articulatory Measurement

The articulatory gestures of /t/ and /i/ have distinct tongue-palate contact pattern. In Figure 4 we defined the first four rows as the Front Region, which was closely related with the closure-formation of the alveolar stops. The linguopalatal contact at the back four rows was closely related with the coarticulatory effect of the following vowel on the preceding consonant, which was not studied in this paper.

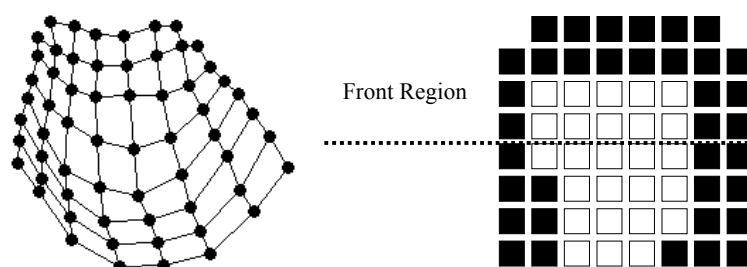


Figure 4 The three-dimensional illustration of the placement for 62 electrodes on pseudo-palate (left) and the division of the two regions of the electrodes (right). The linguopalatal contact on the right chart is the point of maximum contact frame for unaspirated /t/ in syllable /ti/. The shaded squares refer to the contact captured by the WinEPG system.

To study the boundary effect on the post-boundary alveolar stop, two key time points for tongue-to-palate contact were defined: one was the point of maximum contact frame (PMC) and the other the release frame. The PMC was defined as the frame that had the maximal linguopalatal contact during the alveolar closure interval (when the complete alveolar closure was observed), or acoustic closure interval (when no complete alveolar closure was found). The contact pattern at the PMC is considered to directly reflect the magnitude of articulatory excursion for the segment in question (Cho and Keating 2009). The release frame was defined as the last frame of the closure interval of the alveolar stop. If no alveolar closure was observed in the acoustic closure interval (normally in FT or SYL domain), the release frame was defined as the frame immediately before the acoustic release of the stop. In case the alveolar stop was realized as a voiced approximant (see Cho and Keating 2001), no release frame was available. As argued by Cho and Keating (2009, 470), the release frame “might reveal

additional information about the prosodically-conditioned articulatory variation”. For the two key frames, the percent of contacted electrodes in the Front Region was computed, for this region was related with the tongue tip/blade gesture. The seal duration (SD) was taken as the interval between the first and last frame of the alveolar closure. In case no closure frames were found or the alveolar stop was realized as voiced approximant, the seal duration was zero.

To test the locality hypothesis of the domain-initial strengthening, we investigated the linguopalatal contact of the high front vowel /i/ in the domain-initial syllable /ti/. In the previous studies, the high front vowel /i/ was shown to be resistant to the coarticulatory effect, regardless of the boundary strength (Cho 2004). Provided that the boundary effect is restricted at the initial position of the post-boundary syllable, the vowel /i/ in the syllable /ti/ would not show strengthening effect. To reduce the confounding domain-final strengthening/lengthening effect, the syllable /ti/ was designed to be followed by a Syllable boundary. However, the Syllable-initial tokens were followed by a Foot boundary because the bi-syllabic foot was used to construct the test sentences as much as possible. In this case, the syllable was assigned to head a following foot domain. The maximal linguopalatal contact frame for /i/ was selected between the one third and half toward the vocalic interval. By doing so, the coarticulatory effect of the preceding and following segments of the vowel was maximally reduced. The percent of contacted electrodes in the whole region was computed to measure the articulatory magnitude of /i/.

Besides the linguopalatal measure for vowel /i/, we also investigated the vowel-initial f0, open quotient (OQ), and speed quotient (SQ) of the vowels /i/ and /a/, for they tended to be conditioned by the prosodic structure. As shown in Introduction, the f0 manifests the f0 reset in Standard Chinese. The OQ and SQ reflect the voice quality of vowels. The OQ is the inverse of the contact quotient (or closed quotient), which shows the portion of time when the vocal folds are opened in each glottal period. Higher OQ is related with larger glottis opening. The SQ computes the time portion ratio between the opening phase and closing phase as defined in EGG signal in each glottal period. The SQ is associated with how fast the vocal folds adduct. The higher the SQ is, the faster the vocal folds adduct. A recent study on the voice

quality strengthening in English and Spanish had indicated that vowels after the domain-initial glottal stops showed lower contact quotient, or higher OQ, at higher prosodic boundaries (Garellek 2014). In the current paper the f_0 was computed based on the derivative of the EGG signal, and the OQ and SQ were obtained by Hybrid method (Howard 1995). Figure 5 shows the definition of critical moments and intervals, and three equations to obtain the measures.

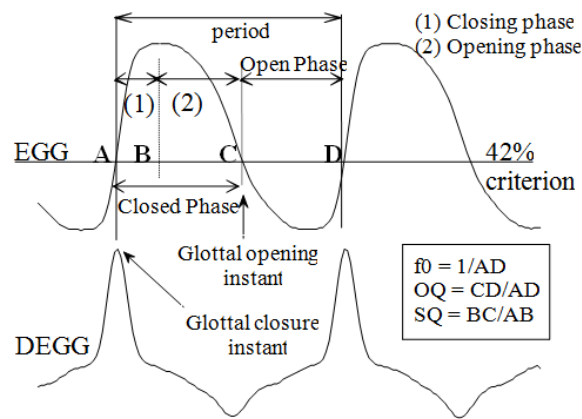


Figure 5 The definition of critical moments and intervals, and three equations to obtain f_0 , OQ and SQ.

2.5 Acoustic Measurement

As the first step, the durational properties of syllables were measured to check whether the utterances were produced appropriately for the current study. One important acoustic parameter that marks the prosodic structure is the pre-boundary vocalic duration (or V1 duration). This measure is used to demonstrate the final lengthening, which refers to the durational variations of the rhyme in pre-boundary syllable (Wightman et al 1992). In this paper the V1 duration was defined by the observable F2 formant trajectory in the pre-boundary vowel. Another acoustic parameter that indexes the prosodic structure is the acoustic closure interval. In this paper it was taken as the interval started from the

termination point for F2 contour in the previous vowel to the stop burst for the following consonant /t/. The acoustic closure duration would not be measured at the utterance-initial position because the better part of the silent interval was not the result of the alveolar closure gesture. These two acoustic measurements were to be compared with the previous results of boundary-induced lengthening.

The acoustic properties of the test segments were investigated because if the domain-initial articulatory strengthening was demonstrated in the acoustic domain, the listeners would possibly pick up those acoustic attributes for phrasing the utterances, and the research of the perceptual relevance of the domain-initial strengthening can be carried out in future. The acoustic measures included: (1) Voice onset time (VOT) of the voiceless unaspirated alveolar stop. The VOT was measured from the point of the stop release to the voice start of the following vowel, signalled by the onset of F2 trajectory in the spectrogram. In case of approximant realization or voice stop, the VOT was zero. (2) Voicing during stop closure and total voiceless interval. These two measures were computed to partially demonstrate the vocal folds state during the oral closure interval (Cho and Keating 2001). The vocal folds state during the oral occlusion of the alveolar stop might as well reflect the conditioning of boundary strength. The voicing during stop closure was the percent of the voicing of the acoustic closure duration. The voicing was represented by the voicing bar at low-frequency of the spectrogram, or cyclical abduction-adduction of vocal folds in EGG signal. The total voiceless interval was the duration of the silent interval in the acoustic closure duration plus VOT. No measurement was conducted at the Utterance-initial position for the two measures. (3) RMS burst energy. Following Cho and Keating (2001) the RMS burst energy for /t/ was defined as the acoustic energy at the burst through calculating the RMS value of frequencies above 500Hz of an FFT spectrum. The low frequency was eliminated because the possible effect of the voicing over the stop release.

3. RESULT

3.1 The Prosodic Hierarchy Indicated by the Utterances

To ensure that the speech material was appropriate for the further analysis, two durational measures were collected and compared with the previous study regarding the acoustic correlates of the prosodic structure in Standard Chinese.

Although the function of the pre-pausal lengthening in indicating the prosodic boundaries is a well-accepted notion, the case for Standard Chinese is far from clear. Cao (2004, 2005) argued that the pre-boundary rhyme lengthening as conditioned by prosodic hierarchy was mainly found at the prosodic phrase level but not at the sentence and paragraph level. Qian, Chu, and Pan (2001) also found that pre-boundary syllable lengthening only existed at the right edges of immediate and intonational phrases, and this effect was unstable at prosodic word level. In their acoustic investigation on the durational pattern of the pre-pausal rhyme, Wang, Yang, and Chen (2004) found the temporal expansion was salient at the lower prosodic boundaries, such as prosodic word and prosodic phrase, whereas its function in marking the prosodic boundary strength became less important in higher prosodic domains when the pause and f_0 reset were the two salient acoustic correlates for marking prosodic boundaries, though at this time the cumulative final lengthening was still observable.

Figure 6 shows the average pre-boundary V1 duration produced by the two speakers. High agreement was achieved regarding the syllable durational pattern domain-finally. For both speakers, the V1 duration was significantly longer at domain-final positions in higher prosodic domains (ip, IP, and U) than in lower ones (SYL, FT). Meanwhile, it was shorter at FT-final position than at SYL-final position, and the difference between the two for the male speaker reached significant level. For the female speaker, the V1 duration was shorter at U-final position than at IP- or ip-final positions, though no significant difference was found. These results basically confirm the findings by Cao (2004, 2005) and Qian, Chu, and Pan (2000). On the one hand, the rhyme at the ip- or IP-final positions tends to be lengthened compared with that at SYL- or FT-final positions, and no cumulative final lengthening is found, on the other.

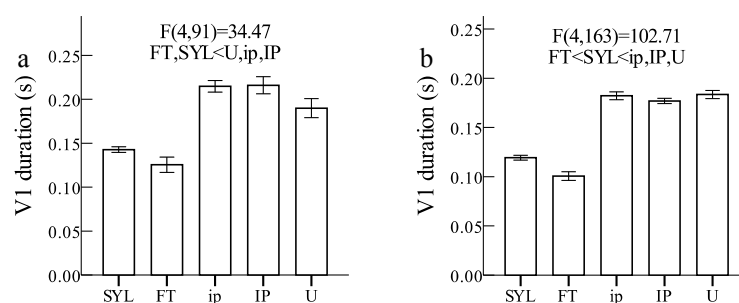


Figure 6 Pre-boundary average V1 duration for female speaker (a) and male speaker (b). The error bar stands for one standard error.

Figure 7 shows the acoustic closure interval at four prosodic boundaries. As predicted, the acoustic closure interval varied depending on the strength of the prosodic boundaries. For both speakers, the longer acoustic closure interval appeared at higher prosodic boundaries. The female speaker produced much longer closure interval at IP- and ip-initial positions, which was attributed to the slower speech rate when the female speaker produced the speech stimuli.

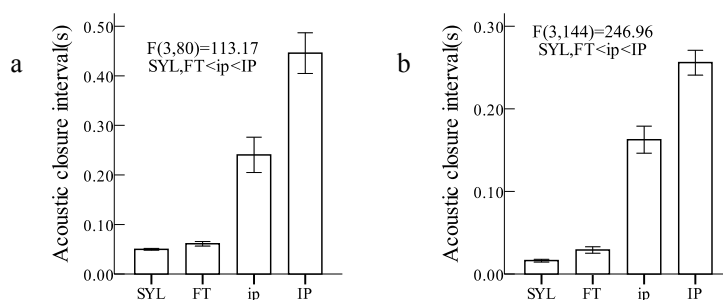


Figure 7 The acoustic closure interval of /t/ at domain-initial positions for (a) female and (b) male speaker. (The measurement at the U-initial position was excluded because it was not the result of alveolar closure gesture on the whole.)

3.2. Articulatory Measures

3.2.1 Linguopalatal contact and seal duration

The one-way analysis of variance (ANOVA) was separately conducted for the articulatory measures produced by the two speakers.

Table 2 shows the results of the ANOVA and the post-hoc comparisons of the articulatory measures in two vocalic contexts. Figures 8 shows the means with one standard error for the linguopalatal contact at two frames.

Table 2 The ANOVA and multiple comparison (Bonferroni method, $p<0.05$) results for the articulatory measures for /t/. Direction of the difference in the measures is indicated by “<” for less percentage or length. The symbol ※ indicates trend effect ($p<0.1$).

Measure	Female speaker	Male speaker
Context a		
% contact(PMC)	F(4,67)=29.91, $p<0.0001$ SYL, FT<ip, IP, U	F(4,118)=65.71, $p<0.0001$ SYL, FT<ip, IP, U
% contact(Release)	F(4,67)=5.48; $p<0.0001$ SYL, FT<ip, U<ip	F(4,118)=12.25, $p<0.0001$ SYL, FT<ip, IP, U
Seal duration	F(4,67)=44.51, $p<0.0001$ SYL, FT<ip<IP, U	F(4,118)=107.81, $p<0.0001$ SYL, FT<ip, IP, U
Context i		
% contact(PMC)	F(4,19)=10.24, $p<0.0001$ SYL<ip, IP, U	F(4,40)=7.64, $p<0.0001$ SYL, FT<ip, IP, U
% contact(Release)	F(4,19)=10.04, $p<0.0001$ SYL<ip, IP, U	F(4,40)=4.29, $p<0.01$ SYL<ip, IP; FT<ip, IP※
Seal duration	F(2,15)=75.34, $p<0.0001$ SYL<FT<ip	F(4,40)=16.08, $p<0.0001$ SYL, FT<ip, IP, U

Generally speaking, the variation of the linguopalatal contact at the PMC and the release frame, and the alveolar seal duration, was larger and longer at higher prosodic constituents than at lower ones. Figure 8 presents the linguopalatal contact of /t/ at the PMC and the release frames. At the PMC frame, a clear tendency of cumulative increase of linguopalatal contact as a function of the boundary strength was found in both vocalic contexts for the female speaker, who distinguished two types of boundaries in /a/ context, namely, lower boundaries of SYL and FT, and higher boundaries of ip, IP, and U. In /i/ context four out of five boundaries were distinguished. For the male speaker, the boundary strength as indicated by the linguopalatal contact at the PMC frame falls into two categories, lower boundaries (SYL and FT), and higher

boundaries (ip, IP, and U). A close look at Figure 8 (b) shows that the linguopalatal contact at SYL-initial position is slightly larger than that at FT-initial position in /i/ context, and the peak contact at IP-initial is the modestly largest among the three higher boundaries.

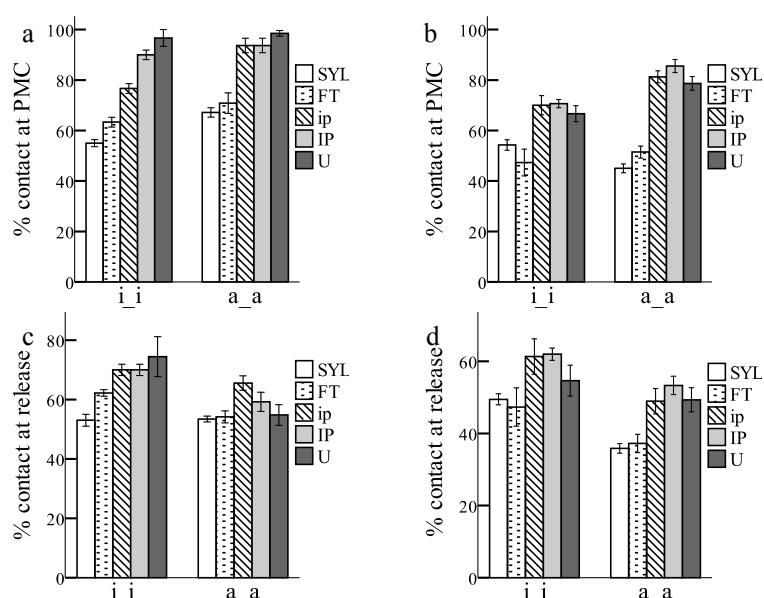


Figure 8 The linguopalatal contact (%) in the Front Region at the PMC and release frame. (The left column is for the female speaker, and the right one male speaker.)

The linguopalatal contact pattern at the release frame was similar to that at the PMC frame except in the /a/ context produced by the female speaker. For the female speaker, the linguopalatal contact varied as a function of the boundary strength in the /i/ context, but no such relation was found in the /a/ context in which condition the significantly lower linguopalatal contact was found at the U-initial position compared with that at the ip-initial position. For the male speaker, the SYL and FT boundaries were distinguished from the higher domains of the ip and IP in the /i/ context, whereas the case in /a/ context was identical to the result obtained at the PMC frame. The linguopalatal contact obtained at the release frame might as well be suggestive of the boundary effect on the segment articulation; however, this effect might fade away toward the plosive release.

Figure 9 shows the sample tokens of linguopalatal contact of /t/ at five domain-initial positions, taken from the PMC frame. As clearly seen in the two rows, the linguopalatal contact in the Front Region, which is closely related with the alveolar closure gesture, is decreasing as the boundary strength becomes progressively weaker, regardless of the vocalic environment. The rightmost frame in the third row shows an incomplete alveolar closure that happened at the SYL-initial position. A careful examination of occurrence ratio of this incomplete alveolar shows that no such tokens existed in the ip-, IP- and U-initial position. For the female speaker 18% tokens have incomplete alveolar closure at SYL-initial position, but no such token is found at FT-initial position. The occurrences of the incomplete alveolar closure produced by the male speaker are 21% and 26%, respectively, at SYL- or FT-initial positions.

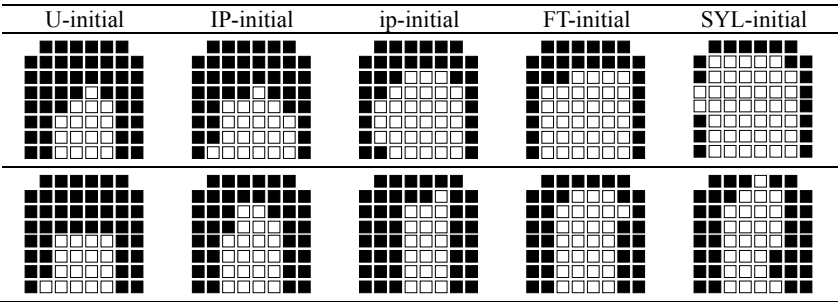


Figure 9 Sample tokens of linguopalatal contact for /t/ produced by the male speaker. (The first row refers to the boundary type. The second and third rows respectively refer to the peak contact frame in /a/ context and /i/ context.)

When measuring the alveolar seal duration, it was not possible to determine the starting point for the alveolar closure in the majority of tokens at IP- or U-initial positions produced by the female speaker in /i/ context. This was because full contact frames that had no linguistic meaning preceded the linguistically meaningful tongue-palate contact in the alveolar region. Thus, the seal duration for /t/ in /i/ context at U- or IP-initial positions was excluded from further analysis. Table 2 and Figure 10 show the alveolar seal duration for the domain-initial alveolar stop. For the female speaker the alveolar closure duration increases as a function of boundary strength in both vocalic contexts. In the /a/ context

three distinctions were identified by the alveolar closure duration. In the /i/ context all three boundaries were distinguished with the alveolar closure duration being shortest at SYL-initial position and longest at ip-initial position. For the male speaker, only two types of boundaries are distinguished: the alveolar closure duration at SYL- and FT-initial positions was significantly shorter than at other higher prosodic domain-initial positions.

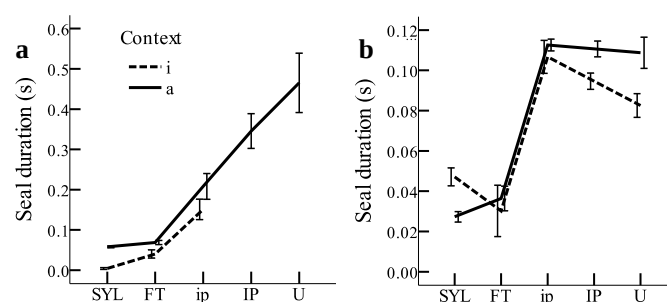


Figure 10 The alveolar seal duration after five prosodic boundaries in two vocalic contexts produced by (a) female and (b) male speakers.

The articulatory magnitude of a consonant at different domain-initial positions might be duration-dependent. Provided sufficient time is given, the consonantal gesture is to be fully realized. In their EPG study on the Korean consonants at different domain-initial positions, Cho and Keating (2001) found that the relationship between the linguopalatal contact for consonants and the seal duration was asymptotic, instead of linear when the linguopalatal contact became larger. Figure 11 shows the scatter plot of the linguopalatal contact in the Front Region against the seal duration, with the curve fitting function that can account for a large portion of variance. As indicated in Figure 11(a, c, d), a polynomial fit was obtained to show that the nonlinear relationship between the linguopalatal contact and the alveolar closure duration. The special case occurred in the /a/ context produced by the female speaker, which shows an exponential relationship between the two measures. A close look at Figure 11b shows the linguopalatal contact increases nearly linearly when the alveolar seal duration was below 0.15 second; however full contact in the Front Region was achieved above 0.15 second. This is an interesting

finding in that a time threshold for a full contact in the front four rows tends to exist. It was observed that below this threshold the alveolar closure started from the first row, or the alveolar ridge, and extended to posterior rows, or the post-alveolar/palatal area, as more alveolar closure time was provided in higher prosodic constituents.

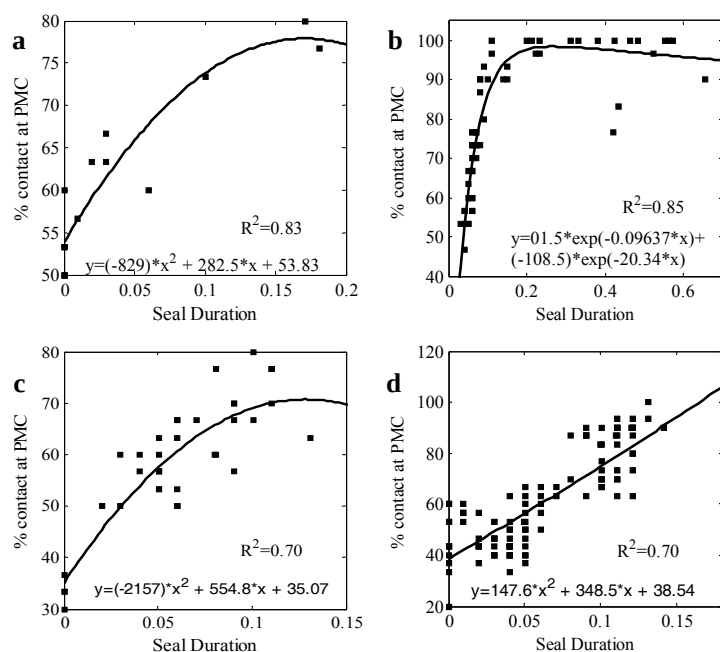


Figure 11 Scatter plot of the linguopalatal contact percent in the front region against the seal duration. (Female speaker, see (a) for /i/ context and (b) for /a/ context. Male speaker, see (c) and (d)).

3.2.2 Vocalic linguopalatal contact

The maximal linguopalatal contact of /i/ in the domain-initial syllable /ti/ was investigated to test the locality hypothesis of the domain-initial strengthening. As predicted in the Introduction, if the domain-initial strengthening only affects the initial segment immediately after the boundary, the articulatory magnitude of /i/ will not vary as a function of the boundary strength. Figure 12 shows the maximal

linguopalatal contact for /i/ uttered by the two speakers.

No clear trend of cumulative boundary effect was found for both speakers. The linguopalatal contact at IP- or U-boundary position was consistently and significantly lower than that at ip-boundary position for the female speaker (Figure 12a) and at FT-boundary position for the male speaker (Figure 12b).

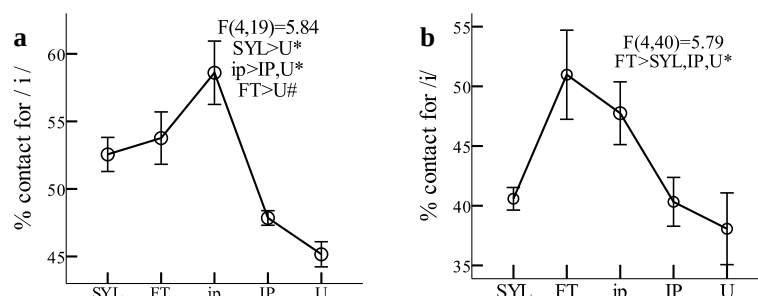


Figure 12 The maximal linguopalatal contact for /i/ in syllable /ti/ after five prosodic boundaries. (female speaker (a) and male speaker (b), * $p < 0.05$, # $p < 0.1$)

3.2.3 F0 and voice quality measures

Figure 13 shows the vowel-initial f_0 , OQ and SQ for vowels /a/ and /i/ in the prosodic domain-initial syllables. The vowel-initial f_0 was consistently higher at the domain-initial position of the higher prosodic boundaries for both speakers (Female: $F(4,121)=27.95$, $p < 0.0001$, SYL < FT, ip < U, SYL < IP; Male: $F(4,276)=30.24$, $p < 0.0001$, SYL < FT, ip < IP, U). Different patterns were observed for the OQ of the two speakers. For the female speaker the OQ progressively increased in higher prosodic domains ($F(4,121)=13.72$, $p < 0.0001$, SYL, FT < IP, U, SYL < ip), which supports the finding by Garellek (2014). But for the male speaker, the OQ tends to be consistent at different positions, except that significantly lower OQ was found at the U-initial position ($F(4,276)=12.48$, $p < 0.0001$, U < IP, ip, FT, SYL). Decreased SQ in stronger prosodic domains was found for both speakers (Female: $F(4,121)=11.87$, $p < 0.0001$, SYL, FT > IP, U, SYL > ip; Male: $F(4,276)=2.93$, $p < 0.05$, SYL > U).

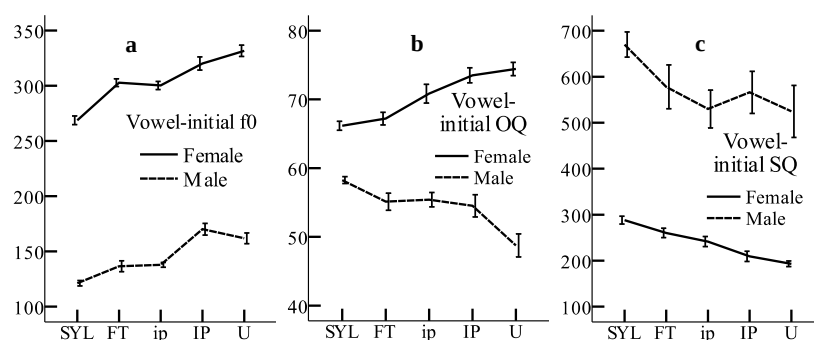


Figure 13 The vowel-initial f0 (a), OQ (b) and SQ (c) for the two speakers.

In summary, the linguopalatal contact and seal duration measures show a general trend in that the segment is strengthened in higher prosodic constituents. However, inter-personal and segmental positional factors are not negligible. As a matter of fact, inter-personal variability has been widely found in previous studies, which show different articulatory strategies adopted by speakers. The slightly different linguopalatal contact at the PMC and the release frames tend to indicate that the strengthening effect might gradually fade away when it is far away from the boundary. The domain-initial strengthening tends not to extend to the vocalic interval that immediately follows the domain-initial stop. The hierarchically structured vowel-initial f0 indicates that the f0 rest is functioning in marking the prosodic structure. The voice quality measures indicate that the vocal folds tend to abduct for a progressively longer portion of time in each abduction-adduction cycle at the edge of higher prosodic constituents. But this gradient variation as a function of boundary strength might be speaker-dependent. In addition, the adduction gesture becomes relatively slower in higher prosodic constituents.

3.3 Acoustic Measures

Table 3 shows the results of the ANOVA and the multiple comparisons for the acoustic measures of the alveolar stop. The acoustic measures at the SYL boundary in the /i/ context produced by the male speaker were excluded because the majority of tokens was either realized as voiced stop or approximant.

Table 3 ANOVA and multiple comparison (Bonferroni method, $p < 0.05$) results for acoustic measures. Direction of the difference in the measures is indicated by “<” for less percentage or length.

Measure	Female speaker	Male speaker
Context a		
VOT	n.s.	n.s.
Voicing during stop closure	$F(4,67)=30.59; p<0.0001$ SYL, FT>ip, IP, U	$F(4,118)=42.25, p<0.0001$ SYL, FT>ip, IP, U
Total voiceless interval	$F(4,67)=156.42; p<0.0001$ SYL, FT<ip<IP<U	$F(4,118)=259.10, p<0.0001$ SYL, FT<ip<IP<U
RMS burst energy	n.s.	n.s.
Context i		
VOT	$F(4,19)=2.35, p=0.09$	$F(3,16)=4.28, p<0.05$ ip>U # ($p=0.07$)
Voicing during stop closure	$F(4,19)=21.85, p<0.0001$ SYL, FT>ip, IP, U	$F(3,16)=65.39, p<0.0001$ FT>ip>IP, U
Total voiceless interval	$F(4,19)=108.96, p<0.0001$ SYL<ip<IP<U FT<IP	$F(3,16)=177.47, p<0.0001$ FT<ip<IP, U
RMS burst energy	$F(4,19)=4.19, p=0.01$ SYL, U<IP	$F(3,16)=10.63, p<0.001$ FT>U

Figure 14 shows the mean values of the VOT at post-boundary position in the two vocalic contexts. In the /a/ context no boundary effect was found, and the VOT stood around 0.015 second, regardless of the boundary types. The boundary types tended to influence the VOT in the /i/ context. For the female speaker the VOT appeared longer at SYL than other domains, but no significant difference was found across boundary types. For the male speaker the boundary effect was significant, and the VOT was marginally longer at the ip boundary than at the U boundary.

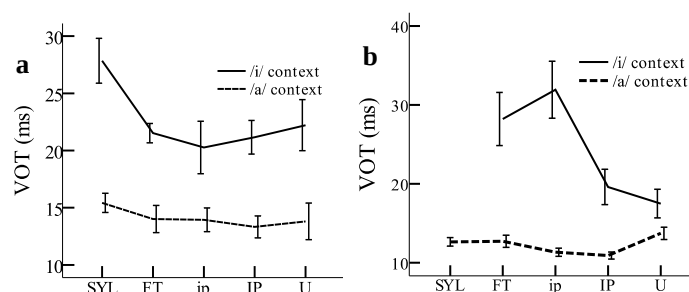


Figure 14 The VOT at post-boundary position for (a) female and (b) male speaker.

The other two acoustic measures that reflect the state of the vocal folds show a clearer cumulative effect of boundary strength across speakers and vocalic contexts. Table 3 and Figure 15 indicate that the voicing takes up the larger portion of the acoustic closure duration at the SYL and FT boundaries than at the three higher boundaries. Voicing during the closure interval was zero at the IP and U boundaries across vocalic contexts and speakers. And at the ip boundary in /a/ context the vocal folds did not vibrate in the acoustic closure duration for both speakers. The variation of the total voiceless interval shows a reverse pattern. It increased as the boundary strength became stronger.

The RMS burst energy shows no significant effect of the boundary strength in /a/ context. In /i/ context no systematic pattern emerges though some significant differences were found in the multiple comparisons.

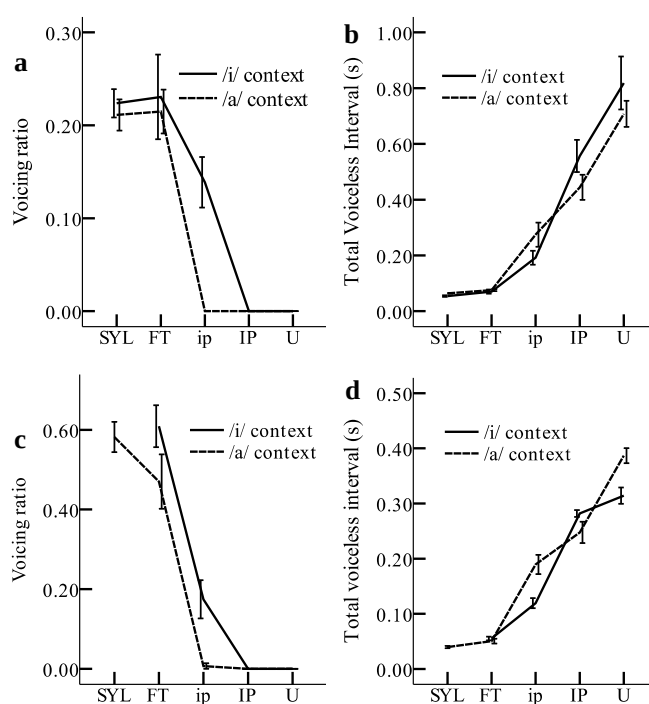


Figure 15 Voicing during the stop closure (a,c) and the total voiceless interval (b,d) produced by the two speakers. (female speaker: a,b; male speaker: c,d)

To summarize, the variation of the four acoustic measures of the alveolar stop shows different pictures under the conditioning of the prosodic structure. The VOT of the unaspirated alveolar stop tends not to be conditioned by the prosodic structure, though the VOT is slightly higher in lower prosodic domains in the /i/ context. The voicing during closure interval and the total voiceless interval are two reliable acoustic measures that show a cumulative strengthening of the post-boundary stop. With the boundary strength becoming stronger, the voicing ratio over the acoustic closure duration decreases, and the total voiceless interval increases. No clear and systematic pattern is found for the RMS burst energy of the stop.

4. DISCUSSION AND CONCLUSION

Two research questions are addressed in this paper. The first concerns whether the domain-initial strengthening is represented as a gradient variation in the articulatory and/or acoustic domains in Standard Chinese? The articulatory evidence shows that articulation for the post-boundary unaspirated alveolar stop varies as a function of boundary strength with increasingly larger linguopalatal contact and longer alveolar seal duration at higher prosodic domains. This cumulative effect on the articulatory magnitude is salient at PMC frame across speakers and vocalic environment. But it fades away toward the release moment, for no robust effect is found at the release frame. The weakening of alveolar closure gesture at the boundary of progressively lower prosodic constituents is accompanied by the reduction of the tongue blade contact on the post-alveolar area. The tongue tip gesture can also be weakened at the initial position of the SYL or FT domains, resulting in the approximant or voiced stop realization.

The vowel-initial f_0 shows a clearly hierarchically-nested pattern, which reflects the global intonational conditioning. The voice quality measures tend to vary with the boundary strength. The vocal folds tend to be abducted in higher prosodic domains for the female speaker, but no clear trend is found for the male speaker. However, both speakers have progressively lower SQ in higher prosodic domains, which indicates that the vocal folds closure gesture becomes slower when the boundary strength is stronger. The results for voice quality supports Kong's finding that the

SQ is negatively correlated with the f_0 , and the result of the OQ produced by the female speaker complies with the recent finding of Garellek (2014).

In terms of acoustic properties of /t/, VOT is not a reliable candidate for marking the boundary strength. But a follow-up acoustic study on stops is worth carrying out because the VOT in the /i/ context shows a tendency to vary as a function of boundary strength. The voicing over the acoustic closure interval and the total voiceless interval are two reliable measures to mark the boundary strength. As the boundary becomes stronger, the former was progressively smaller and the latter progressively longer. These two acoustic measures may be important acoustic correlates for the perception of the boundary strength, which deserves a further perceptual study. The RMS burst energy is expected to show boundary effect with the value becoming smaller at a stronger boundary. However, no clear pattern emerges in both vocalic environments.

The second research question is what is the scope of the domain-initial strengthening effect? In this regard, the high front vowel /i/ in post-boundary syllable /ti/ is used to test the locality hypothesis. No clear trend is found for the linguopalatal contact for /i/. This result shows that the domain-initial strengthening effect is restricted on the temporal domain of the first segment immediately following the boundary, and the strengthening effect quickly fades away toward the release of the first segment. However, the vocalic interval is hierarchically organized by another prosodic device, the intonation. For both speakers, the starting f_0 of the falling tone progressively decreases as the boundary strength becomes weaker. As for the cumulative variation of the two voice quality measures, they might be the by-product of changing f_0 (Kong 2001) or f_0 reset at major boundaries (Garellek 2014).

Though most results in the current paper complied with the previous study, there is one issue worth noting. The articulatory properties of the domain-initial /t/ produced by the two speakers are rather similar in most cases of U and IP. From Figure 11, we can conclude that this might be attributed to the ceiling effect, namely, tongue tip together with the tongue blade gesture are fully realized for the alveolar stop, provided the time threshold is surpassed. At this point, the total voiceless interval could serve as the most important acoustic cue to mark

between U- and IP-boundary. The articulatory properties of /t/ at the ip-initial position tend to pattern with the higher domain in articulatory domains when the time threshold is guaranteed (see Table 2 and Figure 11); however, it is distinguished from the higher domains by the total voiceless interval in both vocalic environments. The articulatory and acoustic properties of /t/ at SYL and FT-initial positions are rather similar; however, a close look at the linguopalatal contact pattern shows a different picture. For the female speaker, the occurrence of incomplete alveolar closure gesture is more frequent in SYL domain compared with the FT domain. This indicates that SYL-initial alveolar gesture is more liable to undergo gestural reduction compared to FT-initial position.

To conclude, the results in the current paper confirm the previous findings in that the domain-initial strengthening is a universally salient effect that cues the prosodic hierarchy in languages. Regarding Standard Chinese, the gradient nature of segment articulation could serve as the important cue that marks the prosodic structure in this language.

NOTES

1. This work was supported in part by the National Natural Science Foundation of China (61073085), Humanity and Social Science Research Fund of the Ministry of Education of China (11YJAZH055) and Yanbian University Social Sciences Research Startup Fund. Professor Kong Jiangping read the draft and offered his insights. Part of the results of this paper was orally presented at the 17th Phonetic Science Congress, 2011, Hong Kong.

REFERENCES

- BOERSMA, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10):341-345.
- BOMBIEN, L., C. Mooshammer, P. Hoole, and B. Kuhnert. 2010. Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics* 38:388-403.
- BYRD, D. 2000. Articulatory vowel lengthening and coordination at

- phrasal juncture. *Phonetica* 57:3-16.
- BYRD, D. and E. Saltzman. 1998. Intra-gestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26:173-199.
- _____. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31:149-180.
- BYRD, D., A. Kaun, S. Narayanan, and E. Saltzman. 2000. Phrasal signatures in articulation. In *Acquisition and the Lexico*, Papers in Laboratory Phonology V, ed. M. B. Broe and J. B. Pierrehumbert, 70-87. Cambridge: Cambridge University Press.
- BYRD, D., J. Krivokapic, and S. Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120:1589-1599.
- CAO, Jianfen 曹劍芬. 2004. Restudy of segmental lengthening in Mandarin Chinese. Paper presented at the Second International Conference on Speech Prosody, Nara, Japan, in its proceedings ed. B. Bel and I. Marlien, 231-234. http://isca-speech.org/archive/sp2004/papers/sp04_231.pdf, (accessed 2/10/2015).
- _____. 2005. Yinduan yanchang de butong leixing jiqi yunlu jiazhi 音段延長的不同類型及其韻律價值 (Different types of segmental lengthening and their prosodic value). *Nanjing Shifan Daxue Wenxueyuan Xuebao* 南京師範大學文學院學報 4:160-167.
- CAO, Jianfen and ZHENG, Yuling. 2006. Articulatory strengthening and prosodic hierarchy. In *Proceedings of the Third International Conference on Speech Prosody*, ed. R. Hoffmann and H. Mixdorff, 289-292. Dresden: Technische Universität Darmstadt Press.
- CHO, T. 2001. Effects of prosody on articulation in English. PhD diss., University of California, Los Angeles.
- _____. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32:141-176.
- _____. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America* 117:3867-3878.
- CHO, T. and P. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29:155-190.

- _____. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37:466-485.
- CHO, T. and J. M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33:121-157.
- FARNETANI, E. and M. Vayra. 1996. The role of prosody in the shaping of articulation in Italian CV syllables. Paper presented at the First ETRW (ESCA Tutorial and Research Workshop) on Speech Production Modelling, Autrans, France, in its proceedings, ed. ESCA, 9-12. http://www.isca-speech.org/archive/spm_96/papers/sps6_009.pdf, (accessed 2/10/2015).
- FOUGERON, C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29:109-135.
- FOUGERON, C. and P. Keating. 1997. Articulatory strengthening at the edge of prosodic domains. *Journal of the Acoustical Society of America* 101:3728-3740.
- GARELLEK, M. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics* 45:106-113.
- HAYASHI, W., C. Hsu, and P. Keating. 1999. Domain-initial strengthening in Taiwanese: A follow-up study. *UCLA Working Paper in Phonetics* 97:152-156.
- HOOLE, P. and F. Hu. 2004. Tone-vowel interaction in Standard Chinese. Paper presented at the International Symposium on Tonal Aspects of Languages, Beijing, China, in its proceedings ed. B. Bel and I. Marlien, 89-92. http://www.isca-speech.org/archive/tal2004/papers/tal4_089.pdf, (accessed 2/10/2015).
- HOWARD, D. M. 1995. Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice* 9:163-172.
- HSU, C. and S. Jun. 1998. Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics* 96:69-89.
- HU, W. X., B. Xu, and T.Y. Huang. 2002. An experimental study on prosodic boundary in Chinese Mandarin. *Journal of Chinese Information Processing* 16(1):43-48.

- JUN, S. 1993. The phonetics and phonology of Korean prosody. PhD diss., Ohio State University.
- KEATING, P., T. Cho, C. Fougeron, and C. Hsu. 2003. Domain-initial articulatory strengthening in four languages. In *Phonetic Interpretation. Papers in Laboratory Phonology VI*, ed. J. Local, R. Ogden, R. Temple, 145-163. Cambridge: Cambridge University Press.
- KIM, S. 2001. The interaction between prosodic domain and segmental properties: Domain initial strengthening of fricatives and post obstruent tensing rule in Korean. M.A. thesis, University of California, Los Angeles.
- KIM, S. and T. Cho. 2011. Effects of prosodic boundary versus accent in the articulation of English /æ/ in #VC and #CVC. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China, in its proceedings ed. W. S. Lee and E. Zee, 1082-1085.
- KONG, Jiangping 孔江平. 2001. Lun yuyan fasheng 論語言發聲 (On language phonation). Beijing: Press of Minzu University of China.
- KUZLA, C. and M. Ernestus. 2011. Prosodic conditioning of phonetic details in German plosives. *Journal of Phonetics* 39:143-155.
- LI, Aijun 李愛軍. 2002. Putonghua duihua zhong yunlu tezheng de shengxue biao xian 普通話對話中韻律特徵的聲學表現 (Prosodic analysis on conversations in Standard Chinese). *Zhongguo Yuwen* 中國語文 6:525-535.
- . 2002. Chinese prosody and prosodic labelling of spontaneous speech. Paper presented at the First International Conference on Speech Prosody, Aix-en-Provence, France, in its proceedings ed. B. Bel and I. Marlien, 39-46. http://www.isca-speech.org/archive/sp2002/papers/sp02_039.pdf, (accessed 2/10/2015).
- LI, Yinghao and Jiangping Kong. 2011. Prosodic boundary effects on segment articulation and V-to-V coarticulation in Standard Chinese. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China, in its proceedings ed. W. S. Lee and E. Zee, 1218-1221.
- LI, Yinghao and Xiaosheng Pan. 2012. The temporal alignment algorithm for electropalatographic and speech signals in long utterance. In *Proceedings of International Conference on Audio, Language and*

- Image Processing*, Volume 2, ed. W. G. Wan, F. L. Luo, and X. Q. Yu, 957-960. Piscataway: IEEE.
- LIN, Maocan 林茂灿. 2002. Putonghua yuju de yunlu jieyou he jipin(F0) gaodixian goujian 普通話語句的韻律結構和基頻(F0)高低線構建 (The prosodic structure in Putonghua and the construction of high- and low-end F0 contour). *Dangdai Yuyanxue* 當代語言學 4:254-265.
- LINDBLOM, B. 1990. Explaining phonetic variation: A sketch of the H & H theory. In *Speech Production and Speech Modeling*, ed. W. J. Hardcastle and A. Marchal, 403-440. Dordrecht: Kluwer Academic Publishers.
- QIAN, Yao 錢瑤, Min Chu 初敏, and Wuyun Pan 潘悟雲. 2001. Putonghua yunlv danyuan bianjie de shengxue fenxi 普通話韻律單元邊界的聲學分析(The acoustic correlates of the prosodic boundaries in Standard Chinese). In *Xinshiji de xiandai yuyinxue: diwujie quanguo xiandai yuyinxue xueshu huiyi* 新世紀的現代語音學: 第五屆全國現代語音學學術會議 (The Phonetic Research in the New Century: Proceedings of the 5th National Conference on Modern Phonetics), ed. Lianhong Cai 蔡蓮紅, Tongchun Zhou 周同春, and Jianhua Tao 陶建華, 70-75. Beijing: Qinghua daxue chubanshe.
- SELKIRK, E.O. 1984. *Phonology and Syntax: the Relationship between Sound and Structure*. Cambridge: MIT Press.
- TSENG, C., S. Pin, Y. Lee, H. Wang, and Y. Chen. 2005. Fluent speech prosody: Framework and modeling. *Speech Communication* 46(3-4):284-309.
- WANG, B., Y. F. Yang, and S. N. Lu. 2004. Acoustic analysis on prosodic hierarchical boundaries of Chinese. *Acta Acoustica* 29(1):29-36.
- WANG, Hongjun 王洪君. 2008. *Hanyu feixianxing yinxixue* 漢語非線性音系學 (*Non-linear Phonology of Mandarin Chinese*). Beijing: Beijing Daxue Chubanshe.
- WIGHTMAN, C. W., S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3):1707-1717.

韻律邊界對漢語普通話音段發音的影響—
基於發音生理和聲學的研究

李英浩*

延邊大學, 延吉 (中國)

提要

本文使用動態電子齶位元(EPG)和聲學分析的方法,考察漢語普通話韻律邊界對韻律單元域首音段的發音生理和聲學特徵的影響。我們選取普通話的清不送氣齒齶塞音/t/和前高母音/i/,從音段產生的空間域和時間域分析域首發音增強現象。普通話的韻律層級包括音節、音步、小韻律短語、大韻律短語和話語。通過比較不同韻律層次域首音段的發音生理和聲學參數,我們發現:(1)單元域首輔音的發音動作受到普通話韻律結構的制約。輔音的舌齶接觸和生理持阻時長與韻律邊界的強度密切相關;輔音的舌齶接觸與生理持阻時長之間呈現出非線性關係;輔音聲學時段的清聲段時長和濁聲時長比能夠有效地標記韻律邊界的強度。(2)輔音除阻時刻的舌齶接觸以及輔音後接母音的最大舌齶接觸受邊界強度的影響較小,且後接母音的嗓音特徵與較大韻律邊界的基頻重設有關。這說明域首發音增強的作用域限於韻律邊界後面的音段。研究結果表明,普通話韻律單元域首音段的發音特徵得到增強,且增強程度與邊界強度密切相關,這是標示普通話韻律結構的一個重要方式。

主題詞

韻律邊界 音段發音 動態電子齶位元 漢語普通話

* 李英浩為延邊大學副教授,執教于延邊大學外國語學院英語系;通訊:中國吉林省延吉市,延邊大學外國語學院英語系 郵編 133002; [youngho_lee@aliyun.com]

VARIATIONS OF LARYNGEAL FEATURES IN JIANCHUAN BAI¹

Feng Wang*

Peking University

ABSTRACT

The tonal system in Jianchuan Bai has attracted much attention for its complex combinations of pitch and phonation type. In this paper, based on EGG signals, three parameters, namely F0, Open Quotient (OQ) and Speed Quotient (SQ), are extracted to examine the tonal quality. It is found that there are two non-modal phonation types, Harsh and Pressed, and roughly four groups of pitch pattern (31/31/41; 33/433; 55/54;35) in the eight tonal categories. One pair of tones can only be distinguished from each other by phonation type since their pitches are the same. As for other pairs, both pitch and phonation type may contribute to the distinction between them. Notably, non-modal phonation types vary across different Bai speakers. For a particular non-modal tone, one speaker may employ harsh voice, while another may use pressed voice. Sometimes, the non-modal phonation type even changed within a syllable. It is then suggested that different strategies may be used to produce non-modal tones in contrast with their modal counterpart. Moreover, based on the Bai data, how to define different phonation types based on the three basic parameters, F0, OQ and SQ, is discussed. Harsh voice is a better term for the type with the laryngeal features [Middle falling F0, -OQ, -SQ] rather than high-pitched voice.

SUBJECT KEYWORDS

Jianchuan Bai EGG Phonation type Tone

1. INTRODUCTION

Half a century ago, non-modal voices in minority languages in China were

* Feng Wang 汪鋒 is an associate professor of Peking University at Center for Chinese Linguistics, Department of Chinese Language and Literature, and Joint Research Center for Language and Human Complexity. The corresponding address: Department of Chinese Language and Literature, Peking University, Beijing 100871, China: [wfwf@pku.edu.cn]

first discovered (e.g. Ma 1948), and then categorized as ‘Tense vowels’ contrasting with ‘Lax vowels’ (of modal voice) in the phonology (See Dai 1958; Hu and Dai 1964; Dai 1979). In the 1980’s, experimental studies on acoustic signals revealed that the ‘Tense/Lax’ contrast is a distinction in phonation types rather than in articulation (See Maddieson and Ladefoged 1985). Later on, experimental studies on different phonation types have been relying on Electroglottography (EGG) signals recently since the amount of reliable parameters to describe the movements of vocal folds can be obtained without any intrusion (See Kong 2001).

Phonation types in the Bai language are rich. Xu and Zhao (1984:12) stated that ‘syllables in Bai bear the Lax/Tense distinction, which is associated with not only vowels but also tones’. Here is their description of Jianchuan Bai (English translation is added).

Table 1. The description of tonal system of Jianchuan Bai

调名 Tone name	调值 Tone value	调号 Tone mark	元音松紧 Vowel quality	例字 Example	
1	33	𐌰	松 Lax	pa1 泡沫 ‘foam’	tei1 拉 ‘pull’
2	42	𐌱	紧 Tense	pa2 奶 ‘milk’	tei2 追 ‘chase’
3	31	𐌲	松 Lax	pa3 闹 ‘noisy’	tei3 田 ‘farmland’
4	55	𐌳	松 Lax	pha4 扒 ‘dig up’	tei4 多 ‘many’
5	35	𐌴	松 Lax	pa5 八(哥鸟) ‘myna’	tei5 急 ‘hurry’
6	44	𐌵	紧 Tense	pa6 倒 ‘collapse’	tei6 蚂 蟥 ‘leech’
7	21	𐌶	紧 Tense	pā7 蹄 ‘hoof’	tei7 手 镯 ‘bracelet’
8	55	𐌷	紧 Tense	pa8(水) 坝 ‘dam’	tei8 寄 (宿) ‘lodge’

Regarding the combination of tone and lax/tense vowels, the table suggests that there are eight phonological categories which can be marked by the eight single numerals. Each category is corresponding to a unique tone value. For example, the tone value of Tone 1 is 33 following the five-tone scale created by Chao (1930/1980). Notably, the above tone letters are creative. The lax ones are marked on the left as usual, while the tense ones on the right. In fact, such letters suggest that two domains of Lax/Tense should be differentiated at first, and under each domain pitches play distinctive roles. Now it is generally known that the lax/tense distinction is not due to

articulation of vowels, but phonation of vocal folds. From the perspective of phonation Kong (2001) proposed the theory of tone quality to incorporate Tiaoshi (调时) ‘temporal tone’ and Tiaosheng (调声) ‘phonation tone’. Temporal tone indicates how fast the vocal folds vibrate in the temporal domain, and its acoustic parameter is F0. Phonation tone indicates how vocal folds vibrate, and open quotient and speed quotient are parameters that are often used (details later). To apply this theory to the Bai data, the question is how these parameters in temporal tone and phonation tone interact internally and externally in an integrated tone system.

2. DATA OF JIANCHUAN BAI

Four native Bai speakers are from the Jinxing village of the Jinhua town, Jianchuan county, Yunnan, two males (M) and two females (F). M1, 55-year old; M2, 37-year old; F1, 65-year old; F2, 78-year old.

Cooledit 2.0 is used to record sound on the left channel and EGG signals on the right channel. The sample rate is 22050Hz.

Regarding tones in Jianchuan Bai, though interpretations of the phonetic realities are still controversial, the eight phonological categories are generally agreed upon. For convenience, this paper will use tone names. To highlight the lax/tense distinction, the two cover symbols, T for Tense and L for Lax, precede the tone names, e.g. L1, T2, L3, L4, L5, T6, T7, and T8.

3. TEMPORAL TONE

For each tone, four examples are given as follow.

Table 2. Recording samples

	Ex.1	Ex.2	Ex.3	Ex.4
L1	拉 tɕi1 ‘pull’	泡沫 pa1 ‘foam’	斧头 pu1 ‘axe’	贩卖 pe1 ‘sell’
T2	追 tɕi2 ‘chase’	奶 pa2 ‘milk’	蹂躏 ta2 ‘tread’	插 pe2 ‘insert’
L3	地 tɕi3 ‘farmland’	闹 pa3 ‘noisy’	偷 ta3 ‘steal’	倍 pe3 ‘double’
L4	多 tɕi4 ‘many’	伯伯 ta4(ta4) ‘elder uncle’	星星 cian4 ‘star’	糠 tshō4 ‘chaff’
L5	急 tɕi5 ‘hurry’	八(哥)pa5 ‘myna’	那 ta5 ‘that’	来 ya5 ‘come’
T6	蚂蟥 tɕi6 ‘leech’	倒 pa6 ‘collapse’	北 pu6 ‘north’	走 pe6 ‘walk’
T7	手 镯 tɕi7 ‘bracelet’	一只 ta7 ‘Classifier for bird’	桃子 ta7 ‘peach’	皮 pe7 ‘skin’
T8	寄 tɕi8 ‘post’	大(哥) ta8(ko33) ‘elder brother’	县 cian8 ‘county’	冲(菜)tshō8 ‘Chongcai leaf mustard’

Each word was pronounced twice. F0 is extracted by the method of auto-correlation².

Thirty points are extracted for each sample, but the three at the beginning and the three at the end will be excluded to ensure stable representation. At first, we deal with F0 data person by person. For the two males, the F0 patterns are almost the same. Taking M2 and F2 as the example, their F0 contours are distributed in Figure 1 and Figure 2.

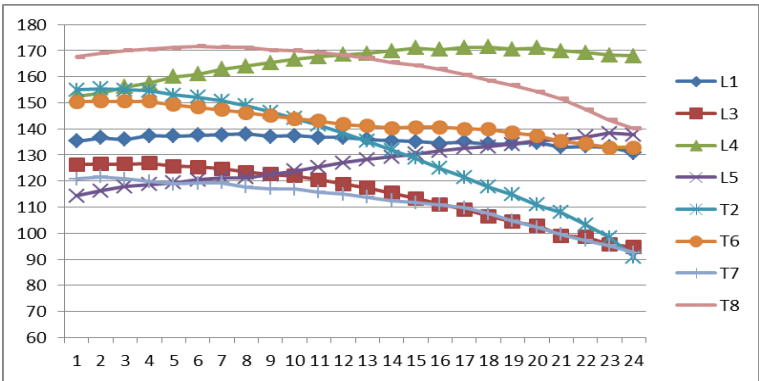


Figure 1. F0 contours of M2 for eight tones

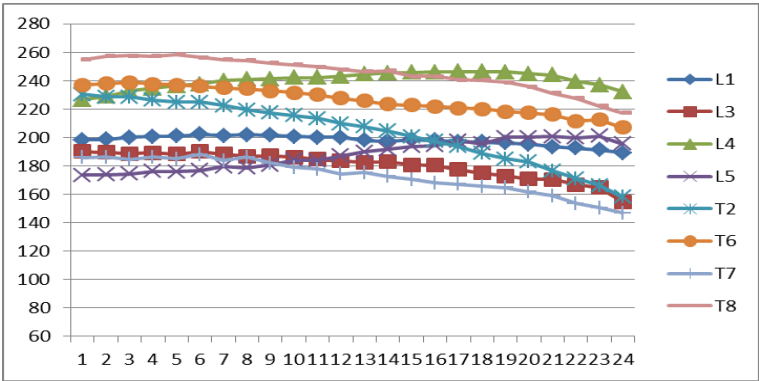


Figure 2. F0 contours of F1 for eight tones

Values of F0 can be converted to relative values (RV) according to Chao (1930) in order to reflect linguistic distinctions. The following fomula can be used.

$$RV(x)=[(lgx-lgb)/(lga-lgb)*4]+1$$

(Note: ‘a’ is the largest F0 value; ‘b’ is the smallest F0 value.)

Table 3. The eight tones of Jianchuan Bai

	RV(M2)	RV(F1)	Xu & Zhao 1984
L1	33	33	33
T2	41	41	42

L3	31	31	31
L4	55	45	55
L5	24	23	35
T6	433	43	44
T7	31	31	21
T8	54	54	55

For the above table, it is impossible to distinguish L3 from T7 based on the F0 parameters. However, F0 does not equal to pitch. It is intriguing to investigate whether and how different phonation types contribute to the pitches in future studies. Anyway, it is expected that the two pairs, L3/T7 and L1/T6, may be differentiated from their phonation types.

Interestingly, if the tone domain splits into two, Lax and Tense, the contrasts in each domain are quite clear taking M2 as the example shown as below.

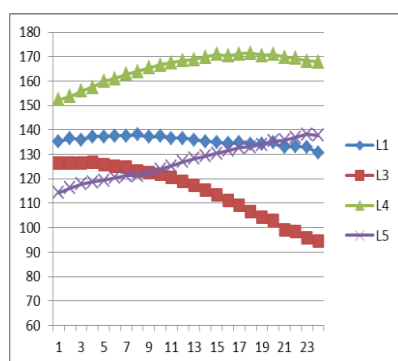


Figure 3. Lax tones of M2

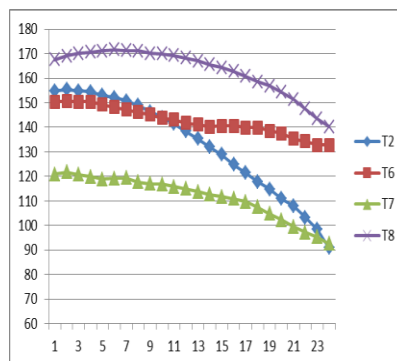


Figure 4. Tense tones of M2

4. PHONATION TONE

4.1 EGG Signals and the Bai Data

A typical EGG signal can be shown as below.

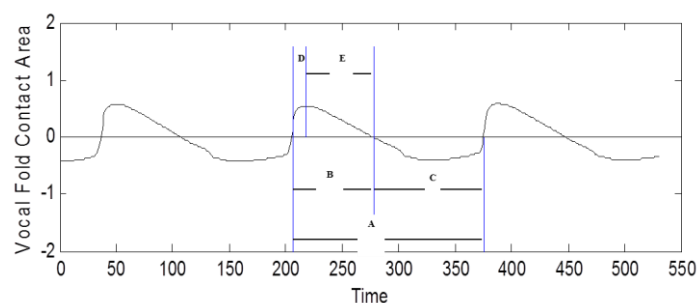


Figure 5. EGG signals

The phase A is a period of the vibrations of vocal folds. B= the close phase. C= the open phase. D= the closing phase. E=the opening phase. Two parameters, Open quotient (OQ) and Speed Quotient (SQ), are frequently used to describe phonation types.

Open Quotient= Open phase (C) / Period (A)×100%

Speed Quotient= Opening phase (E) / Closing phase (D)×100%

By the parameters of temporal tone (Pitch) and phonation tone (OQ, SQ), several typical phonation types can be defined as the following table 4 (see Kong 2001:188). In the following analysis, this table will be the major reference to define the phonation types in Bai.

Table 4. Phonation types and their parameters

	Vocal fry	Breathy	Pressed	Modal	High-pitch
OQ	+	+	-	+ -	-
SQ	+	-	+	+ -	-
Pitch	-	-	-	+ -	+

According to the F0 similarities of the eight tones, three groups can be recognized to investigate their contrasts in phonation types as follows.

- (1) T7 (31) / L3 (31) / T2 (41);
- (2) L1 (33) / T6 (433);
- (3) L4 (55) / T8 (54).

The table below presents minimal pairs for the three groups.

Table 5. Three groups of minimal pairs of phonation types

Group (1)

T7 (31)	L3 (31)	T2 (41)
手镯 tɕi7 ‘bracelet’	地 tɕi3 ‘farmland’	追 tɕi2 ‘chase’
皮 pe7 ‘skin’	倍 pe3 ‘double’	插 pe2 ‘insert’
一只 tɔ7 ‘Classifier for bird’	豆 tɔ3 ‘bean’	草乌 tɔ2 ‘Radix Aconiti Kusnezoffii’
漂浮 pu7 ‘float’	托带 pu3 ‘help to bring’	敷药 pu2 ‘apply medicine’
荞麦 kv7	柜子 kv3 ‘cabinet’	坐 kv2 ‘sit’
桃子 ta7 ‘peach’	偷 ta3 ‘steal’	踩踏 ta2 ‘tread’

Group (2)

L1(33)	T6 (433)
拉 tɕi1 'pull'	蚂蟥 tɕi6 'leech'
贩卖 pe1 'sell'	走 pe6 'walk'
等待 tɕ1 'wait'	得到 tɕ6 'obtain'
斧头 pu1 'axe'	北 pu6 'north'
鬼 kv1 'ghost'	犄角 kv6 'horn'
泡沫 pa1 'foam'	倒 pa6 'collapse'

Group (3)

L4 (55)	T8 (54)
多 tɕi4 'many'	寄 tɕi8 'post'
伯伯 ta4 (ta4) 'elder uncle'	大(哥) ta8(ko1) 'elder brother'
星星 ɕian4 'star'	县 ɕian8 'county'
糠 tshō4 'chaff'	冲(菜) tshō8(tse4) 'Chongcai leaf mustard'

4.2 OQ/SQ and Phonation Types of the Four Tense Tones

Each sample is pronounced twice by each of the four informants. Thirty points of each sample were extracted to calculate their OQ and SQ values. The first three and the last three points were not counted to keep the stable data. At first, the average OQ and SQ values are expected to display a general picture. The average OQ and SQ of samples of each tone at the stable 24 points will be counted, and then the 24 values will be averaged again. Taking M1 as the example, the final average OQ and SQ of the three groups of tense/lax tones can be shown as follows.

M1

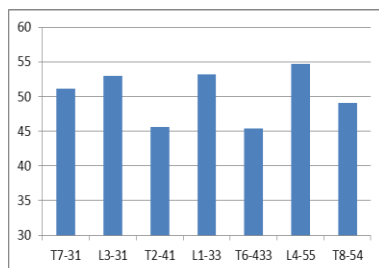


Figure 6. OQ of M1

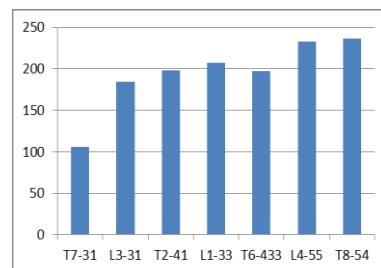


Figure 7. SQ of M1

In each of the three groups, the modal sound will be taken as the baseline to measure the relative value of the non-modal sound. For M1, OQ of L3 is higher than both of the non-modal sounds, T7 and T2. Therefore, the relative values of the two non-modal sounds are T7(-OQ) and T2(-OQ) compared to L3(+OQ). Similarly, the relative values of group 2 can be marked as T6(-OQ) compared to L1(+OQ).

The relative values of group 3 can be marked as T8(-OQ) compared to L4(+OQ). The same procedure could be applied to SQ of M1. For group 1, T7 (-SQ) and T2(+OQ) compared to L3(+SQ) can be obtained. For group 2, T6(-SQ) compared to L1(+SQ) is marked. For group 3, T8(+SQ) compared to L4(+SQ) is marked.

The OQ/SQ contrasts of the four speakers can be summarized as follows.

Table 6. Summary of OQ/SQ contrasts of the four speakers

		M1		M2		F1		F2	
		OQ	SQ	OQ	SQ	OQ	SQ	OQ	SQ
Group 1	T7-31	-	-	-	+	-	-	-	+
	L3-31	+-	+-	+-	+-	+-	+-	+-	+-
	T2-41	-	+	-	+	-	+	-	+
Group 2	L1-33	+-	+-	+-	+-	+-	+-	+-	+-
	T6-433	-	-	-	+	-	-	-	+
Group 3	L4-55	+-	+-	+-	+-	+-	+-	+-	+-
	T8-54	-	+	-	+	-	+	-	-

According to Kong (2001:188, shown in Table 4), the phonation types of the seven tones can be defined as below.

Table 7. Classification of phonation types of Jianchuan Bai

		M1	M2	F1	F2
Group 1	T7-31	High-pitch	Pressed	High-pitch	Pressed
	L3-31	<i>Modal</i>	<i>Modal</i>	<i>Modal</i>	<i>Modal</i>
	T2-41	Pressed	Pressed	Pressed	Pressed
Group 2	L1-33	<i>Modal</i>	<i>Modal</i>	<i>Modal</i>	<i>Modal</i>
	T6-433	High-pitch	Pressed	High-pitch	Pressed
Group 3	L4-55	<i>Modal</i>	<i>Modal</i>	<i>Modal</i>	<i>Modal</i>
	T8-54	Pressed	Pressed	Pressed	High-pitched

However, these definitions are based on OQ/SQ. If the parameter of pitch is considered as in Table 4, some problems may be raised. For instance, the phonation type of T7-31 in M1 and F1 should not be called ‘High-Pitch’ since its pitch, 31, is rather middle-falling, and is almost the same as its lax counterpart, L3-31. Based on some observations, it is realized as Harsh (Li 1992; Edmondson and Li 1994; Edmondson et al 2000, 2001) or ‘Glottal squeezing and friction’ (Li 1992). Based on our hearing, our data would support the term of Harsh. Therefore, we revised the definitions as follows.

Table 8. Revised Classification of phonation types of Jianchuan Bai

			M1	M2	F1	F2
Group 1		T7-31	Harsh	Pressed	Harsh	Pressed
		L3-31	Modal	Modal	Modal	Modal
		T2-41	Pressed	Pressed	Pressed	Pressed
Group 2		L1-33	Modal	Modal	Modal	Modal
		T6-433	Harsh	Pressed	Harsh	Pressed
Group 3		L4-55	Modal	Modal	Modal	Modal
		T8-54	Pressed	Pressed	Pressed	Harsh

From the above table, two points can be made. Firstly, the same tone category does not imply the same phonation type. For instance, T3-31 is associated with Harsh voice in M2 and F1, but with Pressed voice in M2 and F2. Secondly, the tonal distinction may be afforded by F0 or phonation type or both. For instance, T7-31 and T2-41 in M1 can be distinguished by F0 (31 vs 41) or phonation type (Harsh vs Pressed). T7-31 and L3-31 in M1 can be distinguished only by phonation type (Harsh vs Modal). T7-31 and T2-41 in M2 can be distinguished only by F0 (31 vs 41).

4.3 Variations of OQ/SQ within Syllables and Across Speakers

In fact, more variations of OQ and SQ across speakers can be obtained if the whole 24 points of each tone will be shown based on the average value of OQ contours and that of SQ contours. The details are listed speaker by speaker according to the three groups.

Group 1: M1

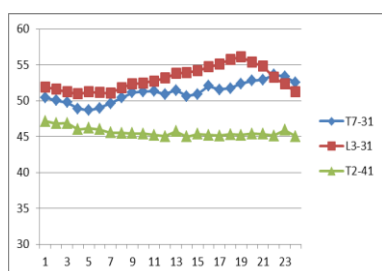


Figure 8. OQ of M1

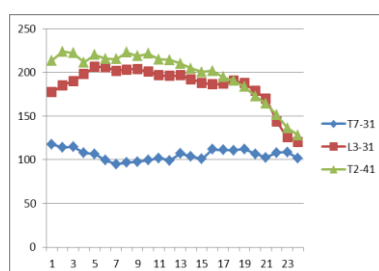


Figure 9. SQ of M1

Regarding OQ, T7-31 is not higher than L3-31 any more at the final three points. Regarding SQ, T2-41 becomes the same with L3-31. Such variations cause the distinctive function of the lower OQ of T2-41 and the lower SQ of T7-31 in contrast with their lax counterpart, L3-31, to become salient. In other words, SQ of T2-41 and OQ of T7-31 in contrast with this group may not be functional at the end

of the syllables.

M2

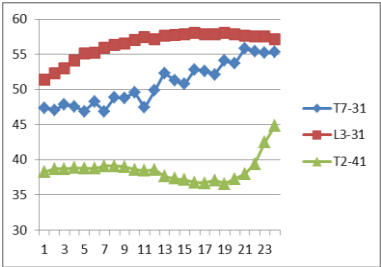


Figure 10. OQ of M2

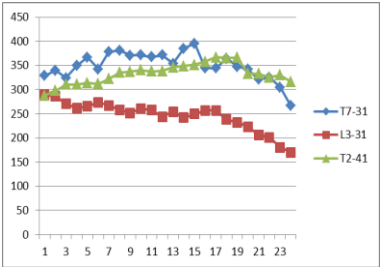


Figure 11. SQ of M2

In contrast with L3-31, the relative value of T7-31 and T2-41 keep stable, except that SQ of T2-41 is not higher than L3-31 at the first two points. Though it does not have any impact on the contrast of this group, the relative positions of T7-31 and T2-41 vary a bit.

F1⁴

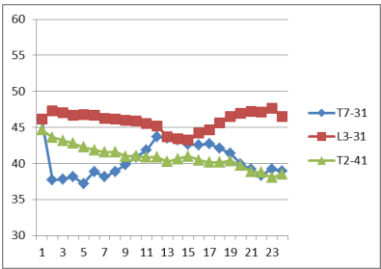


Figure 12. OQ of F1

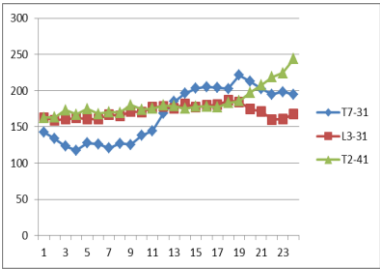


Figure 13. SQ of F1

Variations of F1 are quite salient. In the middle three points, OQ of T7-31 becomes the same with L3-31. Only in the last five points, SQ of T2-41 begins to be larger than L3-31. More interestingly, SQ of T7-31 is smaller than L3-31 in the first half of the syllable, while it becomes larger in the last half. That means the phonation type of T7-31 changes from Hash to Pressed within a syllable.

F2

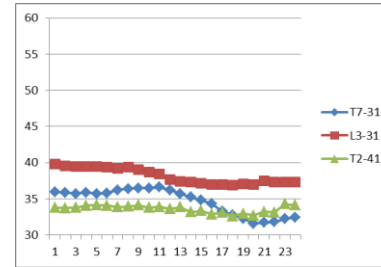


Figure 14. OQ of F2

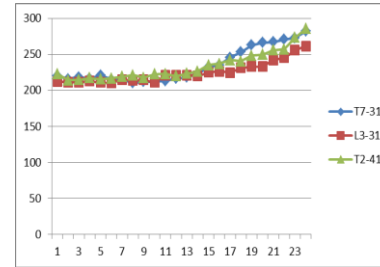


Figure 15. SQ of F2

SQ of the three tones becomes distinguishable a little bit only at the last one-third of the syllable. This would result in the salience of OQ distinction in this group.

Group 2: M1

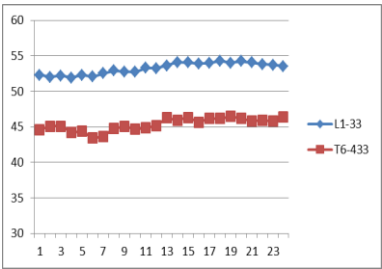


Figure 16 OQ of M1

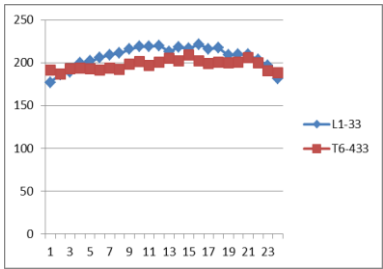


Figure 17 SQ of M1

T6-433 cannot be distinguished from its lax counterpart L1-33 by SQ at the very beginning and the very end, while its OQ is obviously lower.

M2

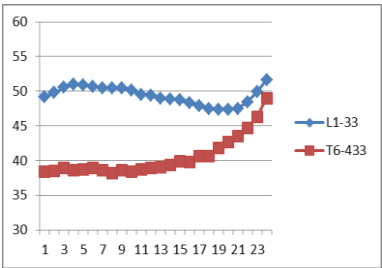


Figure 18 OQ of M2

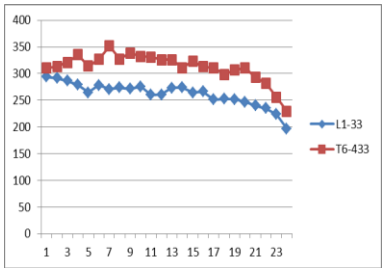


Figure 19 SQ of M2

F1

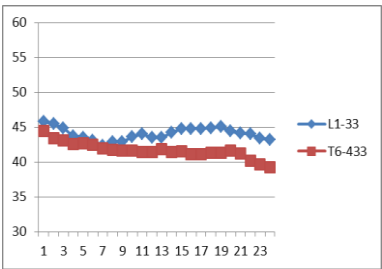


Figure 20 OQ of F1

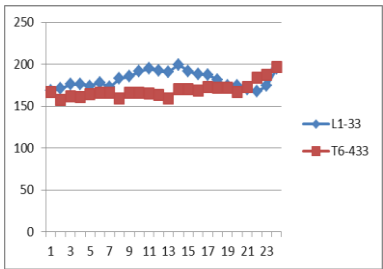


Figure 21 SQ of F1

At the beginning one-third, OQ of T6-433 is almost the same with L1-33. Meanwhile, T6-433 can be hardly distinguished from its lax counterpart L1-33 by SQ at the beginning and the end.

F2

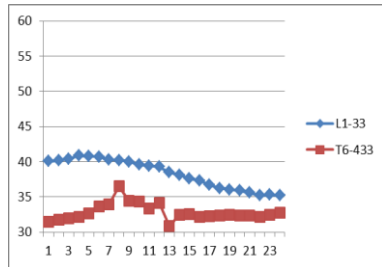


Figure 22 OQ of F2

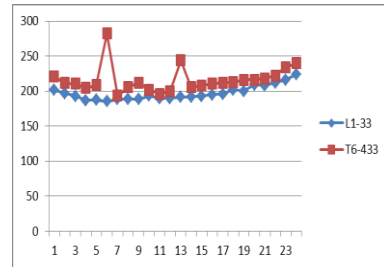


Figure 23 SQ of F2

Group 3:M1

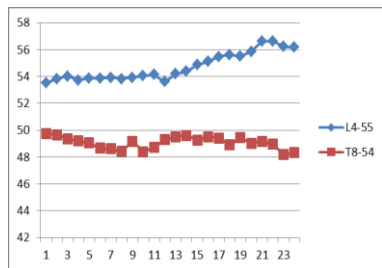


Figure 24. OQ of M1

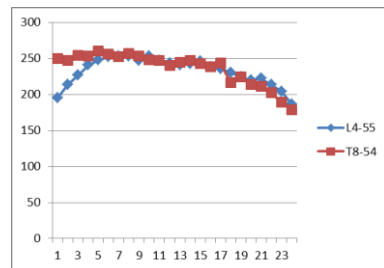


Figure 25. SQ of M1

T8-54 and L4-55 are almost the same in SQ, while T8 is much lower than L4 in OQ.

M2

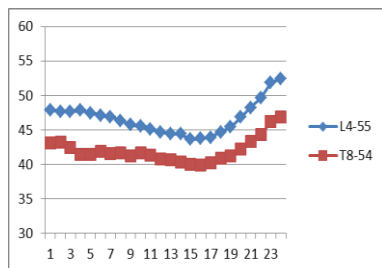


Figure 26 OQ of M2

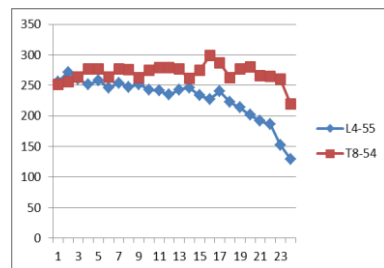


Figure 27 SQ of M2

T8-54 and L4-55 are almost the same in SQ at the first three points.

F1

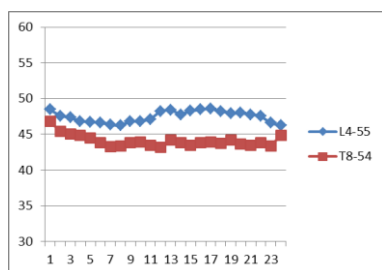


Figure 28. SQ of F1

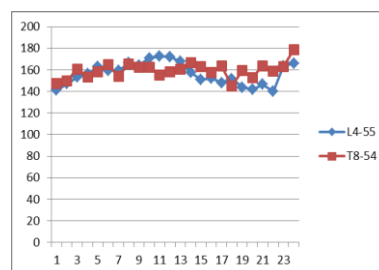


Figure 29. OQ of F1

The contours of T8-54 and L4-55 are almost mixing together.

F2

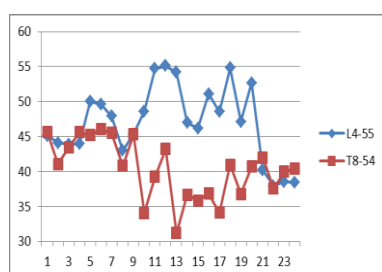


Figure 30. OQ of F2

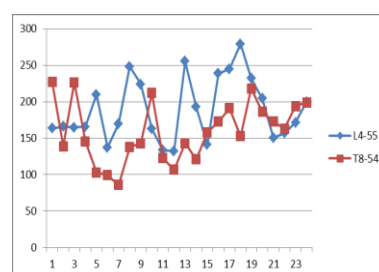


Figure 31. SQ of F2

The fluctuation of OQ and SQ in F2 is notably different from other speakers. However, the major trend is still clear. The values of the lax tone are higher than those of the tense tone.

5. CONCLUSION AND REMARKS

The eight tone categories in Jianchuan Bai have attracted lots of attention. The most intriguing feature of this tonal system is the complex interaction of the two dimensions of tone, namely, temporal tone (pitch) and phonation tone (phonation type). In this study, based on the calculations of F0, OQ and SQ, it is found that there are two non-modal phonation types, Harsh and Pressed, roughly four groups of pitch pattern (31/31/41; 33/433;55/54;35) to make the eight tonal categories shown as below.

Table 9. The eight tone categories in Jianchuan Bai

Tone name	Pitch	Phonation type
T7	31	Harsh/Pressed
L3	31	Modal
T2	41	Pressed
L1	33	Modal
T6	433	Harsh/Pressed

L4	55	<i>Modal</i>
T8	54	Pressed
L5	35	<i>Modal</i>

T7 and L3 must be distinguished by phonation type, since their pitches are the same. As for other pairs, both pitch and phonation type may contribute to the distinction between them. Notably, non-modal phonation types vary across individuals (see 4.3). Sometimes, phonation types even change within a syllable, e.g. T7-31 of informant F1 from Harsh to Pressed. Thus, it is suggested that different persons in a society may use different strategies to produce non-modal phonation types in contrast with their counterparts. As for whether and how different strategies are associated with different social factors, more data are needed. For now it is also difficult to detect the distinctive features of phonation types. In future studies, perception experiments may be helpful to figure out the distinctive features.

The Bai language may be the closest sister language of Chinese (Wang 2006b; 2012; 2013). The nature of the Bai tonal system will shed light on understanding the origin and development of the Chinese tonal system. The complex variation of laryngeal features in Jianchuan Bai hints that phonation types may be one source of the Chinese tonal system, in addition to initials or endings (Wang 2006a).

NOTES

1. This work was supported by the Fok Ying Tung Education Foundation [No. 131102], the Major Project of National Social Science Fund [10&ZD125], the fund from Ministry of Education of China [No.11JJD740004]. During the fieldwork, Prof. Duan Bingchang, Mr. Yang Haichao and Mr. Zhao Guangjian offered their kind help, which is much appreciated.
2. The Matlab programs used in this paper are edited by Li Yonghong and Ye Zehua.
3. Lack of the sample pe42 ‘insert’.
4. The sample tɔ7 ‘classifier (e.g. for rabbit)’ has not been recorded well. It is replaced by ɣɔ7 ‘to mill’.

REFERENCES

CHAO, Y. R.1930/1980. A system of tone letters. Fangyan 2. 1980:81-83.

- DAI, Qingxia 戴慶廈. 1979. Woguo zangmianyu zu songjin yuanyin lai yuan chutan 我國藏緬語族鬆緊元音來源初探 (The origin of tense/lax vowel in Tibeto-Burman languages of China). *Minzu yuwen* 民族語文.1:31-39.
- EDMONDSON, J. A. and Li, Shaoni. 1994. Voice quality and voice quality change in the Bai language of Yunnan Province. *Linguistics of Tibeto-Burman Area* 17.2:49-68.
- HU, Tan 胡坦 & Dai, Qingxia 戴慶廈. 1964 Haniyu yuanyin de songjin 哈尼語元音的松緊(Lax/tense of vowels in Hani). *Zhongguo yuwen* 中國語文. 1: 76-87.
- HUANG, Bufan 黃布凡 et al. 1992. Zang Mian yu zu yu yan ci hui 藏緬語族語言辭彙 (A Tibeto-Burman lexicon). Beijing 北京: Zhongyang minzu xueyuan chubanshe 中央民族學院出版社.
- KONG, Jiangping 孔江平. 2001. Lun yuyan fasheng 論語言發聲(On language phonation). Beijing 北京: Zhongyang minzu xueyuan chubanshe 中央民族大學出版社.
- LI, Shaoni 李紹尼. 1992. Lun baiyu de shengmen hunhe jicayin 論白語的聲門混合擠擦音 (On the combination of glottal squeezing and frication in the Bai language). *Minzu yuwen* 民族語文 4:68-72.
- MA, Xueliang 馬學良. 1948. Luowen zuoji xianyao gongsheng jing yizhu 僦文作祭獻藥供牲經譯注(Notation of Yi classics). *Zhongyang ynjuiyuan lishi yuyansuo jikan* 中央研究院歷史語言所集刊.20:579.
- MADDIESON Ian and Ladefoged Peter. 1985. 'Tense' and 'Lax' in four minority languages of China, *UCLA Working Papers in Phonetics*. 60-99.
- WANG, Feng. 2006a. Rethinking the *-s hypothesis for Chinese qusheng tone. *Journal of Chinese linguistics* 34.1:1-24.
- _____. 2006b. Comparison of languages in contact: the distillation method and the case of Bai. Taipei: Institute of Linguistics, Academia Sinica.
- _____. 2007. Baiyu fangyan zhong teshu fasheng leixing de lai yuan yu yanbian 白語方言中特殊發聲類型的來源與演變 (Origins and development of non-modal phonation types in Bai dialects). *Hanzangyu xuebao* 漢藏語學報.1: 162-170.
- _____. 2012. Yuyan jiechu yu yuyan bijiao- yi baiyu weili 語言接觸與語言比較——以白語為例(Language contact and language comparison- the case of Bai). Beijing 北京: Shangwu yinshuguan 商務印書館.

- . 2013. Hanzang yuyan bijiao de fangfa yu shijian- Han, Bai, Yiyu bijiao yanjiu 漢藏語言比較的方法與實踐—漢、白、彝語比較研究(Methods and implementations of Sino-Tibetan comparison—A comparative study of Chinese, Bai and Yi). Beijing 北京: Beijing daxue chubanshe:北京大學出版社.
- WANG Feng and Kong, Jiangping 孔江平. 2009. Wuding yiyu songjinyin yanjiu 武定彝語鬆緊音研究(A study of Lax/tense voice in the Wuding Yi). Zhongguo yuyanxue 中國語言學 2. Ji'nan 濟南: Shandong jiaoyu chubanshe 山東教育出版社:98-118.
- XU, Lin 徐琳 & Zhao, Yansun 趙衍蓀. 1984. Baiyu jianzhi 白語簡志(A brief description of Bai). Beijing 北京: Minzu chubanshe 民族出版社.

劍川白語的嗓音變異

汪鋒

北京大學*

提要

劍川白語的聲調系統中音高與發聲類型的配合關係複雜，很引人注目。本文基於電聲門信號，提取基頻、開商和速度商三個參數來探討聲調的性質。在八個調類中，有兩種特殊發聲類型-刺耳音和緊喉音。按音高可以大致分為 4 組，即 31/31/41; 33/433; 55/54;35。其中有一對音高完全相同，只能根據發聲類型的不同來區分調類。至於其他對，音高和發聲類型都可能對區分有貢獻。值得注意的是，特殊發聲類型在不同發音人中有變異。某種特殊發聲類型，一個白語發音人可能用刺耳音，而另一個可能發緊喉音。有時，特殊發聲類型可能在一個音節內發生轉換。這說明不同的發音人可能運用不同的發聲策略來與相對的普通發聲調類構成區別。論文根據白語進一步探討了如何根據基頻、開商和速度商這三個基本參數來定義發聲類型的種類。對於白語中具有中降調、低開商和低速度商的特殊發聲類型而言，刺耳音或許是比高音調嗓音更合適的術語。

關鍵詞

劍川白語，電聲門信號，發聲類型，聲調

* 汪鋒為北京大學副教授，執教研究於中國語言學研究中心，中文系，語言與人類複雜系統聯合研究中心。聯繫地址：中國北京，郵編 100871，北京大學中文系；[wfwf@pku.edu.cn]



北京大学

硕士研究生学位论文

题目： 含灯大鼓的声学研究

姓 名： 张春连

学 号： 1201211203

院 系： 中国语言文学系

专 业： 语言学及应用语言学

研究方向： 语音学

导 师： 孔江平教授

二〇一五年五月

摘要

本文利用实验语音学的方法对含灯大鼓这一中国传统曲艺做了分析研究。含灯大鼓是梅花大鼓的一种演唱方式，与传统的梅花大鼓不同，含灯大鼓演唱者演唱时嘴里衔着一根灯架，这就影响了演唱者发音时的整个声道调节：下颌固定，牙齿不能随意开合，唇、舌的动作受限，气息受限。本文对比了同一发音人含灯和正常情况下单元音、语篇和曲目的共振峰、长时共振峰、长时平均谱、基频、开商、速度商等嗓音声学参数，试图找出两者的差异。

研究得出以下结论：

一、含灯后的声学元音图明显比正常时变小，这说明舌位的活动范围受到限制而缩小。/a/、/i/、/u/三个元音的F2都降低了，说明含灯时发音人的舌位发生了整体后移。其中高元音/i/的F2所受影响最大，降低的最为明显；

二、从语篇的长时共振峰来看，含灯后的F1比正常时稍高，F2整体分布范围明显比正常时高，F3没有明显变化。F2的升高可能和含灯后唇的左右开度增加、口腔前腔减小舌头相对位置靠前、圆唇程度降低这几个方面的因素综合相关。

三、/a/、/i/、/u/、/y/四个单元音基频方面，含灯时的均值比正常时更小。开商方面，高元音含灯时的均值要高于正常时的均值，其他元音正常时的均值高于含灯时。速度商方面，前元音正常时的均值大于含灯时的均值，其他元音含灯时的均值大于正常时的均值。通过误差线的比较发现，基频方面含灯时的数据更集中，发音更稳定。开商方面，含灯时的数值整体高于正常时，这说明含灯时声带漏气程度总体大于正常时。同时由于两者的开商误差线长度相近，说明两者的漏气程度都保持在一个平稳状态。速度商方面，含灯与正常的速度商均值没有明显差异，但含灯时速度商的波动范围比正常时略大，这说明含灯时声带松紧的变化范围更大。

四、通过400多音节、语篇、曲目的基频箱式图比较发现，一般情况下正常的基频中位数要高于含灯的。基频在较高频率域上更集中。与朗读时相比，演唱时的基频波动区间更小，发音更为稳定。在开商方面，含灯时的开商数据分布范围没有正常时广，即含灯时声带漏气量更为稳定。朗读语篇时的开商数据分布范围最广，说明朗读时声带漏气量大小最不稳定，而演唱时则较为稳定。从速度商来看，正常和含灯的数据分布都较平稳，这说明两者嗓音的松紧程度都保持在一个稳定的状态。朗读时的箱体比其他语料都要长，这说明朗读时嗓音松紧变化大，而演唱时嗓音的松紧程度保持在一个平稳状态。大致来说，含灯时的速度商箱式

图的箱体比正常时更长，这说明含灯时的嗓音松紧变化比正常时更大。

五、正常和含灯的长时平均谱在 4000Hz-11000Hz 之间有明显差异，正常时的谱曲线高于含灯时约 10dB。曲目的谱曲线和语篇的类似，在高频区也出现了正常时的谱曲线高于含灯时的情况，但这个区间要比语篇的小，两者的差距也比语篇的小。并且两条谱曲线在 9000Hz 之后趋于并拢。另外，曲目含灯和正常时的谱曲线在 2000Hz-4000Hz 之间存在较为清晰的差异，含灯时的谱曲线要明显高于正常时。

本文发现含灯后语篇的长时共振峰 F1 的数值要比正常时稍大。一般来讲，第一共振峰和下颌开度相关，下颌开度大时第一共振峰高。含灯后由于下颌固定，开口度明显比正常时要小，但却出现了第一共振峰稍大的现象。如何解释这一现象还要做进一步研究。

关键字：含灯大鼓，基频，速度商，共振峰，长时平均谱

Acoustics Study of Handeng Opera

ZHANG Chunlian (Phonetics)

Directed by KONG Jiangping

ABSTRACT

In this paper, we studied the singing mode of a kind of Chinese traditional folk – the Handeng Opera using the method of acoustics. Handeng Opera is a performing style of the Plum Opera, which places a stick in the mouth of the performer during the performance. The stick in the mouth would affect the adjust of the entire vocal tract of the singer during pronunciation. For example, the jaw fixed, and it's not free to open and close the tooth and the mouth, the movement of the tongue is restricted, and the breath volume is limited, too. In this paper, we compared the acoustic parameters between normal and Handeng Opera of the same subject. These parameters include formant, Long-Term Formant Distribution(FFT), Long-term-average spectrum(LTAS) characteristics, fundamental frequency, open quotient, speed quotient. We would try to find out the differences of the different pronounce circumstances.

In this paper, we got the conclusions as follows:

1) The formant chart of Handeng Opera is significantly smaller than normal ones. This tells that the movement range of the tongue is reduced in the Handeng Opera. The F2 of /a/, /i/, /u/ decreased, indicating that the body of the tongue shifted to the back of the oral cavity. The F2 of the high vowel /i/ suffered the greatest impact, which reduced obviously.

2) From the FFT, we could find that the F1 of the Handeng Opera is a little higher than the normal ones. The distribution of F2 of the Handeng Opera is significantly higher than normal ones. The F3 doesn't show significant difference. The increase of F2 may be related with the increase degree of the left- right opening of the lips, the narrow of the front part of the vocal track, the reduction of the round degree of the lips.

3) When the object are /a/, /i/, /u/, /y/, the fundamental frequency of the ligh drum is smaller than the normal ones. In respect of open quotient, contrast to other vowel, the average of Handeng Opera /i/ is higher than normal ones. In respect of

speed quotient, the comparison result is the same as the open quotient. From the comparison of the error bars, the fundamental frequency of the Handeng Opera is more concentrated and steady. The open quotient of Handeng Opera is higher than the normal ones, which reflect that the blow-by degree of the Handeng Opera is higher than the normal ones. At the same time, the error bar length of the two kinds of mode is close, which reflect that they are keeping in a steady level. In respect of the speed quotient, there is no obvious difference of the average value between Handeng Opera and normal. But the fluctuation of the Handeng Opera is higher than the normal ones, indicating that the change of the tightness of the vocal cords is more obvious.

4) From the comparison between the Handeng Opera and normal ones of the box chart of the fundamental frequency of more than 400 syllables, the discourse, the tracks, we find that the median of the Handeng Opera fundamental frequency is normally lower than ordinary ones. The fundamental frequency is more focused on the higher frequency domain. Comparing with reading, the fluctuation of the fundamental frequency of the singing is smaller. In respect of open quotient, the range of the data distribution of Handeng Opera is smaller than the normal ones, indicating that the vocal cords blow-by of the Handeng Opera is more steady. The distribution range of open quotient data of reading discourse is the largest, indicating instability of the vocal cords blow-by. In respect of speed quotient, both the Handeng Opera and the normal has a steady data distribution. The body of the box is the largest of reading discourse, indicating that the change range of the tightness of the vocal cords is the largest. Broadly speaking, the body of the box of Handeng Opera is larger than the normal ones, indicating that the vocal cords tightness change of Handeng Opera is larger than the normal ones.

5) There are obvious differences of the LTAS between Handeng Opera and normal during 4000Hz – 11000Hz. The spectral curves of normal ones are about 10dB higher than the Handeng Opera ones. The similar situation ocured in singing. But the distance between the two curve is smaller in singing, and the domain is smaller, too. However, the difference between the two curve disappeared in singing after 9000Hz. In addition, there is obvious difference between the two curve during 2000Hz – 4000Hz in singing, during which the curve of the Handeng Opera is a little higher than the normal ones.

Through this paper, we find that the F1 of Handeng Opera is higher than normal ones. In general, F1 is related with the open degree of the jaw, and the F1 would be

high with a big jaw opening. With the stick in mouth, the open degree of the jaw would be small and fixed, however the F1 is higher than normal ones. We need more study to solve this question.

KEY WORDS: Handeng Opera, Fundamental Frequency, Speed Quotient, Formant, Long-term-average spectrum

目录

第一章 绪论	1
1.1 含灯大鼓背景综述	1
1.2 相关实验语音学综述	3
1.3 歌唱和语言学的关系	6
第二章 实验设备及方法	8
2.1 录音设备及分析软件	8
2.2 发音人及语料	9
第三章 单元音/a/、/i/、/u/共振峰分析	10
3.1 数据提取	10
3.2 单元音/a/、/i/、/u/共振峰分析	12
3.3 本章小结	13
第四章 语篇的长时共振峰分析	14
4.1 数据提取	14
4.2 语篇的长时共振峰分布情况	16
4.3 本章小结	19
第五章 基频、开商、速度商分析	20
5.1 数据提取	21
5.2 单元音 /a/、/i/、/u/、/y/的基频、开商、速度商	22
5.3 不同韵母音节的基频、开商、速度商分析	25
5.3.1 不同韵母音节的基频	26
5.3.2 不同韵母音节的开商	27
5.3.3 不同韵母音节的速度商	29
5.4 语篇和曲目的基频、开商、速度商	31
5.4.1 语篇和曲目的基频箱式图比较	31
5.4.2 语篇和曲目的开商箱式图比较	32
5.4.3 语篇和曲目的速度商箱式图比较	33
5.5 本章小结	33
第六章 语篇和曲目的长时平均谱分析	35
6.1 数据提取	35
6.2 语篇的长时平均谱	36

6.3 曲目的长时平均谱.....	37
6.4 本章小结.....	39
第七章 总结与展望	40
参考文献.....	41
附录 A 400 单音节	43
附录 B 曲目	46
附录 C 语篇《北风和太阳的故事》	49
附录 D Matlab 数据归一化时用到的程序	50
附录 E 归一化后的单元音共振峰.....	52
附录 F 长时平均谱数据.....	54
后记.....	68

第一章 绪论

含灯大鼓是一种始于清末民初的曲艺表演形式，它以梅花大鼓唱腔为主，本身并不是一个独立的曲种。含灯大鼓表演形式独特，演唱时演员口衔一根灯架，灯架另一端竖着几根点燃的蜡烛。演唱时演员不能将蜡烛吹灭。

“含灯”这种艺术表演形式具有一些特殊之处，演员因为嘴里含着灯架，发音时会受到很大影响。比如：

1. 牙齿咬着灯架，口不能随意开合，唇、齿的发音动作受到影响。这时演员整个口腔形状、下颌开度、声道长度都会受到很大影响。

2. 为使蜡烛不灭，发送气音时演员需要调节送气量，同时做到咬字清晰。

“含灯”演唱时，发声（phonation）和调音（articulation）都会发生一些改变。下颌开度会影响共振峰的高低，下颌开度增大，第一共振峰升高。唇形面积与第二共振峰关系密切，唇形面积较小时，第二共振峰较低，并且当唇形面积增大时，第一、第二、第三共振峰都会受到影响。

黄强，黄平（2001）^[1]对含灯大鼓进行过声学特征的初步分析。文中选取了“a、i”作为发音语料，分别采集“含灯人”和“普通人”说、唱形式的语音，用以观察第一、第二共振峰的变化情况。文章得出的结论是口衔竹筷会使元音的共振峰发生变化，但共振峰模式并没有发生改变，所以演唱者仍然能发出清晰的元音。

本文希望对“含灯”所使用的特殊发声、调音方式做更深一步研究。对“含灯”所带来的代偿性发音方式的研究，可以增加我们对发音生理上的认识，同时可以对人类在语音方面的感知做进一步探索。

1.1 含灯大鼓背景综述

随着历代说唱艺术表现形式的不断发展，中国说唱艺术逐渐形成了“南词北鼓”的格局：南方弹词，北方大鼓。大鼓由元明词话与各地民间曲调结合形成，一般认为清初形成于山东、河北等地的农村，主要流行于北方。唱词基本为七字句、十字句。早期曲目以中、长篇居多，有说有唱。后期多以短篇为主，只唱不说。曲目故事多取材于古典小说和戏曲。北方的大鼓又可细分为京韵大鼓、西河大鼓、梨花大鼓、梅花大鼓等曲种。

梅花大鼓又名“梅花调”。对梅花大鼓名字的由来有不同说法。其一，梅花大鼓创始人是旗籍子弟玉瑞，他精通诗词歌赋，别号梅花馆主。他创作演唱的鼓曲自称“梅花调”，也叫清口大鼓，为清客串（票友）演唱的意思。^[2]其二，此

曲种演出时，有三弦、四胡、琵琶、鼓、板等五位伴奏者坐在台上，形似梅花的五个花瓣，因得名梅花大鼓。另有说法认为，这一曲种兴起之初所唱曲目内容与梅花相关，故取名梅花大鼓。还有说法认为此曲种脱胎于全堂八角鼓的五音大鼓，而五音大鼓所使用的五种乐器犹如梅花的五个花瓣，故名梅花大鼓。^[3]总的来看，梅花大鼓在清代中期由满清子弟创作产生，使用五种乐器进行伴奏。^[2]

梅花大鼓长于叙事中抒情，节奏缓慢，字少腔多，缠绵哀怨色调较浓重。^[4]句式基本为七字句，偶尔有三字头。多为一人演唱，传统曲目有《湘子上寿》等，新编节目《秋江》、《张羽煮海》等也可以二人对唱。演出时，演员自击鼓板，伴奏乐器通常为三弦、四胡和琵琶。有时也加入扬琴（如五音联弹）、笙、笛等乐器。梅花大鼓传统曲目保留下来的共有三十三段，大都为清末梅花大鼓艺人王文瑞提供，经单弦艺人德寿山修改后交给金万昌演唱流传下来的。梅花大鼓传统曲目不少取材于《红楼梦》，如《探晴雯》、《黛玉悲秋》、《黛玉葬花》等，此外还有选自其它题材的曲目，如《鸿雁捎书》等。^[2,5]

梅花大鼓在清光绪年间从北京流入天津。当时天津已有艺人演唱北板梅花调，声音洪大，质朴无华。^[6]

民国六年（1917），南板梅花调创世金万昌应天津四海升平茶社之约，到天津首次演出经过改革的南板梅花调，贴演时名为梅花大鼓，首演曲目为《大观园》（带四季）。他嗓音洪亮宽厚，唱腔凄婉悲怨，不同于老梅花调的新颖而优美的唱腔大受观众欢迎。后他又多次到天津演出，但起初尚无演员正式学演。直至二十世纪二十年代中期，常州正规杂耍园演唱这一曲种的仍只有金万昌一人。

天津摩学梅花大鼓者多是在落子馆中教唱或是弹弦的师傅，他们掌握后再传授给落子馆的唱手和投师学艺者。天津相继出现的诸多梅花大鼓名角多是由邱玉山传授的。花四宝即为其学生。

在天津，卢成科、花四宝等人对梅花大鼓进行了改革，融入了时调旋律、唱法和特点，风格趋向天津化。此次改革突出的一点是丰富了段落之间的过门音乐。改革成功的梅花大鼓，被认为是“津派梅花大鼓”，又称为“卢派梅花大鼓”或“花派”，成为与“金派”梅花大鼓并提的新流派。四十年代，天津曲艺舞台的梅花大鼓在唱腔风格上日趋本土化，这引起了天津、北京曲艺评论者的关注，他们指出了北京、天津两地梅花大鼓的不同，认为北京的梅花大鼓更为温文儒雅，而在天津则更强烈、放得开。^[6]

50年代，著名弦师白凤岩为了说唱新的内容，又一次对梅花大鼓的板式、唱腔、唱法和过门进行革新，丰富了变调的艺术手法，灵活地插入曲牌和小曲曲调，称为“新梅花调”。梅花大鼓各个艺术流派的特点，是长于在叙事中抒情。它的慢板、中板，声腔婉转动听；快板、紧板活泼有力；收束时的慢板稳重而又

有余音。^[7]

梅花大鼓的演出形式通常为艺人自击鼓板站立演唱,唱段皆为短篇曲目。演唱者不必有表情和做工,因此梅花调又有“文大鼓”之称。此外,还有换手联弹、含灯大鼓、双鼓合音的表演形式,多是艺人用以展示其高难技巧以招徕观众而为。

含灯大鼓是梅花大鼓的特殊表演形式之一,主要展示演员口齿的特殊技能。演员口含装有多盏灯的灯柄演唱梅花大鼓,必须做到既含住灯不使其坠落,又要将字和腔唱得准确清楚。^[6]

含灯大鼓的灯具形制多种,其一:取三十余厘米长细竹,将细竹一段烤制弯折向上,放置一个点燃的小灯盏,内蓄香油。演员口衔细竹另一端,双手分执鼓槌、檀板演唱。另一种为竹木与金属合制:金属制成如眼镜架的形状,当中有竹木条,一段与架的前部相连,横担细竹或细木条,上置两个小灯盏,另一端衔在演员口中。亦有在架的两侧饰以绢花者。还有直接将瓦制小灯盏排放在窄木板上,演员口衔窄木板的中央演唱。演员演唱时,气息不得过大到吹灭灯焰;更不得使灯盏倾倒、坠落。行腔、吐字亦不能含混不清。这是难度很大的一种特技。^[8]

含灯大鼓自 20 世纪 20 年代后很少有人表演。1987 年著名弦师白奉林和陈邵武共同挖掘,将含灯大鼓这一技艺传授给了北京曲艺团的梅花大鼓演员王玉兰女士。

以往针对梅花大鼓的研究并不多,其中张静(2007)^[9]对梅花大鼓做了较为全面的介绍。黑杨(2011)指出“金”派演唱时嗓音洪亮宽厚,苍劲有力,韵味醇厚。“花”派为充分发挥女声特长,将唱腔特色发展为高亮,形成了“悲、媚、脆”的风格。^[10]

1.2 相关实验语音学综述

现代语音学研究主要分为两大部分:调音(articulation)研究和发声(phonation)研究。^[11]前者主要是关于元音音色、辅音发音部位及方法的研究,后者主要是对发声类型的研究。对于调音的研究基本上是看得见的唇形、开口度和舌位,对于发声的研究主要是通过声学分析的方法来实现的。发声研究的声学分析方法主要有频谱分析、多维嗓音分析等,语音信号可采用言语声波或喉头肌电波等。^[11]本文也是从这两个角度入手来研究含灯与不含灯的差异。

口腔前腔和后腔的变化都会影响到共振峰的变化。唇形面积的变化、舌位收紧点前后位置的变化、声道长度的变化都与共振峰有关。“F1 频率升高同开口度直接相关。频率越高,开口度越大;F2 降低同舌头后缩直接相关。频率越低,舌头越靠后;F2 与圆唇有关,唇越圆,F2 也越低;F3 的升降与软腭的升降有关,软腭降得越低,腭咽部的面积越大,F3 也越低;F3 的降低同-r 化音色有关,舌

尖上翘越高，-r 化音色越重，F3 就越低。”（Delattra,1951）。^[30]

吴宗济 等（1989）指出，下颌开度是指上、下门齿间的距离，颌骨运动与元音发音直接有关。通过元音 X 光图片的观察，文章认为下颌开度能够反映出元音音色的变化，并能表征元音舌位的高低。^[30]

在嗓音方面，声带振动方式的不同会形成不同的发声类型。发声器官由呼吸器官、声带和声道组成。肺部呼出的气流冲击关闭的声门引起声带振动，由此产生的气流脉冲即为声源，之后经由声道的调音、共鸣和唇辐射形成我们能听到的语音。

《论语言发声》^[12]中给出了汉语单音节发声变量模型，指出基频的变化与开商、速度商之间有密切关系。一般来说，基频的上升会带来开商的上升和速度商的下降。该模型使用的数据取自单独录制的单音节，并未受语流中韵律的影响，可以作为嗓音发声的基本模型。

董理（2012）中以昆曲为研究对象，分析了同一个人在使用昆曲不同唱法（老生唱法、小生唱法）以及按昆曲字音朗读时的语音、嗓音差异。文章主体分为三部分，第一部分为长时平均谱分析，第二部分为单元音共振峰分析，第三部分为 EGG 信号的基频、开商、速度商分析。^[13]

文章计算了长时平均谱和共振峰，“长时平均谱是一种常用而有效的分析连续语流或唱歌的工具。通常截取 30-40s 的语音数据，用其声压平均值来再现其频谱特性。”在进行计算之前，需要将辅音和空白段切除。

古典歌手的谱曲线在 2000Hz 到 3000Hz 之间存在明显的歌手共振峰，4000Hz 附近有演讲者共振峰。歌手共振峰是在 3000Hz 附近出现的大的频谱包络，之前研究表明，它主要是由第三、第四、第五共振峰簇产生的共鸣现象。歌手共振峰的水平 and 第一共振峰有关，这又和元音、声音响度和其他因素有关。它和元音共振峰频率相关是已经被证实过的。应用了语音产生的声学理论，已有对不同元音第一、第三共振峰差异水平的研究。研究发现专业歌手所唱的元音之间的差异要比未经过歌唱训练的人所唱的元音之间的差异大。^[25]歌唱共振峰的存在可以增强歌唱者嗓音的明亮度和穿透力，研究发现元音对歌唱共振峰的形态有直接的影响。^[14]歌唱共振峰的一个副作用是常常让人听不出所唱的元音是什么。文章在分析谱特征后认为，昆曲老生和小生唱法存在演讲者共振峰，而不存在歌唱共振峰；和朗读昆曲相比，演唱昆曲时的发音更中立化，朗读和演唱时口腔位置差异显著；在声道长度方面（F3 的大小和声道长度成反比），朗读时声道最短，老生次之，小生最长。

在嗓音方面，文章通过对基频、开商、速度商这三个参数进行差异性、相关性分析，发现昆曲这三种形式的三个参数在声学空间内的分布区域不同。

速度商大意味着声带紧,开商大意味着漏气较多。文章推测老生使用的是紧的气嗓音,小生使用的是松的气嗓音。最后,文章从语音、嗓音角度分别得出了较多结论。

曹洪林等(2009)通过分析/a/、/i/、/u/三个元音在正常和大声两种不同说话状态下基频、共振峰、谐波振幅差值等参数的不同,发现大声说话时的语音并不是正常说话时语音的简单放大。在声强上的差异是显而易见的,但两者在频率域上也有重要变化。状态不同时,同一个人语音的频谱特性也有较大的差异。^[16]

下颌开度(jaw height)对元音的影响较大,而对辅音的影响则不明显。舌位收紧点的位置受音量大小的影响较小,但由于舌体与下颌相关部位相连,舌位与下颌开度的变化常会相互影响。两者对元音的发音都会产生影响。

杨锋(2012)采集了四位发音人吟诵时的视频、语音、嗓音 EGG、胸腹呼吸信号,使用的语料包括五言、七言近体诗、古体诗、词和古文共 118 篇。以分析言语呼吸韵律为出发点,编写了 Matlab 程序,使用实验语音学的技术手段和语言学理论方法,比较了古诗词文在朗读和吟诵时发声、呼吸、韵律和旋律各方面的异同。^[15]

在嗓音方面,吟诵和朗读的差别主要在于吟诵时句末拖腔的发音特点较为明显。平声字时,吟诵的发音比朗读的要松,高频能量弱,表现为吟诵平声字拖腔时基频低、开商大、速度商小。入声字时,吟诵的拖腔呈现出紧嗓音的特点,表现为基频高、开商小。

在呼吸方面,无论是近体诗还是词,吟诵中的一级呼吸重置较为常见,而朗读时则多为一二级呼吸交替。具体来说,在近体诗朗读中,腹呼吸重置分为两级,一二级呼吸交替出现,而吟诵中基本为一级呼吸重置。在词的朗读中,一级呼吸重置只出现在上、下阕开始处,而在吟诵中,每个长句开始处都有一个一级呼吸重置。文章还指出,吟诵中的气息量要比朗读时增加数倍。

吉永郁代在《能剧嗓音发声研究——兼论汉传、藏传佛教诵经的声学特征》^[16]一文中对演唱中的嗓音发声机制和特征做了较为详细的分析研究。文章先讨论了嗓音发声的生理基础,之后介绍了从生理(电声门图)、声学(时域分析、频率域分析)角度出发对嗓音发声的研究方法,解释了电声门图的工作原理及其声学特征、电声门图各参数所对应的波形的物理意义。文章在提取嗓音参数时使用了 35%的尺度法。文章分别录制了能剧、佛教诵经和持续元音的语音、嗓音参数,以此为基础提取了基频、开商和速度商,使用统计分析的方法来观察不同参数的嗓音特性。

黄强,黄平(2001)中选用了两位发音人,一位是专业的含灯大鼓演员(简称“含灯人”),一位是普通人,两人分别说和唱普通话的/a/、/i/两音。^[1]结果显

示,与正常发音相比,含灯人口衔竹筷唱/a/时第一、第二共振峰都下降了,说/i/时第一、第二共振峰都上升了;普通人说、唱/a/时第一、第二共振峰都下降了,说/i/时第二共振峰上升了,唱/i/时第一、第二共振峰都下降了。文章认为以上数据符合以往研究成果,即第一共振峰频率越高,开口度越大,第一、第二共振峰数值随唇形面积增大而提高。但含灯人说/a/时,第一、第二共振峰都提高了,唱/i/时第一共振峰提高、第二共振峰降低了,文章认为这是与以往研究结果不一致的。

1.3 歌唱和语言学的关系

歌唱声学是五十多年来,特别是近一二十年来很活跃的一个主要由语言声学
与音乐声学交接而生长出来的新研究分支。^[17]

对于歌唱嗓音声学在西方音乐声学界已有较为广泛的研究。歌唱声学的研究
者们在研究时主要使用的是频谱分析的方法,不过随着计算机技术的迅速发展,
又有相关软件等技术出现。在研究内容上有歌手共振峰、歌唱声学基础、艺术嗓
音声学、歌唱技巧、听感和声学参量之间的关系、不同唱法存在的差异、声乐教
学等。

要研究歌唱声学,首先要先了解歌唱时的生理变化。歌唱时起到调节声道关
键作用的是放低喉头。^[26]包紫薇一文中指出,喉头降低会产生三种生理效果,
即咽底部扩张、喉腔和咽腔分立、声道拉长。这样的生理变化可能造成的频谱变
化为歌手共振峰的出现和元音 F1、F2 频率的降低。^[18]Soundberg 认为歌手共振
峰的成因是歌者降低喉头造成的,即在歌唱中歌者降低喉位,进而导致咽喉扩张,
扩张到使咽腔、喉腔借口处两腔截面积之比大于六分之一时,喉室会变成一个独
立于声道其他部分的共鸣腔,这个共鸣腔会为声道增加一个额外的共振峰。
Sundberg 对歌手共振峰的定义有过几次变化:1) 第四元音共振峰 F4; 2) 额外
峰 Fe; 3) F3、F4、F5 的聚集峰。西方和我国对歌手共振峰的研究基本建立在
Sundberg 的研究基础之上。^[33,34]

在歌唱声学的研究中,“歌手共振峰”是学界较为关注的一点。古典歌手的
谱曲线在 2000Hz 到 3000Hz 之间存在明显的歌唱共振峰,4000Hz 附近有演讲者
共振峰。歌唱共振峰的存在可以增强歌唱者嗓音的明亮度和穿透力,研究发现元
音对歌唱共振峰的形态有直接的影响。^[14]歌唱共振峰的一个副作用是常常让人
听不出所唱的元音是什么。

李小武(2012)利用语音分析技术提取了 30 名青年声乐专业学生歌声声学
参数的第一共振峰、第三共振峰、基频、音域、基频微扰、第一共振峰微扰、第
三共振峰微扰、平均能量,使用 BP 神经网络的方法评价了歌声质量。BP 神经

网络先用训练样本建立模型，用该模型预测待评价的样本。文章认为由训练样本建立的多参数艺术嗓音客观评价模型与主观听感评价结果有较好的一致性。^[19]

为了探讨我国戏曲唱法（如越剧、昆曲等）优秀的嗓音声谱中是否也存在歌手共振峰，于善英(2010)通过截取不同类型唱法的元音声谱进行了共振峰分析，并且还使用 CT 成像技术对不同类型唱法的歌手的喉腔进行了分析。研究的发音人为不同类型唱法的专业演唱者。分析时主要选取了元音/a/、/i/、/u/、/e/、/o/，研究发现歌手共振峰不仅在美声歌手的嗓音中出现，而且在民族唱法、昆曲及越剧演员的嗓音声谱中也存在。但不同类型唱法的歌手共振峰存在的高度、频域、密度又有各自的特点。^[22]

文章中给出的结果是：美声歌手的女高音歌手共振峰峰值为 3200Hz，女中音为 2800Hz，男高音为 2900Hz，较少达到 3200Hz，男中音约为 2700Hz，男低音约为 2600Hz；中国民族唱法的歌手共振峰集中在 2600Hz-3200Hz 之间，男高音为 2700-2900Hz，女高音为 2700-2900Hz；昆曲演员的歌手共振峰集中在 2500Hz-2800Hz 之间，不同角色又有不同；越剧演员的歌手共振峰集中在 2600-2800Hz，不同角色又有不同。

文章通过 CT 扫描发现，美声唱法中喉腔扩大至静态的 1.5-2 倍，越剧唱法中喉腔明显比静止状态缩小，昆曲中喉腔整体似乎缩小至静态时的二分之一，声区发声时比静态缩小近一倍，高声区发声时咽腔扩张接近于喉腔静止状态，民族唱法中咽腔的动态改变与美声唱法非常接近，发声中喉腔明显扩大。

陈彧（2012）利用 Praat 软件对美声唱法专业学生唱元音/a/的样本提取了共振峰参量、绘制了三维语图、频谱切片等，对其歌手共振峰的共振峰频率、共振峰强度、共振峰带宽做了声学测量。^[20]

第二章 实验设备及方法

2.1 录音设备及分析软件

本文从语音多模态研究的理念出发，除对发音人发音过程进行录像外，还同步采集发音人的语音、噪音和呼吸信号，组成一个小型的数据库。此次实验中用于分析的有四路信号。第一路是语音信号，用麦克风采集。第二路是噪音信号，由电子声门仪（EGG）采集。第三路是胸呼吸信号，第四路是腹呼吸信号，两者分别由两条呼吸带采集。

数据采集集中用到的仪器有：摄像机、麦克风、电子声门仪、呼吸带、调音台、Powerlab 十六通道采集器。其中，摄像使用的是索尼高清数码摄像机（HDR-PJ790E）；麦克风使用的是 Sony 公司的 Sony Electret Condenser；电子声门仪使用的是 VoceVista 公司的 ElectroGlottograph M7050A 型号电子声门仪；呼吸带使用的是澳大利亚 ADInstruments 公司的 MLT1132 压电呼吸带传感器（MLT1132 Piezo Respiratory Belt Transducer）；调音台使用的是德国 Behringer 公司的 XENYX302 型调音台。

录制过程中，语音和喉头仪信号经由调音台连入电脑。胸呼吸和腹呼吸信号则直接通过十六通道采集器连入电脑，如图 1 所示：

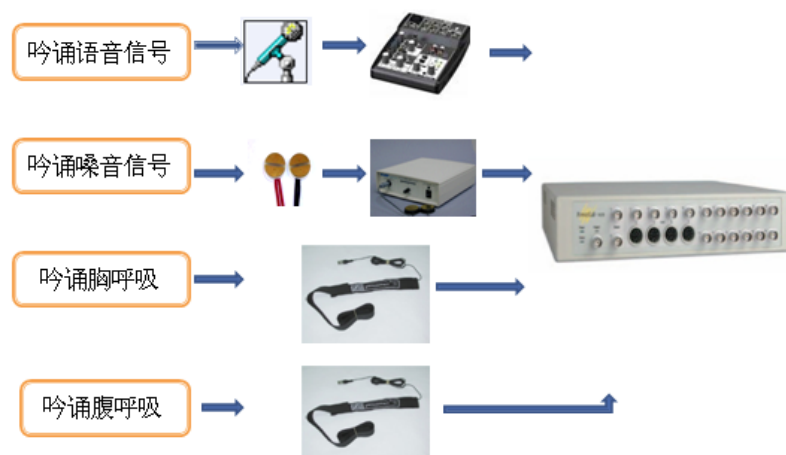


图 2.1 设备连接示意图（杨锋，2012）

在软件使用方面，信号采集时，使用 ADInstruments 公司的 Powerlab Chart5 for Windows；噪音信号提取基频、开商、速度商时，使用美国 Kay 公司的 Multi-Speech Signal Analysis Workstation Model 3700；语音信号共振峰提取时，使用的是由荷兰阿姆斯特丹大学人文学院语音科学研究所的主席保罗·博尔斯马教授和大

卫·威宁克助教授研发的 Praat 软件；长时共振峰提取时使用瑞典皇家理工学院开发的 Wavesurfer 软件；提取长时平均谱时使用的是 Wavesurfer 软件；数据统计分析时使用的是 Excel 软件、Matlab 软件和 SPSS 软件。

录制时的数据采样频率为 20kHz，采样精度为 16 位。

2.2 发音人及语料

本文实验所有语料的发音人均均为著名含灯大鼓表演艺术家王玉兰女士。王女士师从白奉林，早年从事梅花大鼓的表演，后转为含灯大鼓表演。王女士录音时 57 岁，身体健康，无嗓音疾病或感冒。

第三章使用的语料为汉语普通话中的单元音/a/、/i/、/u/，每个音发音时持续 3 秒以上，每个音连续发 5 遍。第四章使用的语料为语篇《北风和太阳的故事》以及普通话中的单元音/a/、/i/、/u/、/y/。第五章使用的语料为汉语普通话单元音/a/、/i/、/u/、/y/，汉语普通话 400 多音节，语篇《北风和太阳的故事》，曲目《万里春光》、《重整山河待后生》、《鸿雁捎书》、《黛玉悲秋》、《两头忙》、《梅花的故事》、《前门情思大碗茶》。（详细参见附录）

由于含灯录音时需要先将灯架上的蜡烛点燃，所以我们先录制含灯的发音再录制不含灯的发音。具体流程如下表所示：

1. 单元音/a/，/i/，/u/，/y/。
2. 汉语拼音 400 多个单音节。
3. 语篇《北风和太阳的故事》。
4. 曲目。

表 2.1 录音项目及流程

项目	使用设备	采集信号类型	提取数据	录制类型
1	麦克风、喉头仪	语音、嗓音	基频、开商、速度商	含灯读五遍 正常读五遍
2	麦克风、喉头仪	语音、嗓音	基频、开商、速度商	含灯读一遍 正常读一遍
3	麦克风、喉头仪、呼吸带	语音、嗓音、呼吸	基频、开商、速度商	含灯读两遍 正常读两遍
4	麦克风、喉头仪、呼吸带	语音、嗓音、呼吸	基频、开商、速度商、呼吸	含灯唱一遍； 正常唱一遍

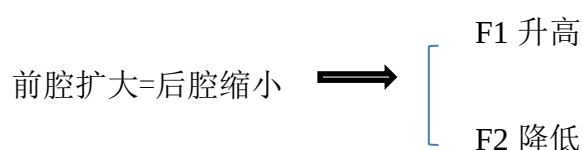
第三章 单元音/a/、/i/、/u/共振峰分析

共振峰本义指声腔的共鸣特性，在语图上是某个特定频率区域聚集较大声能的语音表现。对元音来说，前三个共振峰对元音音色有决定性作用，可以根据这三个共振峰的不同来区分不同元音。其中 F1、F2 和舌位、唇形的变化有密切关系，因而在语音学上常使用 F1、F2 的数值作为描写元音音色的依据。共振峰在频率轴上的排列形式被称为“共振峰模式”。元音音色不同，共振峰模式也会不同。

在语音实验中，元音的音质主要决定于第一共振峰(F1)和第二共振峰(F2)的数据。利用语音实验测得元音的共振峰频率，可以绘制声学元音图，又称声位图。^[21]声学元音图可以表现元音的主要声学特征。绘制声学元音图时，以 F1 为一维，以 F2 为另一维。声学元音图虽然不能完全和生理舌位图相同，但是在相对位置上大致对应。^[21]本文通过提取单元音/a/、/i/、/u/的共振峰数据，绘制出含灯与正常发音的声学元音图，以此来进行比对，观察“含灯”对发音造成了哪些影响。

Delattre (1951) 提出，口腔“开度”指的是上下门齿间的距离。^[30]他认为 F1 频率升高同开口度直接相关。频率越高，开口度越大。F2 降低同舌头后缩直接相关。频率越低，舌头越靠后。F2 与圆唇有关，唇越圆，F2 也越低。F3 的升降与软腭有关，软腭降得越低，腭咽部的面积越大，F3 也越低。F3 的降低同-r 化音色有关，舌尖上翘越高，-r 化音色越重，F3 就越低。

口腔前腔和后腔的变化都会影响到共振峰的变化。实际发音中，前腔和后腔面积往往是互补的，声学效果大致是：



另外，唇形面积的变化、舌位收紧点前后位置的变化、声道长度的变化都会影响到共振峰的变化。^[12]

本文提取了含灯与正常时的共振峰参数，将含灯与正常情况的声学元音图进行比对。同时对语篇和曲目在含灯和正常状态下的语音进行了长时共振峰分析。

3.1 数据提取

我们使用 Praat 软件来进行共振峰提取。选择发音人含灯和正常时各 5 遍的发音作为分析对象，排除听觉和/或语图中显示的共振峰差别过大的音频（至少有两两次以上发音相近才进行分析）。

具体操作步骤为：

1) 参数设置：通过 Open-read from file 打开一个语音文件，之后通过 View&edit 进入分析界面。在 Formant 菜单里勾选 Show formant, 之后选择 Formant settings 进行共振峰参数设置。此处的 Maximum formant (Hz) 虽然在分析成年女子的语声时原则上设定为 5500Hz，但我们将实际观察到的共振峰位置与程序所画的共振峰位置进行比对，通过将共振峰上限分别设置为 4500Hz、5000Hz、5500Hz、6000Hz 来得到最为准确的共振峰标记位置。Number of formants 也根据情况分别设置为 4、5、6 等，选择能够得到最佳共振峰位置的数目。Window length (s) 设置为 0.025，Dynamic range (dB) 设置为 30，Dot size (mm) 设置为 1.0。点击 Apply 键将以上设置应用。

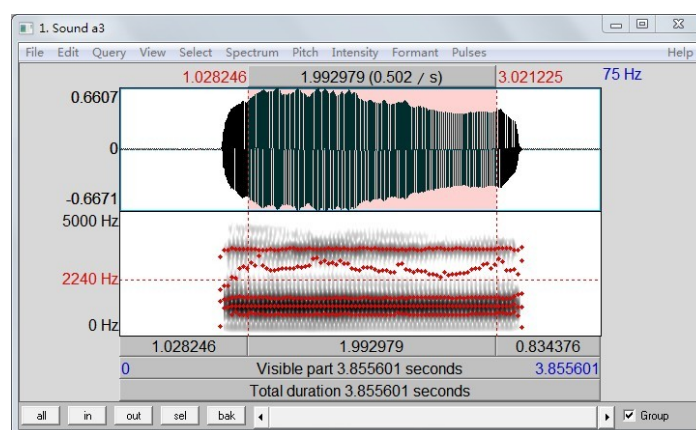


图 3.1 Praat 软件提取单元音共振峰示意图

2) 共振峰提取：完成参数设置之后用鼠标选择共振峰比较稳定的语音段，通过 Formant-Formant listing 得到选中段的共振峰数据，之后复制到 Excel 表格中进行使用。由于每个音频文件提取共振峰数据时选择的时间长度都不相同，所以我们将所有共振峰数据进行了归一化处理，最终每个音得到 30 个点。（使用的 Matlab 归一化程序详见附件）

Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
1.028800	800.657702	1147.928691	1494.983492	2684.226807
1.059800	866.281606	1149.386419	1511.176432	2888.485803
1.090800	843.521110	1146.199553	1495.504942	2801.846795
1.121800	842.520270	1149.050671	1494.094590	3159.130623
1.152800	833.161248	1144.628864	1487.136284	2786.625251
1.183800	814.453289	1153.082571	1488.582208	2668.817656
1.214800	811.308313	1147.402125	1474.020734	2576.763112
1.245800	814.506555	1143.282296	1470.111658	2591.319141
1.276800	830.047785	1142.220871	1477.770511	2629.978524
1.307800	797.113313	1148.843780	1475.217559	2675.665655
1.338800	791.712429	1156.680284	1481.391104	2686.349157
1.369800	797.571289	1164.066548	1490.165606	2669.902808

图 3.2 Praat 软件提取单元音共振峰时的数据列表表示图

3.2 单元音/a/、/i/、/u/共振峰分析

经过处理后，分别得到了/a/、/i/、/u/三个元音的共振峰数据，它们的均值如下表所示：

表 3.1 单元音/a/、/i/、/u/的共振峰平均值

正常	F1 (Hz)	F2 (Hz)	含灯	F1 (Hz)	F2 (Hz)
a	828.4107	1196.946	a	764.0205	1137.121
i	398.5099	2950.13	i	442.5861	2430.982
u	586.3989	1007.365	u	578.9736	909.9308

将上表中的数据做成声学元音图可以更直观的看到含灯与正常时的差异，如下图所示。图中点右侧的文字分别标记了正常或含灯的状态，图中的数据是由全部同类音节的共振峰数据平均得来的。

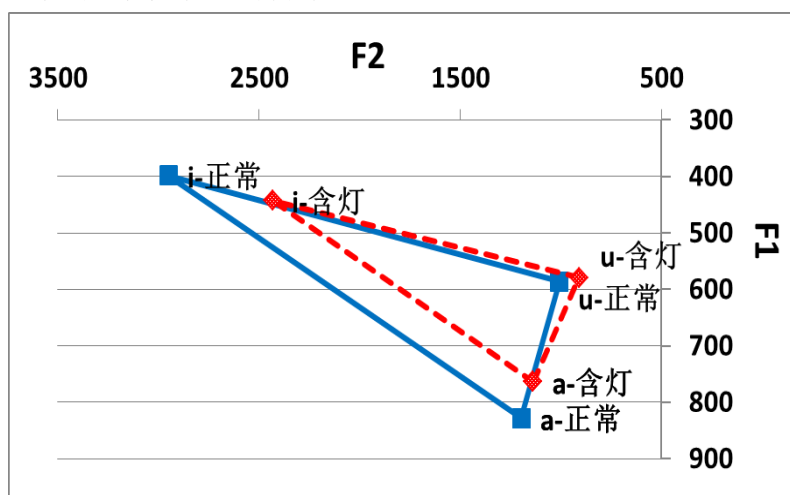


图 3.3 通过 Praat 提取的正常和含灯情况下单元音/a/、/i/、/u/的声学元音图（蓝色实线代表正常时的声学元音图，红线虚线代表正常时的声学元音图）

从上图中可以直观的看到，整体来说含灯后的声学元音图面积明显比正常时变小了。其中，/a/、/i/、/u/三个元音的位置都退后了，即 F2 的值都降低了。F2 的降低与舌头的后缩直接相关，由此可见发音人含灯时发音的舌位整体发生了后移。同时，F2 的升降不仅关系到发音位置的前后，它还与唇形面积的大小有关，唇形面积缩小，F2 降低。

在/a/、/i/、/u/三个元音中，/i/音在含灯时 F2 降低的最明显，降低了约 520Hz。/a/和/u/的 F2 也有一定降幅，分别下降了约 60Hz 和 100Hz。从 F1 的变化来看，含灯后/i/音 F1 数值有小幅提升，约 45Hz。含灯后/a/和/u/的 F1 数值虽然都有降低，但降幅不大，分别下降了约 64Hz 和 10Hz。一般来讲，F1 的大小和开口度有直接的关系，开口度越大，F1 越大。正常的/i/音是齐齿呼，开口度不大，而含灯时由于齿间衔着灯架，唇的左右开度增加，进而使得 F1 数据提高。/a/和/u/

音的 F1 都有小幅下降，这可能是由于含灯后牙齿咬着灯架，进而使得唇的上下面积减小造成的。

3.3 本章小结

本章通过使用 Praat 软件分析了普通话单元音/a/、/i/、/u/正常和含灯的共振峰情况，通过观察声学元音图发现，含灯后的声学元音图面积明显比正常时缩小，并且有整体后移趋势，高元音/i/的 F2 降低最为明显。

第四章 语篇的长时共振峰分析

目前,长时共振峰分布特征是声纹鉴定中的一项重要研究特征。长时共振峰分布测量法(Long-Term Formant Distribution, LTF)由 Nolan 和 Grigora 提出。^[28]该方法分析一段语音中的所有元音部分的共振峰分布情况。长时共振峰分布情况可以体现发音人的整体共鸣特点,也可以反映一个人的发音习惯,并且个体稳定性较强。已有研究表明长时共振峰可以较好的区分不同发音。^[23,27]

此方法在德国联邦调查局语音分析实验室得到了广泛的应用。该方法的优点为:比较省时;不同鉴定人员对同一语料分析的一致性较高;不同语言间的差异较小等。有研究表明,朗读时的 LTF 要比自然说话时稍高。^[25]目前对汉语的 LTF 进行的研究还比较少。

这一方法的缺点是它集中了所有元音的信息。这不同于分析单个元音,可能会忽略一些更有价值的个性特征,并且在解释它的结果时会变得更加困难。另外需要注意的是,电话通道的带宽一般是 300-3400Hz,所以多数情况下只能显示出前三条共振峰,如果录音质量差的话只能显示出两条共振峰。而 F3、F4 这两个高次共振峰的稳定性最强,个体差异最大,这在鉴定工作中是很有价值的。所以使用此方法的前提是录音质量要好,可以呈现出清晰的共振峰结构。

4.1 数据提取

本节使用 WaveSurfer 软件对语料提取共振峰参数,提取语篇《北风和太阳的故事》的前 3 条共振峰稳定段的中心频率(F1、F2、F3)。

我们使用 Wavesurfer 软件对语音文件进行共振峰数据提取。在处理语篇录音时,首先将语音文件中的空白段、辅音段以及共振峰不明显的元音部分切除,只保留共振峰清晰稳定的部分。在这个步骤中尤其要注意将浊辅音[m]、[n]、[ŋ]切除掉,它们在语图上表现为类似于元音的共振峰。之后使用 Wavesurfer 软件提取第一、第二、第三共振峰。提取共振峰时参数设置如下:FFT 窗长:512 个点,共振峰数量:3,分析窗口类型:哈明窗,LPC 系数:12,截止采样频率:10kHz。如果软件自动提取的数据不正确,通过手动操作进行调整。为了进行对比参照,我们使用 10kHz 作为提取样本的截止采样频率,将 11kHz 作为参考样本。

具体操作步骤如下:

1) 使用 Adobe Audition 将音频中的空白段、浊辅音段、共振峰不明显的部分切除(由于在 Wavesurfer 中每进行删除操作一次程序都会重新进行共振峰计算,

所以会影响电脑运行时间)。将已经切除空白段的音频保存后用作 Wavesurfer 程序使用。

2) 打开 Wavesurfer1.8.8p4, 导入上个步骤中处理过的数据。鼠标右键单击后选择 Create Pane-Spectrogram, 之后右键选择 Apply Configuration, 选取“LTF 切音_female”(详见附录)将整个音频切分为几个小段, 每段时长为 4s, 此时被切分音频所在的文件夹会有一个新的文件夹, 里面存储的是切分好的音频文件。(将音频切分为 4s 长的小段音频是由于如果音频时间较长的话, Wavesurfer 在提取共振峰时会有小的偏差, 而小段音频提取出的共振峰则更为精确。)

3) 再次将切分为 4s 的音频文件导入 Wavesurfer, 鼠标右键单击选择 Create Pane-Spectrogram, 之后右键选择 Apply Configuration, 选取“LTF_female”(详见附录), 此时会出现上下两栏共振峰提取示意图, 我们以上面一栏(11kHz)为参考样本, 对下面一栏(10kHz)提取出来的三条共振峰曲线进行手动调整, 使每条共振峰提取的都比较准确。

4) 调整完共振峰曲线之后, 鼠标右键单击后选择 Save data file, 将数据存储为.frm 格式, 之后可复制到 Excel 文件中使用。由于我们将整个音频文件切分成了几个小文件, 所以最后又将所得数据进行了累加。

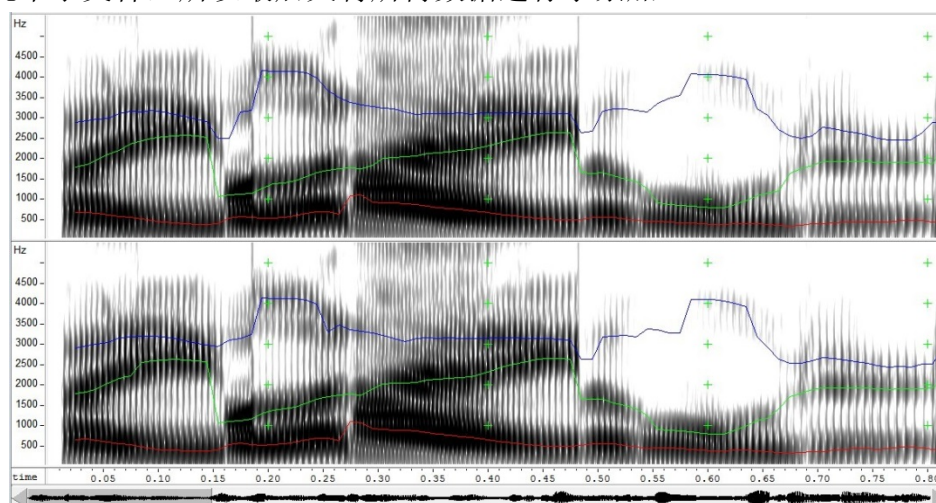


图 4.1 使用 Wavesurfer 自动提取的一段音频的前三个共振峰曲线从下至上的红线、绿线、蓝线分别代表第一、第二、第三共振峰(F1、F2、F3)

图 1 是使用 Wavesurfer 程序自动提取的共振峰数据曲线示意图, 可以观察到, 曲线在拐角处会有小的误差, 我们可以用鼠标将该线条拖住进行改动。

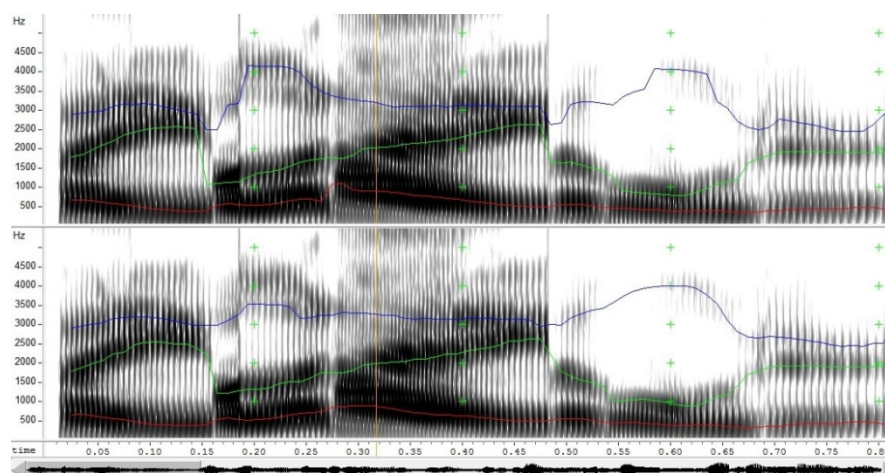


图 4.2 使用 Wavesurfer 手动修改后的一段音频的前三个共振峰曲线从下至上的红线、绿线、蓝线分别代表第一、第二、第三共振峰 (F1、F2、F3)。

通过鼠标拖动线条，我们将提取出的共振峰曲线有偏差的部分进行了调整，得到了图 2 中的曲线。其中，下面一栏为我们最终要保存的数据。

通过以上步骤得到三条共振峰的数据之后，将数据导入 Matlab 进行统计分析，画出一定频率范围（本文使用分箱间隔为 45Hz）数据出现频次的拟合曲线。我们画图时使用的 Matlab 命令为：

```
x=(0:100:4500);%将 0-4500 分为 100 分，即确定分箱间隔为 45
```

```
y=histc(a,x);%a 为正常时 F1、F2、F3 的数据
```

```
z=histc(b,x);%b 为含灯时 F1、F2、F3 的数据
```

```
plot(x,y);
```

```
hold on
```

```
plot(x,z,'--');%为了与正常发音时的长时共振峰频次拟合曲线相区别，将含灯时的长时共振峰曲线设置为点线。
```

4.2 语篇的长时共振峰分布情况

我们将含灯与正常情况各自的两遍录音的共振峰数据进行平均，得到两个文件。它们的长时共振峰拟合曲线如图 3 所示。

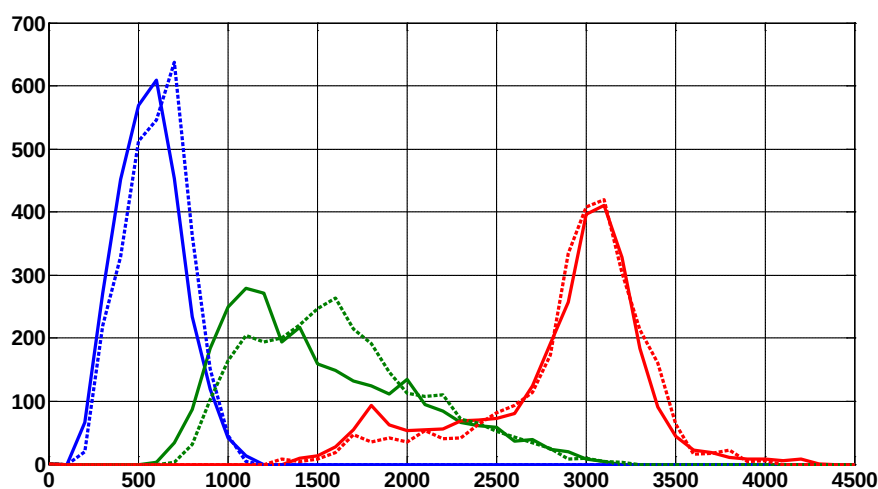


图 4.3 同一发音人含灯与正常情况下语篇的 LTF1-LTF3 分布情况(图中蓝线、绿线、红线分别代表 F1、F2、F3 的长时分布情况，其中实线为正常时的共振峰数据，虚线为含灯时发音的共振峰数据)

从图中可以观察到，F1 含灯时整体比正常时稍高，F2 整体分布范围明显比正常时要高，F3 没有明显变化。

F2 的明显升高与以下几个因素有关：首先，含灯后由于灯架处于上下齿之间，唇的左右开度明显增加，尤其在发齐齿呼和撮口呼时更为明显。其次，F2 的大小和舌头的前后位置有关，舌头越靠前 F2 越大。含灯后由于口腔前腔减小，舌头所处的相对位置比正常时更靠前。另外，F2 的大小还和圆唇程度有关，唇形越圆 F2 越低。含灯后由于灯架的阻碍，发音人在发音时几乎不可能做出圆唇动作，所以这也造成了含灯时的 F2 比正常时更高。

值得注意的是，含灯后的 F1 比正常时略高。以往研究发现 F1 与下颌开度成正比，下颌开度大时 F1 更高。含灯后为了保持灯架平稳，下颌处于固定位置，所以下颌开度只会比正常时更小，但此时得到的 F1 数据比正常时略高。

另外从整体来看，含灯和正常时的第一、第三共振峰的 LTF 分布范围更为集中，呈现出明显的“尖峰”(leptokurtic)，而第二共振峰的 LTF 分布范围较大，呈现“扁平”(platykurtic)状态。Nolan (2005) 指出，LTF 分析中共振峰的分布范围不仅反映了说话人的声道面积，而且还反映了说话人的习惯，比如倾向于使用腭化或圆唇。^[28]对于分析来说，各个共振峰的分布形状可能也是有用的。比如，一些发音人的高次共振峰更稳定，所以会形成更尖的峰 (leptokurtic)，相反则会形成扁平的分布 (platykurtic)。在图 2 中可以观察到，发音人第二共振峰的 LTF 分布较为扁平。这反应了说话人发音时的发音习惯。

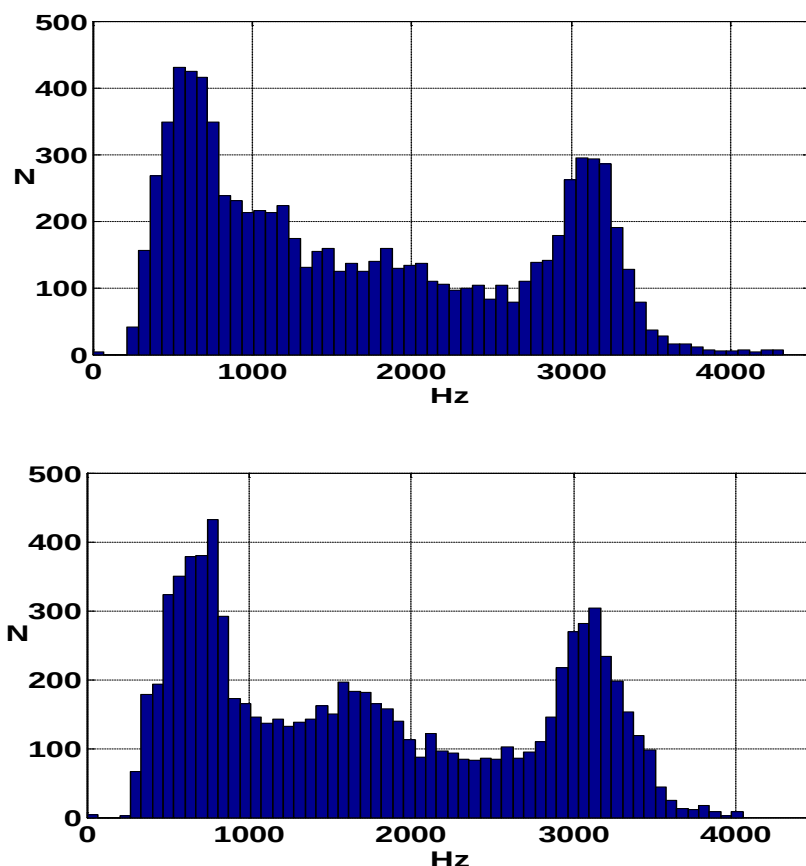


图 4.4 正常（上图）和含灯（下图）情况下的共振峰频次直方图（正常时第一、第二、第三共振峰的众数分别为 542Hz、1203Hz、3017Hz；含灯时的第一、第二、第三共振峰的众数分别为 734Hz、1749Hz、3138Hz）

对比图 4.4 中正常和含灯情况下共振峰出现频次的直方图可以发现，正常时三个共振峰的众数（分别为 542Hz、1203Hz、3017Hz）都要低于含灯时的众数（分别为 734Hz、1749Hz、3138Hz）。其中 F2 的众数值相差最大，含灯时要高于正常时 546Hz。含灯时 F1 的众数比正常时高出 192Hz，F3 的众数比正常时高出 121Hz。

上文中已经通过拟合曲线讨论了 F1、F2 的相关情况，此处通过具体数据可以得知含灯后 F3 也发生了变化。含灯后的 F3 虽然没有 F1、F2 变化幅度大，但众数还是有 100Hz 多的提升。F3 的提高和软腭上升有关，含灯后伴随着前腔变小，后腔增大，软腭比正常时有所抬升。

表 4.1 同一发音人正常和含灯情况下 LTF1-3 的均值（m1）、中位数（m2）、众数（m3）和标准差（sd）

LTF (Hz)	LTF1				LTF2				LTF3			
	m1	m2	m3	sd	m1	m2	m3	sd	m1	m2	m3	sd

正常	614	609	542	177	1573	1455	1203	525	2893	3033	3017	500
含灯	655	658	734	170	1671	1615	1749	474	2957	3054	3138	447

上表中给出了正常和含灯情况下各长时平均共振峰详细的均值、中位数、众数和标准差。对正常和含灯情况的数据进行比较后,对较大的数值进行了加粗。从表中可以清晰的观察到,均值、中位数、众数的数值都是含灯时高于正常时,而标准差都是正常时高于含灯时。其中 F2、F3 的标准差差距较大,正常时比含灯时分别高出 51 和 53。标准差反应了组内个体间的离散程度,标准差越大,表明样本数据的波动越大。由上表中的数据可以得知,正常和含灯时的 F1 标准差相近,说明两种状态的 F1 波动相近。但是正常时的 F2、F3 要远大于含灯时,可见含灯时 F2、F3 的波动较小更稳定。

4.3 本章小结

本节分析了同一发音人含灯和正常时语篇、单元音的长时共振峰数据。

通过三个单元音的声学元音图可以发现,含灯后声学元音图的分布范围明显比正常时缩小,同时三个元音发音时舌位所处的位置也发生了改变,总体来看一方面明显后移,一方面更向中部集中。三个元音的 F2 都出现了下降的情况,其中 /i/ 音 F2 下降最明显。

通过观察 LTF 分布情况可以看出,含灯时的 F2 总体分布范围比正常时显著提升, F1 有少量上升, F3 虽然从拟合曲线上不能直观的看出变化,但数据分析后发现也存在小幅的上升。

从语篇共振峰直方图的“尖峰”和“扁平”区域可以看出发音人在含灯和正常时的发音习惯不同。含灯时 F2 所占区域的数值比正常时明显上升,这可能是由于唇形面积和 F2 大小成正比造成的。含灯时唇形面积受灯架影响而变大,因而 F2 升高。含灯时的 F3 众数数值大于正常时,这可能是由于声道变短、软腭提升、腭咽部面积缩小造成的。

本文在数据分析方面还有一些问题没有解决。F1 与下颌开度或者说开口度有关,开口度变小时 F1 通常会减小。但图 4.3 中含灯时 F1 的均值为 655,众数为 734Hz,明显比正常时的 614Hz、542Hz 要高。含灯时由于牙齿咬着竹棍,开口度明显比正常时更小,所以 F1 频率理论上应该比正常时更低。对于出现这一现象的原因还需要进一步研究。另外,单元音分析中含灯时的 F2 比正常时要低,而语篇的长时共振峰则是含灯时 F2 比正常时更高。这可能和语篇中受辅音音轨影响、存在协同发音、存在多个韵母有关,两者关系较为复杂,并非简单的正相关关系,需要做进一步分析才能得出结论。

第五章 基频、开商、速度商分析

我们对声音信号频率的感知表现为音调的高低，称为音高。音高是语音的特性之一。研究声调时一般都以分析音高为主。

通常我们说的语音中的音高实际是指基频的高低。基音是每个乐音中频率最低的纯音，其强度最大，基音的频率即为基频（Radical frequency），又叫 F_0 。基频 = $1/\text{周期}$ 。

我们在研究嗓音发声类型时主要采用生理和声学的方法。生理方面的研究主要通过分析电声门图所采集的数据来进行，声学方面则从时域、频率域（频谱测量方法）方面进行探讨。

在嗓音研究中使用的仪器主要是电声门仪。电声门仪又叫 Electroglottography，简称 EGG。它在语音、医学研究领域的广泛应用是在 1957 年经过 Fabre 改良之后。

EGG 测量的是声带接触面积，工作时将一对电子感应片固定在喉结两侧，其中一个电子感应片发送电流，另一个接收电流，通过微弱的高频交流电测量人体组织的阻抗。在声门打开时，喉部充满气体，电阻抗很大，电流不易通过；声门关闭时，电阻抗较小，电流较容易通过。电流信号不仅只通过声带，它还通过身体的其它组织。声带开关时，电子声门仪的电子感应片之间的阻抗会产生差异。

电子声门图上显示的信号和声带开合之间有密切关系，如图 5.1 所示：

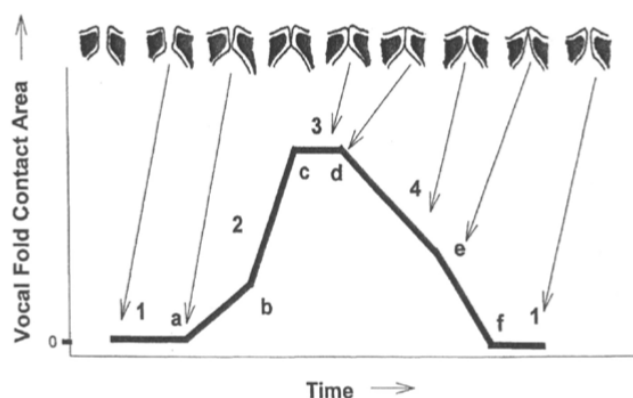


图 5.1 电声门信号和声带开合之间的关系 (Baken, 2000) ^[32]

电声门图的波形与声带的不同运动状态紧密相关。电声门图信号周期信号可以分为两个部分：开相和闭相。闭相又包括关闭相和开启相。

在下图中，ad 为一个声门运动周期，其中 ac 为闭相，cd 为开相，ab 为关闭相，bc 为开启相。

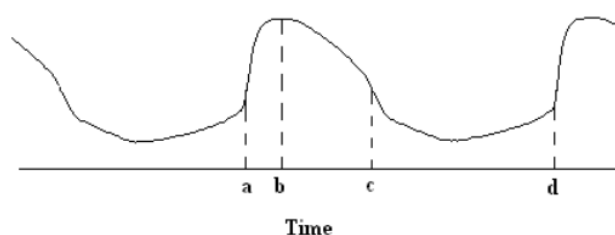


图 5.2 EGG 信号示意图

开商（Open Quotient）等于开相和周期之比（开商=开相/周期）。

速度商（Contact Quotient）等于开启相和关闭相之比（速度商=开启相/关闭相）。

速度商大意为着声带紧，开商大意味着声门漏气多。基频的升高会伴随开商的提高、速度商的下降。速度商与音调高低成反比，开商与音调高低成正比。

《论语言发声》中给出了汉语单音节发声变量模型，指出基频的变化与开商、速度商之间有密切关系。一般来说，基频的上升会带来开商的上升和速度商的下降。^[13]

在本文中，我们通过 Kay 公司的 Multi-Speech Model 3700 程序提取了汉语普通话单元音/a/、/i/、/u/、/y/，38 个不同韵母音节，1 个语篇和 7 个曲目的基频、开商、速度商，通过计算均值、标准差，绘制误差线、箱式图等方式，用以观察含灯和正常状态下三个参数的差异。

5.1 数据提取

我们前期通过电子声门仪采集了语料的语音和嗓音信号，之后将双通道的信号使用美国 Kay 公司的 Multi-Speech Model 3700 软件进行分析，对单元音/a/、/i/、/u/、/y/进行了基频、开商、速度商的参数提取。

操作步骤如下：

1) 打开 Multi-Speech Model 3700，选择 Real-Time EGG Analysis。先进行参数设置，选择 Options-Select Real-Time EGG Options，在 Analysis 选项下设置 Pitch Analysis Range (Hz)，将 Minimum 设置为 70，将 Maximum 设置为 800。在 EGG Quotient 里使用 “Use Peak Amplitudes-30”。在 Quotient Type 中，如果提取的是开商的话，选择 Open Quotient，如果提取的是速度商的话，则选择 User defined 里的 Opening Phase 和 Closing Phase。最后将设置应用。

2) 打开一个语音和嗓音双通道的音频文件，选择 Analyze-Produce Numerical Results，将所得的 EGG Analysis Results 数据保存为.OUT 文件（之后可以用.txt

程序打开)。

5.2 单元音 /a/、/i/、/u/、/y/的基频、开商、速度商

通过数据提取,分别得到了这四个元音正常和含灯时各参数的数据平均值,如下表所示:

表 5.1 /a/、/i/、/u/、/y/四个单元音在含灯和正常状态下的基频、开商、速度商均值比较(加粗的为两种情况中较大的数值)

	基频均值 (Hz)	开商均值 (Hz)	速度商均值 (Hz)
正常/a/	288.5	44.9	491.0
含灯/a/	286.1	43.6	520.9
正常/i/	296.6	40.1	532.7
含灯/i/	295.7	41.2	522.4
正常/u/	304.8	44.9	503.1
含灯/u/	298.9	46.9	518.8
正常/y/	303.8	39.1	548.9
含灯/y/	303.6	41.0	493.5

从上表可以观察到, /a/、/i/、/u/、/y/四个单元音的基频在正常情况下的平均值都高于含灯时的平均值,其中/u/音含灯和正常时基频的均值差异最大,含灯时约高出 6Hz, /y/音的基频均值差异最小。但在开商、速度商方面四个音又有各自的特点:开商方面,高元音在含灯时的均值要高于正常发音时的均值,其他元音则是正常时的均值高于含灯时的均值;速度商方面,前元音在正常时的均值要大于含灯发音时的均值,其他元音则是含灯时的均值大于正常时的均值。

为了进一步比较单元音正常和含灯情况下嗓音的差异,我们将/a/、/i/、/u/、/y/四个元音含灯和正常发音的基频、开商、速度商的数据做了误差线分布图。数据处理的方法:每个单元音含灯和正常状态各有 5 次发音,先将所有音提取出来的基频、开商、速度商的数据通过 Matlab 程序归一化为 30 个点,之后分别将含灯和正常这两个类型的所有音的基频、开商、速度商进行序列上的平均,最后两个类型分别得到 30 个平均后的点,利用均值和标准差画出误差线。此处使用的是一倍标准差的误差线,代表了数据主要的分布范围。

下表显示的为/a/、/i/、/u/、/y/四个元音正常和含灯时基频的均值和标准差。为了便于呈现 30 个点,我们在表格中每行放置 10 个点,每个类型共三行。

表 5.2 单元音/a/、/i/、/u/、/y/在含灯和正常状态下的均值、标准差比较(加粗的为两种情况中较大的数值)

正常基频均值	300.6	300.5	300.5	300.6	301.0	300.4	299.8	300.1	299.2	299.7
标准差	10.7	10.4	11.0	10.7	9.9	10.5	11.2	10.0	11.8	12.2

正常基频均值	299.1	298.7	297.5	298.1	298.2	299.2	297.7	298.3	297.1	298.1
标准差	11.5	12.3	12.1	12.2	11.9	11.5	12.3	12.2	11.9	12.2
正常基频均值	297.5	297.7	296.6	297.4	297.1	297.1	296.4	297.2	297.0	297.8
标准差	12.6	12.6	12.3	12.6	13.7	12.5	12.9	12.7	12.0	10.9
含灯基频均值	297.3	297.7	297.5	297.4	297.0	297.6	296.7	296.7	296.5	296.7
标准差	8.2	8.1	8.6	7.7	7.9	7.4	6.8	8.1	7.6	7.3
含灯基频均值	296.3	296.7	296.9	297.5	296.8	296.1	296.7	296.0	295.7	296.3
标准差	7.0	7.1	6.9	7.7	7.6	7.9	7.3	7.9	6.9	6.9
含灯基频均值	296.1	295.7	295.4	295.2	296.6	295.5	295.8	296.2	296.1	295.7
标准差	6.7	8.5	6.6	7.1	7.3	7.3	6.5	7.1	6.7	7.3

将上表中基频的数据以散点图误差线的方式来呈现可以比较直观的看到含灯和正常时的差异，如下图所示：

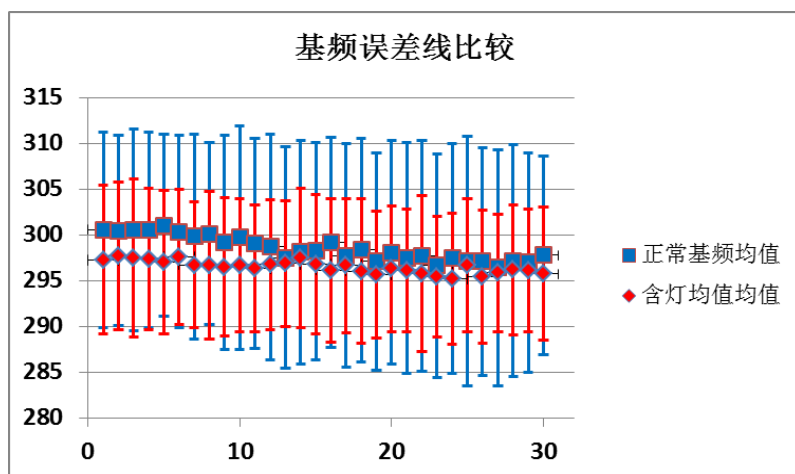


图 5.3 /a/、/i/、/u/、/y/四个单元音正常和含灯状态下的基频散点图误差线比较

从上图可以观察到，正常发音情况下的基频均值总体高于含灯时发音的基频均值。含灯时的误差线要比正常时的短，这说明含灯时的基频数据相对更集中，发音更为稳定。误差线上限和下限都出现在正常状态下。

/a/、/i/、/u/、/y/四个元音正常和含灯时的开商均值和标准差数据如下表所示：

表 5.3 单元音/a/、/i/、/u/、/y/在正常和含灯时的开商均值和标准差

正常开商均值	43.3	43.5	42.9	44.0	43.7	42.1	43.0	43.0	43.5	42.4
标准差	3.8	4.5	4.7	5.8	5.1	4.2	4.5	4.8	4.8	4.8
正常开商均值	43.3	42.5	42.8	41.5	41.6	41.6	42.2	41.6	41.9	41.6
标准差	5.2	4.4	4.2	3.7	3.4	4.2	4.0	2.7	4.4	3.9
正常开商均值	41.6	41.6	41.2	41.6	41.2	41.2	41.1	41.6	41.8	42.4
标准差	3.5	3.9	2.8	3.3	3.2	3.2	3.7	3.3	3.3	3.4
含灯开商均值	44.4	43.8	43.8	43.9	44.3	44.6	44.1	44.2	43.7	43.5
标准差	4.0	5.1	4.4	4.0	4.4	4.6	4.6	4.2	3.8	4.2
含灯开商均值	43.1	43.0	43.0	42.6	42.6	42.7	42.8	43.6	42.7	42.5

标准差	4.1	3.5	3.8	2.8	3.5	3.7	3.3	4.7	2.9	3.0
含灯开商均值	42.4	42.9	42.1	42.2	42.6	42.7	43.4	43.2	43.1	43.1
标准差	3.8	4.2	2.7	3.5	3.0	2.6	3.8	3.6	4.0	3.3

将上表中的开商的数据以散点图误差线的方式呈现出来，如下图所示：

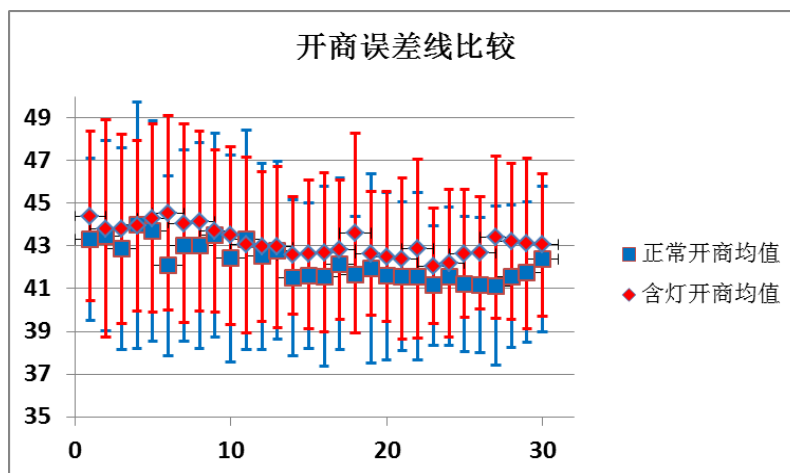


图 5.4 单元音/a/、/i/、/u/、/y/正常和含灯状态下的开商散点图误差线比较

上图为含灯和正常状态下/a/、/i/、/u/、/y/四个单元音的开商散点图误差线比较图。通过观察上图可以看出，含灯时的开商总体高于正常时，两种情况下的误差值范围差别不大，但误差线的下限基本都出现在正常状态下，上限基本都出现在含灯状态下。开商的大小和声带送气的多少有关，上图误差线的情况反应了含灯时的送气总体比正常时多，同时含灯和正常发声时声带的送气量的大小都保持在一个平稳状态。

/a/、/i/、/u/、/y/四个元音正常和含灯时的速度商均值和标准差数据如下表所示：

表 5.4 单元音/a/、/i/、/u/、/y/正常和含灯时速度商均值和标准差

正常速度商均值	560.7	565.2	557.3	533.7	539.4	561.7	537.7	539.3	506.2	545.8
标准差	99.2	102.4	94.4	102.0	87.5	93.3	88.8	82.1	95.4	113.4
正常速度商均值	548.0	502.9	545.7	541.3	520.0	522.2	546.4	535.8	538.7	540.8
标准差	121.6	82.5	87.3	95.1	90.5	101.7	103.0	86.0	62.6	82.1
正常速度商均值	543.5	508.0	535.8	546.5	516.4	525.3	523.1	512.8	486.7	499.7
标准差	90.6	81.3	85.6	102.9	85.6	70.1	94.4	123.4	120.4	157.4
含灯速度商均值	491.7	474.9	475.5	463.9	504.0	516.1	496.3	519.7	528.7	518.8
标准差	149.4	177.8	131.8	159.4	124.4	109.1	87.6	87.5	84.0	87.1
含灯速度商均值	510.7	508.1	520.7	538.2	553.7	527.5	559.1	546.0	554.8	535.5
标准差	72.7	87.4	118.8	116.4	79.0	77.7	86.5	66.5	86.5	75.1
含灯速度商均值	559.9	548.7	547.4	575.4	566.1	569.8	543.6	539.0	526.0	546.9
标准差	82.8	89.9	86.5	103.3	84.9	81.1	137.6	136.6	101.4	116.7

将上表中的速度商的数据以散点图误差线的方式呈现出来，如下图所示：

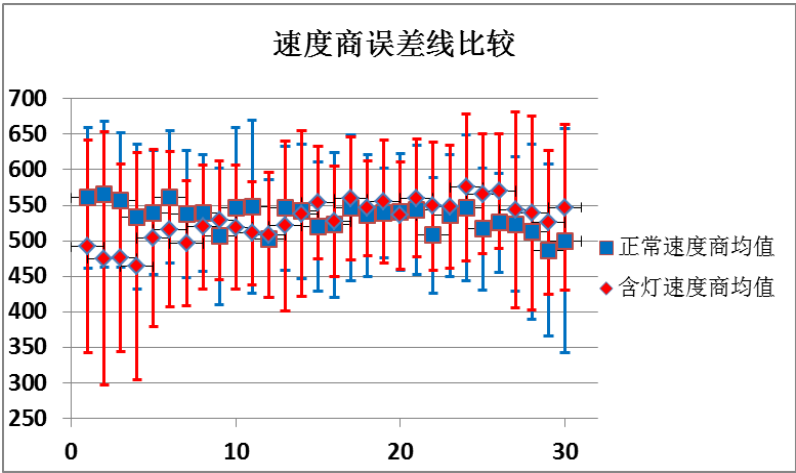


图 5.5 单元音/a/、/i/、/u/、/y/正常和含灯状态下的速度商散点图误差线比较

上图为含灯和正常状态下/a/、/i/、/u/、/y/四个单元音的速度商的散点图误差线比较图。从图中可以观察到，含灯和正常状态下的速度商均值分布相近，两种状态下的误差线总体来看含灯时比正常时略长，误差线的频率上限和频率下限都是出现在含灯时。这说明含灯时的速度商波动范围较正常时略大。速度商和发声时声带的松紧有关，上图误差线的情况反应了含灯时和正常时嗓音松紧并没有显著差异，但含灯时声带的松紧变化范围更大，声带松紧状态与正常时相比更不稳定。

5.3 不同韵母音节的基频、开商、速度商分析

我们前期录制了 400 多个音节含灯和正常情况下的语音和嗓音信号。这 400 多个音节包含了汉语普通话中所出现的所有音节，在录制时这些音节一般取阴平调，没有阴平调的音节取其他调类。我们选取了部分音节来观察不同韵母情况下含灯与正常时的嗓音参数的差异。在选取音节时一方面考虑了尽量包含所有韵母，另一方面考虑了我们所录制的信号是否可用，选择噪音干扰较小的信号进行分析。其中大部分音节为阴平调。

下面是我们分析时所用到的所有音节：

表 5.5 不同韵母音节分析时用到的音节

guang	gong	kua	sai	gou	gan	sen	si	er
sui	suan	sang	seng	guo	kuai	hei	bao	jia
jie	jue	biao	jiu	jian	jìn	jiang	jìng	xiong
xuan	xun	ueng	sun	da	jì	ju	ge	bo
gu	shi							

我们采用了和上节提取单元音嗓音数据的相同的方法，通过 Kay 公司的 Multi-Speech Model 3700 软件提取了以上音节的嗓音参数。由于每个音发音时长有差异，我们得到每个音的参数后通过 Matlab 将它们归一化为 30 个点。得到数据后使用 Excel 画出了正常和含灯时同一音节的对比图。

5.3.1 不同韵母音节的基频

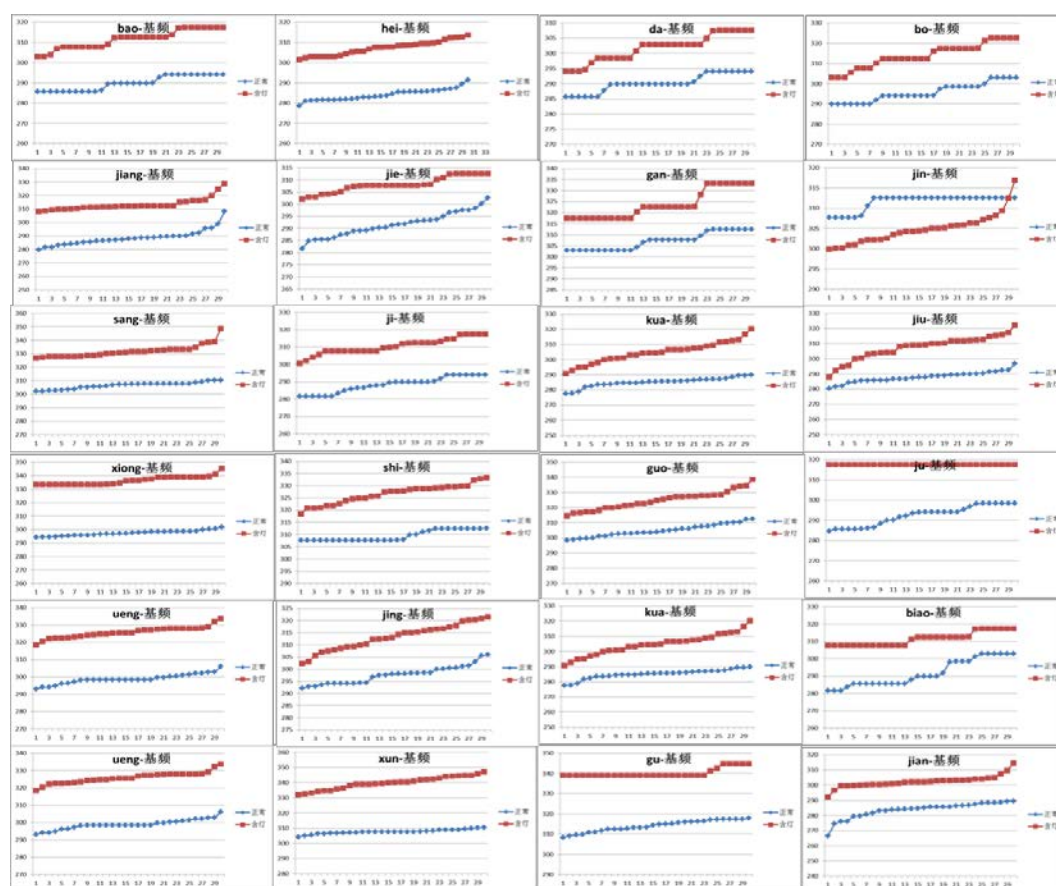


图 5.6 部分音节正常和含灯时基频比较示意图 A

上面列出了正常和含灯基频对比时差异较大的 24 个音的情况，这也是出现最多的情况，即含灯时的基频远高于正常时。在上图中，蓝色代表正常时的基频曲线，红色代表含灯时的基频曲线，可以观察到，以上各图中在走势一致的情况下，含灯时的基频曲线显著高于正常时。这与之之前单元音中正常时的基频大于含灯时明显不同。这些音节包括：bao, hei, da, bo, jiang, jie, gan, jin, sang, ji, kua, jiu, xiong, shi, guo, ju, ueng, jing, kua, biao, xun, gu, jian。

上图出现的是大部分音节所表现出来的情况，但也有部分音在正常和含灯时基频的差距不像上图那么大，并且有的出现了交叉现象，如下图所示：

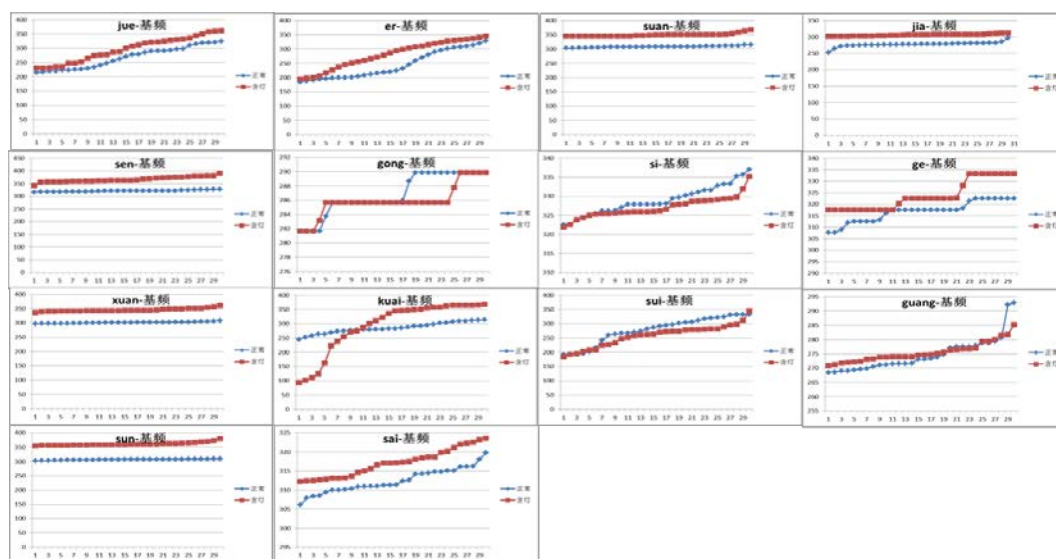


图 5.7 部分音节正常和含灯时基频比较示图 B

在上图中，蓝色代表正常时的基频曲线，红色代表含灯时的基频曲线。从上图可以发现，在这部分音节中含灯和正常的基频值差异没有图 5.6 中那么大，并且像 gong、kuai、guang 这样的音节的基频值出现了交叉现象。但从大的趋势上来看，此部分的音节仍大致可以保持含灯时基频高于正常时。上图中包括的音节为：jue, er, suan, jia, sen, gong, si, ge, xuan, kuai, sui, guang, sun, sai。

5.3.2 不同韵母音节的开商

正常和含灯时开商的分布比较有规律，大致分为以下几种情况：

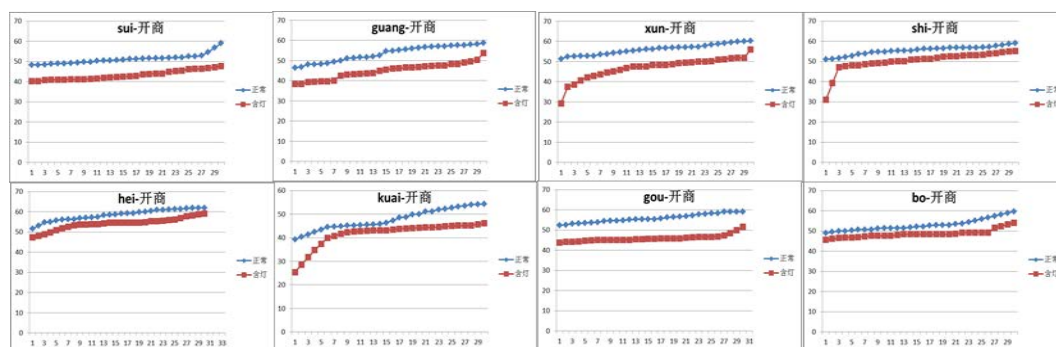


图 5.8 部分音节正常和含灯时开商比较示图 A

1) 上图为正常和含灯时开商曲线大致平行的几个音节。蓝色代表正常时的开商曲线，红色代表含灯时的开商曲线。总体来看，正常时的开商值大于含灯时，两者差距不大，但在整个发音过程中两者都保持了持续上升的状态，即正常和含灯发声时声带漏气程度都在不断增加。上图包括的音节有：sui, guang, xun, shi, hei, kuai, gou, bo。

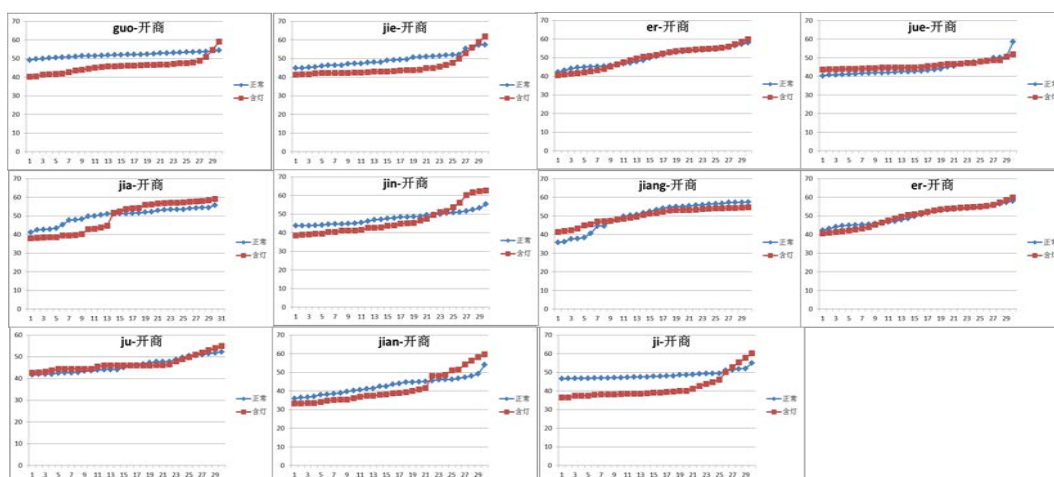


图 5.9 部分音节正常和含灯时开商比较示图 B

2) 上图为正常和含灯时开商数据出现交叉的几个音节。蓝色代表正常时的开商曲线，红色代表含灯时的开商曲线。两者开商在都保持上升趋势的情况下，在发音结束前开商数据出现了交叉。出现较多的情况是前期正常时的开商大，后期含灯的开商大。这说明在前期正常时的声带漏气大于含灯时，而后期含灯的声带漏气有了较大程度的增加，进而两种情况的开商在曲线上发生了交叉。

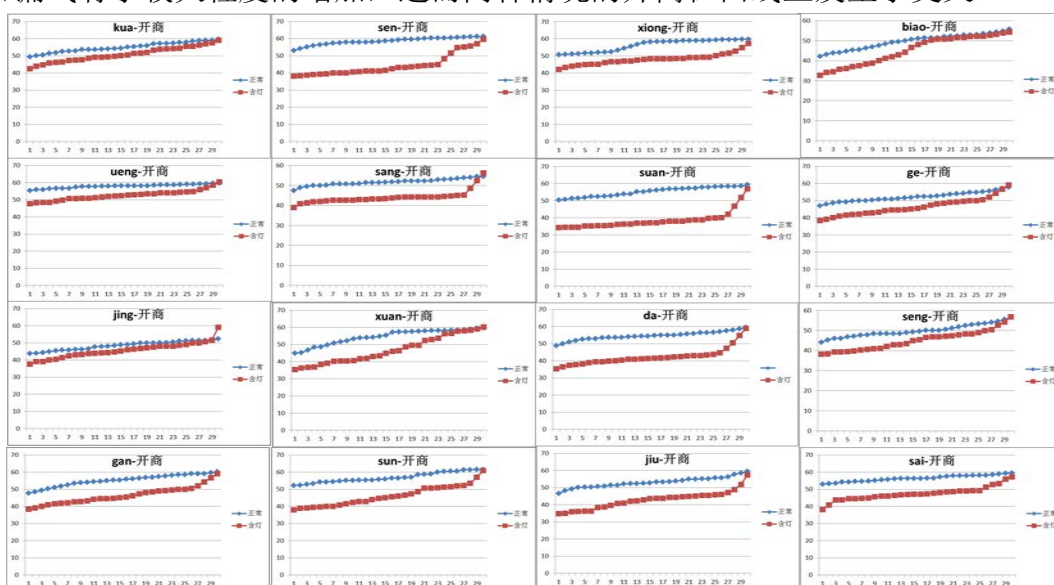


图 5.10 部分音节正常和含灯时开商比较示图 C

3) 第三种情况出现的比较多，如上图所示。蓝色代表正常时的开商曲线，红色代表含灯时的开商曲线。这类情况的整体趋势是前期正常的开商大于含灯的开商，在音节发音末尾时含灯的开商突然增大，出现了正常开商和含灯开商大小相同的情况。这说明这些音节发音时整体都是正常时的声带漏气程度大于含灯时的声带漏气程度，而在音节发音结束前很短的时间内含灯的声带送气突然增加，

达到和正常时相同的水平。上图包括的音节有：kua, sen, xiong, biao, ueng, sang, suan, ge, jing, xuan, da, seng, gan, sun, jiu, sai。

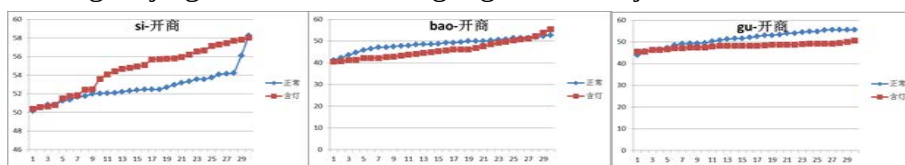


图 5.11 部分音节正常和含灯时开商比较示意图 D

4) 最后一类情况是在音节前端正常和含灯的开商相同，发音过程中两者开商差距增大，最后开商值出现靠拢倾向。在这类情况中，有含灯的开商整体大于正常的情况，也有正常的开商整体大于含灯的情况。此类音节，发声开始的时候正常和含灯声带漏气程度都是一样的，之后两者发生差异，其中一方漏气程度增大，最后两者漏气程度又达到一致。

5.3.3 不同韵母音节的速度商

下面是部分音节正常和含灯时速度商的对比情况。

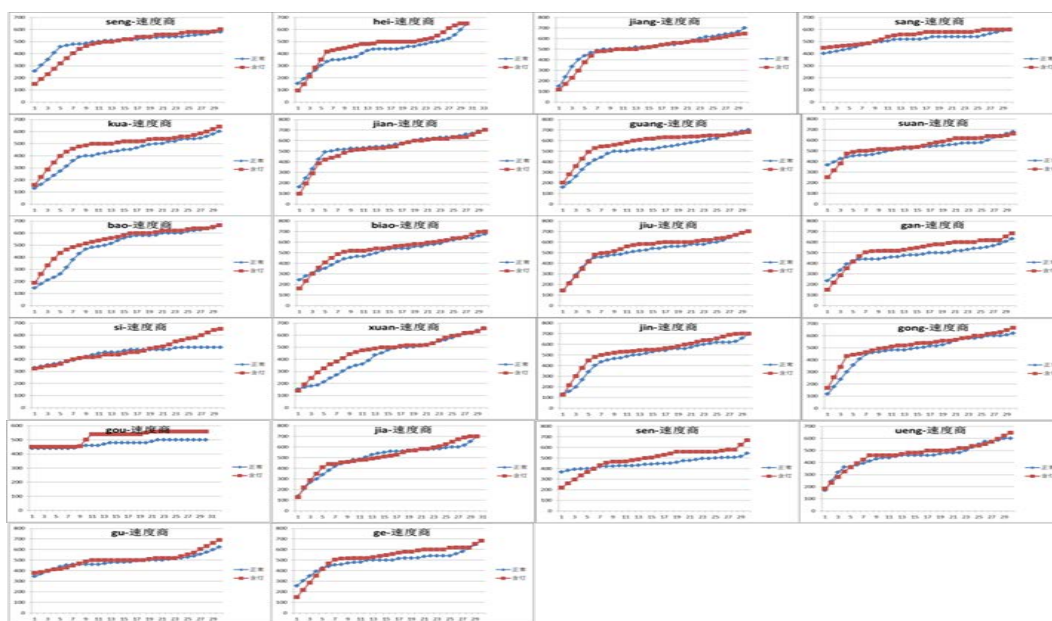


图 5.12 部分音节正常和含灯时速度商比较示意图 A

1) 这些音节正常和含灯时速度商的比较是很有规律的。上图显示的是大部分音节出现的情况。蓝色代表正常时的速度商曲线，红色代表含灯时的速度商曲线。从整体来看，含灯（红线）和正常（蓝线）发声时的速度商都呈上升趋势，同时含灯时的速度商基本大于正常时的速度商。在上图的这些音节中，含灯和正常时的速度商起点相近，终点也没有太大差异，但发声过程中含灯时的速度商整体大于正常时。这说明正常和含灯发声开始时声带松紧相当，但发声过程中含灯

状态的声带相对更紧一些，在音节末尾时两者的声带松紧又趋于一致。

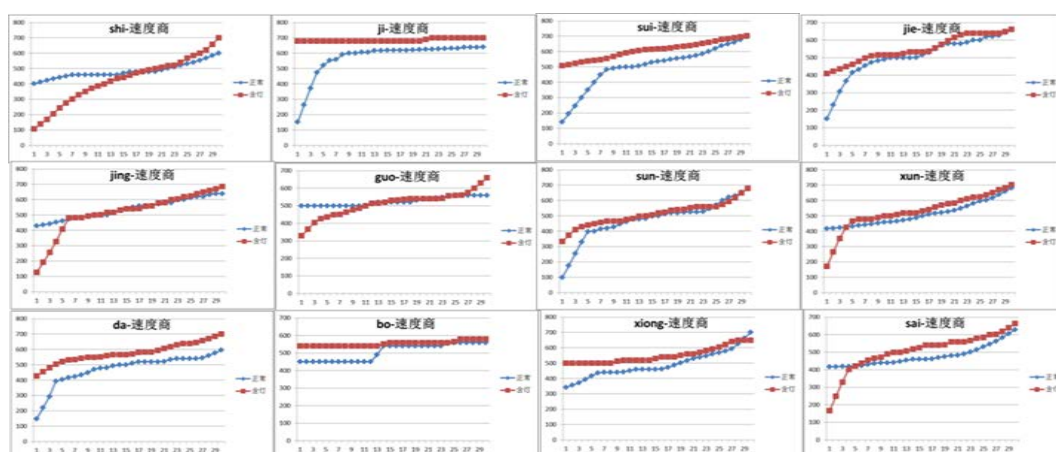


图 5.13 部分音节正常和含灯时速度商比较示图 B

2) 图 5.13 是第二类速度商对比模式，蓝色代表正常时的速度商曲线，红色代表含灯时的速度商曲线。在这几个音节中，正常和含灯时的速度商在音节开始时相差较大，通常是含灯时的速度商高于正常时，之后两者有趋近趋势。这说明在发声开始的时候含灯的声带要远比正常时更紧，之后二者声带松紧差异减小。

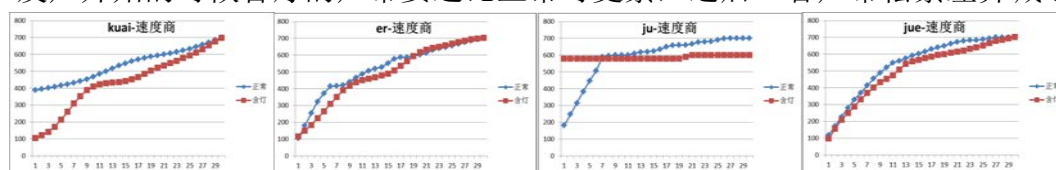


图 5.14 部分音节正常和含灯时速度商比较示图 C

3) 我们发现虽然整体来说含灯时的速度商比正常时要大，但也有正常时的速度商大于含灯时速度商的情况，如图 5.14 中的 4 个音节。蓝色代表正常时的速度商曲线，红色代表含灯时的速度商曲线。在这里可以发现，有的情况是正常发声时声带一开始就比含灯时紧，之后两者声带松紧趋近，最后趋于一致。也就是说，也存在正常时声带比含灯时声带更紧的情况。

本节小结：通过对部分音节正常和含灯时速度商比较我们发现，基本规律是含灯时的速度商大于正常时的速度商，很多情况下两者差异不大，也有部分情况在发声开始时正常和含灯的速度商差距较大，之后两者速度商趋近。即，总体来讲含灯时的声带要比正常时紧，有部分情况是在发声初期含灯时的声带明显比正常时紧，之后两者差异减小，大部分情况音节末尾时正常和含灯声带的松紧程度一致。

5.4 语篇和曲目的基频、开商、速度商

5.4.1 语篇和曲目的基频箱式图比较

我们通过 Kay 公司的 Real Time EGG Analysis Model 5138 提取了 400 多音节、语篇《北风和太阳的故事》和 7 个曲目的基频、开商、速度商数据，为了更直观的观察正常和含灯状态下各参数的分布情况，我们将同类参数导入 SPSS 画出了箱式图。

箱式图主要利用了数据中的五个统计量：最小值、第一个四分位数、中位数、第三个四分位数、最大值，以此来描述数据的分布情况。从箱式图可以粗略的观察到数据是否对称以及数据分布的离散程度。

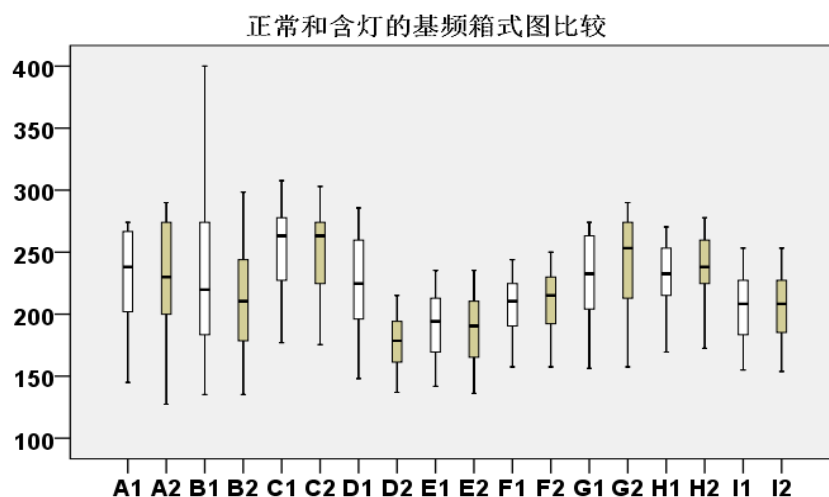


图 5.15 正常和含灯的基频箱式图比较（A-I 分别代表不同语料，相同字母代表同一语料，1 代表正常时的数据，2 代表含灯时的数据。A 语料为 400 多音节，B 为语篇《北风和太阳的故事》，C 到 I 代表不同曲目。）

上图为 400 多音节、语篇《北风和太阳的故事》以及 7 个曲目的基频参数的箱式图分布情况。图中空白箱代表正常时的基频值，灰色箱代表含灯时的基频值。通过观察基频的总体分布区域可以发现，这几组语料的基频值都集中在 150Hz 到 300Hz 之间。A、B、C、D、E 组正常时的基频中位数要高于含灯时，F、I 组正常和含灯时的基频中位数相近，G、H 组含灯时基频的中位数要高于正常时。从箱体四分位点所处的位置来看，基频在较高频率域上更为集中，其中可以观察到 C 组箱图的中位数也集中在箱体顶部，这说明数据在所占区域的上部更为集中。从箱体长短来看，除了 A、B、D 两组有明显差异外，其他几组箱体长短都较相近，也就是说它们的数据主体分布范围没有太大差异。

另外值得注意的是，A、B 两组为朗读的基频值，而其他组则是演唱时的基

频值。从箱形图可以发现，朗读时的箱体普遍比演唱时的箱体要长。也就是说演唱时基频波动的区间更小，发音更稳定，而朗读时基频波动区间较大。

5.4.2 语篇和曲目的开商箱式图比较

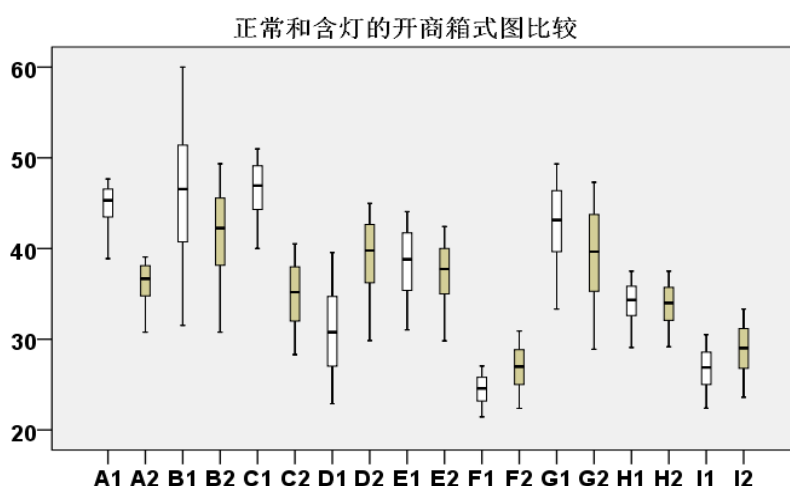


图 5.16 正常和含灯的开商箱式图比较（A-I 分别代表不同语料，相同字母代表同一语料，1 代表正常时的数值，2 代表含灯时的数值。A 语料为 400 多音节，B 为语篇《北风和太阳的故事》，C 到 I 代表不同曲目。）

上图为400多音节、语篇《北风和太阳的故事》以及7个曲目开商的箱式图分布情况。空白的箱体代表正常时的开商值，灰色的箱体代表含灯时的开商值。从上图可以观察到，虽然开商值都集中在20到50之间，但开商在不同组别间差异较大。从箱式图的数据主体来看，正常时的开商在上下坐标上所占区域要比含灯时广，也就是说正常时的开商波动范围要大于含灯时，即含灯时声带的漏气程度比正常时更稳定。A、B、C、E、G组正常时的开商箱体分布区域明显高于含灯时，D、F、I组含灯时的开商箱体分布区域明显高于正常时，H组正常和含灯的箱体分布区域没有显著差异。与基频的箱式图类似，开商的数据主体集中在偏上部分，可以看到箱体下方的线段要长于箱体上方的线段。从数据主体可以观察到，开商的中位数基本处于箱体中央位置，即数据主体的分布较均匀。

和之前在基频的箱式图中观察到的现象类似，B组（即语篇《北风和太阳的故事》）的开商分布区域最广，并且开商最大值出现在此组的正常发音时。开商和声带漏气有关，声带漏气多时开商大。这说明朗读语篇时开商波动最大，同时正常朗读时的漏气量最大。相比而言，演唱时声带的漏气程度普遍较低，漏气量也最稳定。

5.4.3 语篇和曲目的速度商箱式图比较

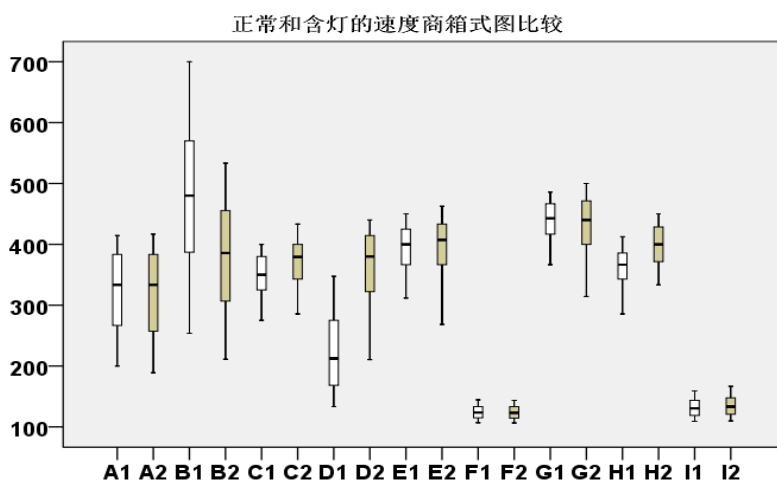


图 5.17 正常和含灯的速度商箱式图比较 (A-I 分别代表不同语料, 相同字母代表同一语料, 1 代表正常时的数值, 2 代表含灯时的数值。A 语料为 400 多音节, B 为语篇《北风和太阳的故事》, C 到 I 代表不同曲目。)

上图为400多音节、语篇《北风和太阳的故事》以及7个曲目速度商的箱式图分布情况。空白箱体代表正常时的速度商, 灰色箱体代表含灯时的速度商。与开商整体分布有较大波动不同, 无论是正常还是含灯时的速度商整体分布较为平稳, 主要集中在250到400之间。但值得注意的是F、I两组数据, 这两组数据的箱体长度很短, 主要集中在100到150之间, 这说明它们的速度商数据非常集中, 发音时嗓音的松紧程度很稳定。与开商的中位数分布类似, 速度商的中位数基本位于箱体中央位置, 这说明数据主体的分布范围较为均匀。

值得注意的是, B组(语篇《北风和太阳的故事》)数据的箱体比其他组明显更长, 这说明朗读语篇时速度商要比演唱时波动大。速度商的大小和嗓音松紧有关, 这说明朗读时嗓音的松紧变化要比演唱时更丰富。

速度商的箱式图与基频和开商的箱式图有一些不同之处, 即速度商有明显的组内差异。可以看到A、E、F、G、H、I组含灯时的速度商箱体要比正常时更长, 这说明含灯时嗓音松紧变化的范围比正常时更大。

5.5 本章小结

本节分析了单元音、语篇和曲目的基频、开商、速度商这三个嗓音数据, 分别通过误差线、箱式图的方法来进行比较直观的呈现。

通过分析发现, /a/、/i/、/u/、/y/四个单元音的基频均值在正常条件下都要比

含灯时更大。在开商方面，高元音含灯时的均值要高于正常发音时的均值，其他元音则是正常时的均值高于含灯时的均值。速度商方面，前元音正常时的均值要大于含灯时的均值，其他元音则是含灯时的均值大于正常时的均值。

与单元音正常和含灯的基频比较结果不同，通过观察多个音节的嗓音参数发现，含灯时的基频普遍明显高于正常时，这可能是由于发单元音时口腔并没有明显成阻，而增加了辅音之后的音节在发音时口腔形成了明显的阻碍，所以声门下压增大进而导致了基频的升高造成的。多个音节的开商的基本趋势是，正常时的开商稍大于含灯时，速度商方面是含灯时的速度商稍大于正常时的速度商。

通过单元音误差线的比较可以发现，基频方面，含灯时的误差线要比正常时更短，即含灯时基频数据相对更集中，发音更稳定；开商方面，含灯时和正常时的误差线长度差别不大，但含灯的开商均值总体高于正常时，这说明含灯时声带的漏气程度总体大于正常时，同时两者发声时的漏气程度都保持在平稳状态。速度商方面，含灯时和正常时的速度商均值分布并没有显著差异，但含灯时的误差线更长，即含灯时声带的松紧变化范围更大，声带松紧状态与正常时相比更不稳定。

通过400多音节、语篇和曲目的基频箱式图比较发现，大部分情况是正常时的基频中位数要高于含灯时。从四分位点来看，基频在较高频率域上更为集中，含灯和正常的基频数据主体分布范围没有太大差异。朗读时的箱体普遍比演唱时的箱体长，这说明演唱时基频波动区间小，发音更为稳定。

从开商的箱式图来看，不同语料组别间的差异较大。但整体来说，正常时开商数据的分布范围要比含灯时广，这说明正常时开商的波动范围要比含灯时大，即含灯时声带的漏气量比正常时更稳定。朗读语篇时开商的分布范围最广，这说明此时声带的漏气量的大小最不稳定。相比而言，演唱时声带的漏气程度普遍较低，漏气量也最稳定。

从速度商的箱式图来看，正常和含灯时的速度商整体分布都较为平稳，这说明嗓音的松紧程度都较稳定。正常和含灯时的中位数基本位于箱体的中央位置，这说明数据整体的分布较为均匀。语篇朗读时的箱体比其他组语料的都要长，这说明朗读语篇时速度商比演唱时波动大，即嗓音松紧变化更多，而演唱时嗓音的松紧程度处于较为平稳的状态。与基频、开始的箱式图不同，速度商的箱式图有明显的组内差异，大致来讲含灯时的速度商箱体要比正常时更长，这说明含灯时嗓音松紧变化要比正常时更大。

第六章 语篇和曲目的长时平均谱分析

长时平均谱又叫长时平均 FFT（fast fourier transform）功率谱，是一种常用而有效的分析连续语流或唱歌的工具。它利用声压平均值来再现频谱特性。数据提取时，通常截取 30-40s 的语音数据。在进行计算之前，需要将辅音和空白段切除。每个人的长时平均谱都具有明显而稳定的个性特征。

长时平均谱利用快速傅立叶变换计算出每帧的短时频谱，再对一定时间里的短时频谱进行平均。这类功率谱可以比较直观的反应出基频周期、各谐波能量、频率、形态等情况。其中能量较高的尖峰对应了共振峰。^[24]

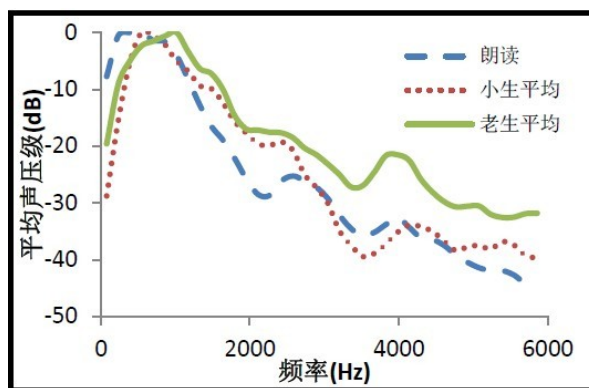


图 6.1 昆曲三种发音方式的长时平均谱（董理，2012）^[14]

古典歌手的谱曲线在 2000Hz 到 3000Hz 之间存在明显的歌唱共振峰，4000Hz 附近有演讲者共振峰。歌唱共振峰的一个副作用是常常让人听不出所唱的元音是什么。董理（2012）认为昆曲老生和小生唱法存在演讲者共振峰，而不存在歌唱共振峰。^[14]

6.1 数据提取

本文录制时使用的语篇为《北风和太阳的故事》，含灯和正常每组各录制了两遍。在此部分的参数设置为：语音采样频率为 20kHz，分析带宽为 312.5Hz，FFT 点数为 128 个，分析频率域为 0-10000Hz。

操作步骤为：

1) 先将音频文件导入 Adobe Audition，切除空白段。

2) 将音频文件导入 Wavesurfer 软件，鼠标右键单击选择 Create Pane-Spectrogram，之后右键单击选择 LTAS。此时在弹出的窗口中按照上文的参

数进行设置。之后选择 **Export** 将音频的长时平均谱数据以.txt 格式导出。最后将数据导入 Excel 程序中进行画图。

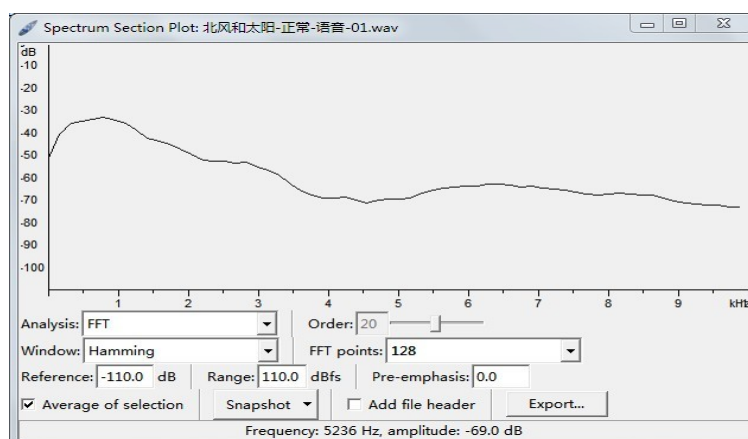


图 6.2 Wavesurfer 长时平均谱提取界面示图

6.2 语篇的长时平均谱

下图为我们得到的语篇《北风和太阳的故事》的长时平均谱分布情况。

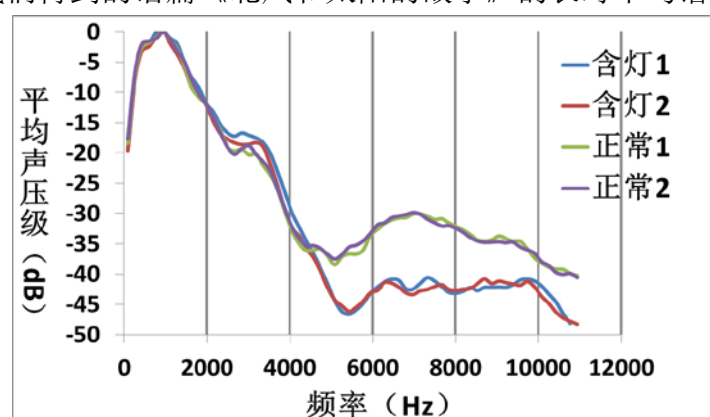


图 6.3 语篇《北风和太阳的故事》长时平均谱示图

在图 6.3 中，在高频区域较高的两条线（深蓝色和绿色）代表正常时的长时平均谱，较低的两条线（浅蓝色和红色）代表含灯时的长时平均谱。

从上图来看，含灯和正常的长时平均谱在 800Hz 附近都出现了一个频谱高峰区，即为能量最大区。在 0Hz-4000Hz 之间时，含灯与正常的频谱曲线大致相同，在 4000Hz-10000Hz 之间时，正常时的声压级持续高于含灯时，其中在 4700Hz 到 8500Hz 之间基本高出 10dB 左右。

具体来说，含灯与正常时的频谱曲线在 2500Hz 到 3500Hz 之间都出现了明显拐点。但含灯的谱曲线在此处起伏程度不大，整体较为平直。而正常的谱曲线在此时出现了较明显的小波峰。在此频率段内，含灯的谱曲线一直高于正常状态。

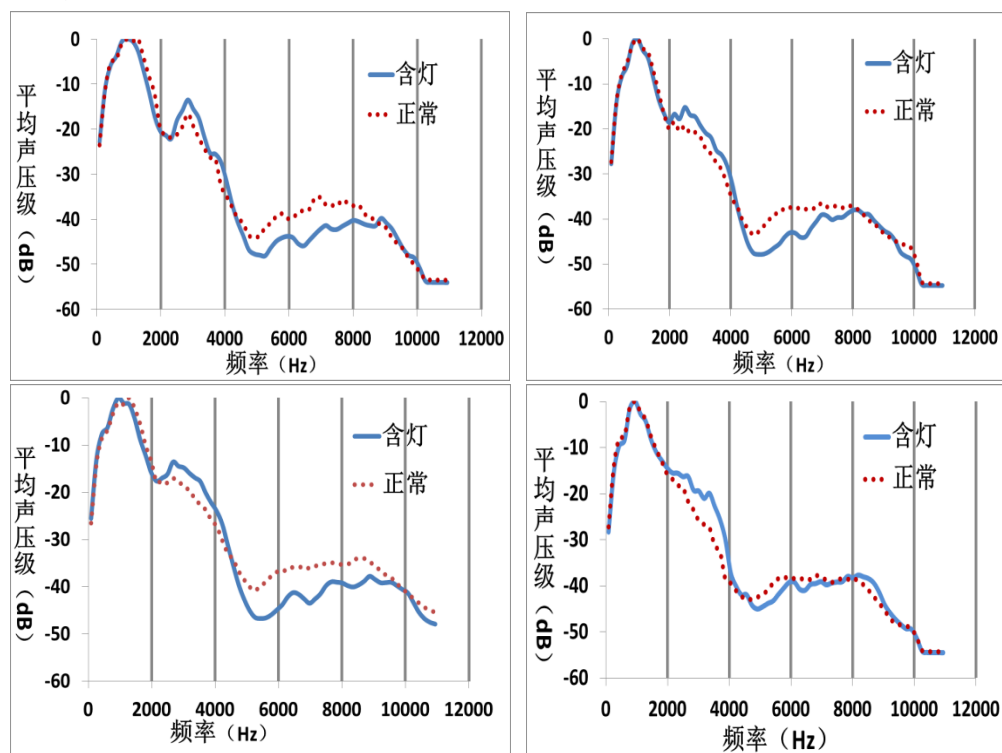
含灯与正常时的频谱在 4000Hz 后开始产生较大区别。其中，正常情况的频谱曲线下降速度开始减慢，之后在 4700Hz 附近到达最低点。含灯时的谱斜率在 4000Hz 之后仍保持较大的绝对值，当谱曲线下降至 4700Hz 时到达最低点，即 4000Hz 后含灯时的能量随着频率的提高迅速下降，至 4700Hz 左右时又快速提高。

4000Hz 之后，含灯时的声压值持续低于正常时。在 4700Hz 之后，两者在大致保持 10dB 稳定差距的情况下又都开始上升，在 6000Hz 左右两者都出现了一个小的峰值。含灯时的谱斜率在 6000Hz 之后一直处于平稳状态。在 8500Hz 左右时，两者的频谱曲线出现了明显下降。正常时的频谱曲线在 10000Hz 时达到比 4700Hz 附近出现的值还小的最低值，含灯时的频谱曲线在 10000Hz 时达到接近 4700Hz 时出现的最低值。

从上图可以观察到，虽然在 2000Hz-3000Hz 之间并未形成大的歌唱共振峰，但此处的谱曲线也在一定程度上发生了改变。

6.3 曲目的长时平均谱

通过对比我们发现，曲目和语篇的长时平均谱存在很大差异。下面是曲目的长时平均谱示图。



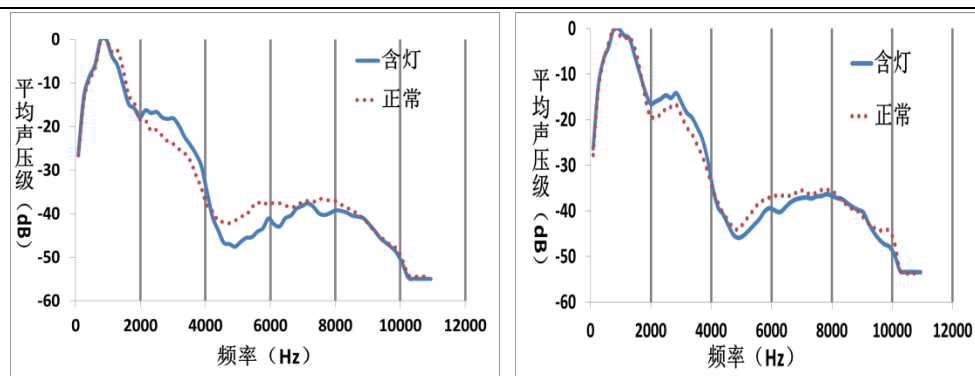


图 6.4 曲目长时平均谱示图（上面几个图从左到右从上到下的顺序依次为《黛玉悲秋》、《梅花的故事》、《两头忙》、《前门情思大碗茶》、《四世同堂》、《一声祝福》。）

观察上图中曲目的长时平均谱可以发现，和语篇谱曲线的走势一致，曲目的长时平均谱在 2000Hz-4000Hz 的低频区域时含灯的谱曲线要高于正常的谱曲线，在 4500Hz-8500Hz 之间的高频区域时正常的谱曲线高于含灯的谱曲线。

具体来看，曲目正常和含灯时的谱曲线在 800Hz 附近都出现了一个频谱高峰区，即能量最大区。和语篇的长时平均谱不同，在 2000Hz-4000Hz 之间时，曲目正常和含灯的谱曲线有较为明显的差异，含灯的谱曲线整体位置明显高于正常时。其中，正常时的谱曲线下降较直接，像《梅花的故事》、《前门情思大碗茶》、《四世同堂》中正常的谱曲线在 2000Hz 之后并没有一个上升的过程，而是直接下降。而含灯时的谱曲线在此区域一般先升后降。

在 4000Hz-8500Hz 左右的高频区间内，几个曲目正常时的谱曲线虽然也和语篇中一样高于含灯时的谱曲线，但两者之间的距离都没有语篇中那么大，尤其在《前门情思大碗茶》和《一声祝福》中，正常时的谱曲线与含灯时的谱曲线相差很小。

对比了语篇和曲目的长时平均谱之后还可以发现，语篇中含灯和正常频谱的谱曲线在高频区域一直到 11000Hz 之后仍然保持有较大距离，而曲目中两者在 8500Hz 之后就几乎相同。这说明朗读和演唱时的语音在高频区域会有很大不同，朗读时在 8500Hz 之后正常语音和含灯语音仍然存在很大差异，而演唱中正常语音和含灯语音在 8500Hz 之后差异几乎消失。即朗读时含灯与否带来的对高频能量带来的影响更大。

另外观察上面几个图可以发现，含灯时的长时平均谱在 5000Hz-9000Hz 之间有几个较大的波动，而正常时的长时平均谱在此区间较为平稳，即使有波动也不大。

目前对 5000Hz 以上的语音频谱研究较少，但它们对歌唱和言语的音质感知可能会有影响。在进行将高频能量加入到语音合成和声道模型这样复杂的工作之前，我们有必要先研究高频能量的细微变化是否对人的感知产生影响。高频能量

可以提供稳定的个人音色差异方面的信息。Sten（2011）改变 8000Hz 水平的八度音阶的听辨实验，表明多余 90% 的被试可以感知到高频能量的改变。^[31]

从图 6.3、图 6.4 可以看到，含灯后的频谱曲线在高频区域与正常时有明显不同，这是否会造成听众对语音感知的不同？我们认为可以对此进行感知实验。

6.4 本章小结

本节通过 Wavesurfer 软件提取了语篇和曲目的长时平均谱，观察分析了同一发音人在正常和含灯情况下的谱差异。结果发现：1）语篇的长时平均谱正常和含灯的情况下在 4000Hz-11000Hz 之间有明显差异，正常时的平均谱要高于含灯时约 10dB。目前学界对 5000Hz 以上的语音频谱研究较少，但目前有研究表明人们可以感知高频能量的改变。^[29]2）曲目的长时平均谱也表现出了和语篇类似的情况，即在高频区间正常时的谱曲线要高于含灯时。但这个区间在曲目中大概保持在 5000Hz-9000Hz，9000Hz 以上含灯和正常的长时平均谱的谱曲线又趋向并拢。3）曲目中含灯和正常的长时平均谱在高频的差距没有语篇中大。4）另外，曲目的长时平均谱在 2000Hz-4000Hz 之间存在较为清晰的差异，含灯时的谱曲线要高于正常时，这和语篇中的情况也是不同的。

本节的分析结果可以反应正常和含灯情况下无论语篇还是曲目的长时平均谱谱曲线都存在差异，这为进一步做听辨实验提供了基础。

第七章 总结与展望

本文从共振峰、长时共振峰、基频、开商、速度商、长时平均功率谱等方面对含灯和正常的语音、嗓音进行了对比分析，从以上参数可以发现含灯和正常存在的一些差异。

由声学元音图在含灯后的改变可知，含灯对舌位产生了很大影响，其中高元音/i/受到的影响最大，其F2在含灯后发生了明显下降。从语篇的长时共振峰来看，可以明显的发现含灯后F2的整体分布范围比正常时更高，这可能和含灯后唇的左右开度增大、口腔前腔变小、圆唇程度降低有关。此处仍存在问题需要解决：单元音分析中含灯时的F2比正常时要低，而语篇的长时共振峰则是含灯时F2比正常时更高。这可能和语篇中受辅音音轨影响、存在协同发音、存在多个韵母有关，两者关系较为复杂，并非简单的正相关关系，需要做进一步分析才能得出结论。

单元音/a/、/i/、/u/、/y/在基频方面含灯时比正常时小，开商方面高元音含灯时高于正常时，速度商方面前元音正常时大于含灯时。含灯时的基频比正常时更稳定，声带漏气程度总体大于正常时，嗓音的松紧变化范围比正常时更大。

400多音节、语篇、曲目中，含灯时的基频中位数要低于正常时。基频在较高频率域上更集中。演唱时的基频比朗读时更稳定。含灯时声带漏气量比正常时更稳定，演唱时声带漏气量比朗读时更稳定。正常和含灯的嗓音松紧程度变化都较平稳，而朗读时嗓音松紧的变化范围要比演唱时大，含灯时的嗓音松紧变化比正常时大。

正常和含灯的长时平均谱谱曲线在高频有明显差异，正常的谱曲线明显高于含灯时。朗读时正常和含灯谱曲线的差异比演唱时更大，并且存在差异的频率范围比演唱时更广。

本次实验得到的结果可以为语音高频的研究提供一些支持，比如含灯和正常时的长时平均谱在高频区间的差异。以往对于5000Hz以上的语音图谱研究的较少，但已有研究表明人们可以感知高频能量的变化。我们目前已经从长时平均谱的观察得知正常和含灯时在高频存在显著差异，这就为进一步做听辨实验提供了基础。

参考文献

- [1].黄强.“含灯大鼓”唱声声学特征的初步分析.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].中国曲艺志编辑部,中国曲艺志,天津卷.2001,中国 ISBN 中心出版社:北京.
- [3].中国艺术研究院音乐研究所中国音乐词典编辑部,中国音乐词典.1985,人民音乐出版社:北京.
- [4].东方音乐学会编,中国民族音乐大系,曲艺音乐卷.1989,上海音乐出版社:上海.
- [5].北京百科全书总编委员会,北京百科全书.2001,奥林匹克出版社:北京.
- [6].中国曲艺志编辑部,中国曲艺志,北京卷.2001,中国 ISBN 中心出版社:北京.
- [7].中国大百科全书出版社编辑部,中国大百科全书,戏曲曲艺卷.1983,中国大百科出版社:北京.
- [8].中国曲艺音乐集成编辑部,中国曲艺音乐集成,天津卷.1996,中国 ISBN 中心出版社:北京.
- [9].张静,梅花大鼓论略,2007,天津音乐学院.第134页.
- [10].黑杨,试析梅花大鼓的琵琶伴奏艺术,2011,中央音乐学院.第48页.
- [11].孔江平,哈尼语发声类型声学特性研究及音质概念的讨论,1996,民族语文.
- [12].吴宗济等,实验语音学概要.1989,高等教育出版社:河北.
- [13].孔江平,论语言发声.2001,北京:中央民族出版社.
- [14].董理,昆曲老生和小生的语音和嗓音声学初探,语音乐律报告2012,北京大学语音乐律实验室.
- [15].郑茂平,声学语音学.2007,上海:上海音乐出版社.
- [16].曹洪林,刘建伟,正常、大声两种说话状态下语音的声学差异及对声纹鉴定的影响,第二届证据理论与科学国际研讨会2009:中国北京.第12页.
- [17].杨锋,中国传统吟诵研究,2012,北京大学:北京.
- [18].吉永郁代,能剧嗓音发声研究,2012,北京大学.
- [19].王士谦,关于歌唱共振峰及其有关问题的探讨——兼评Sundberg的歌唱共振峰概念及其降喉解释.南京艺术学院学报(音乐与表演版),1986(04):第2-8+41页.
- [20].包紫薇,歌唱和音乐的声学问题——为全国高等音乐院校学生声乐比赛大会而作.音乐艺术,1981(01):第52-59+63页.
- [21].李小武,罗兰娥,BP神经网络在评价歌唱艺术嗓音中的应用.现代电子技术,2012(14):第146-148页.
- [22].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.2010(02).
- [23].陈彧,歌手共振峰Praat软件测量分析初探.天津音乐学院学报,2012(02):第87-93页.
- [24].石锋.语音格局的分析方法.第六届全国现代语音学学术会议.中国天津.

- [25].曹洪林, 孔江平, 长时共振峰分布特征在声纹鉴定中的应用. 中国司法鉴定, 2013(01): 第62-67页.
- [26].王英利, 李敬阳, 曹洪林, 声纹鉴定技术综述. 警察技术, 2012(04): 第54-56页.
- [27].庄琳, 利用长时平均FFT功率谱进行话者识别. 山西警官高等专科学校学报, 2011(01): 第80-82页.
- [28].Baken, R.O.R.F., Clinical Measurement of Speaking and Voice. 2000, New York: Delmar learning.
- [29].Brian B. Monson, A.J.S.T., Detection of high-frequency energy changes in sustained vowels produced by singers, Journal of the Acoustical Society of America. 2011.
- [30].Delattre, P., The Physiological Interpretation of Sound Spectrogram, in Publication of the Modern Language Association of America. 1951. p. 66(5).
- [31].McDougall, K., Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies. International Journal of Speech Language and the Law, 2006. 13(1): p. 89 - 126.
- [32].Nolan, F., .G.C., A case for formant analysis in forensic speaker identification. International Journal of Speech Language and the Law, 2005. 12(2): p. 143 - 173.
- [33].Sundberg, J., Level and Center Frequency of the Singers Formant. Journal of Voice - Official Journal of the Voice Foundation, 2001. 15(2): p. 176 - 186.
- [34].Sundberg, J., The acoustics of the singing voice, Scientific American. 1977.

附录 A 400 单音节

图 A.1 单音节（带调 400 多音节，有些音节没有代表字，只需要发出音节即可）

a ⁻	guang ⁻	o ⁻	gong ⁻	e ⁻	ka ⁻	ai ⁻	ke ⁻	ei ⁻	kai ⁻
阿	光	哦	工	饿	卡	哀	科	诶	开
ao ^ˊ	kao ^ˋ	ou ⁻	kou ^ˋ	an ⁻	kan ⁻	en ⁻	ken ^ˋ	ang ⁻	kang ⁻
熬	靠	欧	扣	安	刊	恩	肯	肮	康
eng ⁻	keng ⁻	er ^ˊ	ku ⁻	yi ⁻	kua ⁻	ya ⁻	kuo ^ˋ	ye ⁻	kuai ^ˋ
	坑	儿	哭	一	夸	压	扩	耶	快
yao ⁻	kui ⁻	you ⁻	kuan ⁻	yan ⁻	kun ⁻	yin ⁻	kuang ⁻	yang ⁻	kong ⁻
邀	亏	优	宽	烟	昆	音	筐	央	空
ying ⁻	ha ⁻	yong ⁻	he ⁻	wu ⁻	hai ⁻	wa ⁻	hei ⁻	wo ⁻	hao ^ˋ
应	哈	拥	喝	乌	嗨	挖	黑	窝	好
wai ⁻	hou ^ˊ	wei ⁻	han ^ˊ	wan ⁻	hen ^ˊ	wen ⁻	hang ^ˊ	wang ⁻	heng ⁻
歪	喉	威	含	湾	痕	温	杭	汪	亨
weng ⁻	hu ⁻	yu ⁻	hua ⁻	yue ⁻	huo ^ˋ	yuan ^ˋ	huai ^ˊ	yun ⁻	hui ⁻
翁	忽	淤	花	约	或	怨	怀	晕	灰
ba ⁻	huan ⁻	bo ⁻	hun ⁻	bai ^ˊ	huang ⁻	bei ⁻	hong ⁻	bao ⁻	ji ⁻
八	欢	播	婚	白	荒	杯	烘	包	击
ban ⁻	jia ⁻	ben ⁻	jie ⁻	bang ⁻	jiao ⁻	beng ⁻	jiu ⁻	bi ⁻	jian ⁻
搬	加	奔	街	帮	胶	绷	究	逼	尖
bie ^ˊ	jin ⁻	biao ⁻	jiang ⁻	bian ⁻	jing ⁻	bin ⁻	ju ⁻	bing ⁻	jue ^ˊ
别	金	标	江	边	京	宾	居	兵	绝
bu ^ˋ	juan ⁻	pa ⁻	jun ⁻	po ⁻	jiong ^ˋ	pai ⁻	qi ⁻	pei ^ˊ	qia ⁻
捕	捐	趴	均	坡	囧	拍	七	赔	掐
pao ^ˋ	qie ⁻	pou ⁻	qiao ⁻	pan ⁻	qiu ⁻	pen ⁻	qian ⁻	pang ^ˊ	qin ⁻
跑	切	剖	敲	潘	秋	喷	千	旁	亲
peng ⁻	qiang ⁻	pi ⁻	qing ⁻	pie ⁻	qu ⁻	piao ⁻	que ⁻	pian ⁻	quan ⁻
烹	枪	批	青	撇	屈	飘	缺	偏	圈
pin ⁻	qun ^ˊ	ping ^ˊ	qiong ^ˊ	pu ⁻	xi ⁻	ma ⁻	xia ⁻	mo ⁻	xie ⁻
拼	群	平	穷	扑	锡	妈	瞎	摸	歇
me ⁻	xiao ⁻	mai ^ˊ	xiu ⁻	mei ^ˊ	xian ⁻	mao ⁻	xin ⁻	mou ^ˋ	xiang ⁻
么	萧	埋	修	梅	先	猫	心	某	香
man ^ˊ	xing ⁻	men ^ˊ	xu ⁻	mang ^ˊ	xue ⁻	meng ^ˊ	xuan ⁻	mi ⁻	xun ⁻
蛮	星	门	需	忙	靴	盟	宣	眯	勋
mie ^ˋ	xiong ⁻	miao ⁻	zha ⁻	miu ^ˋ	zhe ⁻	mian ^ˊ	zhi ⁻	min ^ˊ	zhai ⁻
灭	熊	喵	扎	谬	哲	面	知	民	宅

灭	兄	喵	扎	缪	着	棉	之	民	斋
ming ˊ	zhao ˊ	mu ˊ	zhou ˊ	fa ˊ	zhan ˊ	fo ˊ	zhen ˊ	fei ˊ	zhang ˊ
名	招	模	州	发	粘	佛	真	非	张
fou ˋ	zheng ˊ	fan ˊ	zhu ˊ	fen ˊ	zhua ˊ	fang ˊ	zhuo ˊ	feng ˊ	zhui ˊ
否	争	翻	朱	分	抓	方	桌	峰	追
fu ˊ	zhuan ˊ	da ˊ	zhun ˋ	de ˊ	zhuang ˊ	dai ˊ	zhong ˊ	dei ˋ	zhuai ˋ
符	专	搭	准	得	装	呆	中	逮	拽
dao ˊ	zhei ˋ	dou ˊ	cha ˊ	dan ˊ	che ˊ	dang ˊ	chi ˊ	deng ˊ	chai ˊ
刀	这	兜	插	单	车	当	吃	灯	拆
di ˊ	chao ˊ	die ˊ	chou ˊ	diao ˊ	chan ˊ	diu ˊ	chen ˊ	dian ˊ	chang ˊ
低	超	爹	抽	刁	搀	丢	抻	颠	昌
ding ˊ	cheng ˊ	du ˊ	chu ˊ	duo ˊ	chuo ˊ	dui ˊ	chuai ˋ	duan ˊ	chui ˊ
丁	称	都	出	多	戳	堆	揣	端	吹
dun ˊ	chuan ˊ	dong ˊ	chun ˊ	den ˊ	chuang ˊ	ta ˊ	chong ˊ	te ˋ	chua ˊ
吨	川	东	春		窗	她	充	特	
tai ˊ	sha ˊ	tao ˊ	she ˊ	tou ˊ	shei ˊ	tan ˊ	shi ˊ	tang ˊ	shai ˊ
胎	沙	涛	奢	偷	谁	滩	诗	汤	筛
teng ˊ	shao ˊ	ti ˊ	shou ˊ	tie ˊ	shan ˊ	tiao ˊ	shen ˊ	tian ˊ	shang ˊ
疼	烧	梯	收	帖	山	挑	身	天	商
ting ˊ	sheng ˊ	tu ˊ	shu ˊ	tuo ˊ	shua ˊ	tui ˊ	shuo ˊ	tuan ˊ	shuai ˊ
听	声	突	书	托	刷	推	说	团	衰
tun ˊ	shui ˊ	tong ˊ	shuan ˊ	tei ˊ	shun ˋ	na ˊ	shuang ˊ	ne ˊ	re ˋ
吞	谁	通	栓		顺	拿	双	呢	热
nai ˋ	ri ˋ	nei ˋ	rao ˋ	nao ˊ	rou ˋ	nan ˊ	ran ˊ	nen ˋ	ren ˊ
奈	日	内	饶	孬	肉	男	然	嫩	人
nang ˊ	rang ˋ	neng ˊ	reng ˊ	ni ˊ	ru ˊ	nie ˊ	ruo ˋ	niao ˋ	rui ˋ
囊	让	能	扔	尼	如	捏	若	尿	锐
niu ˊ	ruan ˋ	nian ˊ	run ˋ	nin ˊ	rong ˊ	niang ˊ	za ˊ	ning ˊ	ze ˊ
妞	软	蔫	润	您	荣	娘	杂	宁	则
nu ˊ	zi ˊ	nuo ˊ	zai ˊ	nuan ˋ	zei ˊ	nun ˊ	zao ˊ	nong ˊ	zou ˋ
奴	资	挪	栽	暖	贼		遭	农	走
nv ˋ	zan ˊ	nue ˋ	zen ˋ	nou ˋ	zang ˊ	la ˊ	zeng ˊ	le ˊ	zu ˊ
女	咱	虐	怎	耨	脏	拉	增	了	租
lai ˊ	zuo ˊ	lei ˊ	zui ˋ	lao ˊ	zuan ˊ	lou ˋ	zun ˊ	lan ˊ	zong ˊ
来	昨	雷	最	捞	钻	搂	尊	栏	宗
lang ˊ	ca ˊ	leng ˊ	ce ˋ	li ˊ	ci ˊ	lia ˋ	cai ˊ	lie ˋ	cao ˊ
郎	擦	棱	册	梨	词	俩	猜	咧	糙
liao ˊ	cou ˊ	liu ˊ	can ˊ	lian ˊ	cen ˊ	lin ˊ	cang ˊ	liang ˊ	ceng ˊ
疗	凑	溜	参	连	岑	林	仓	凉	层
ling ˊ	cu ˊ	lu ˊ	cuo ˊ	luo ˊ	cui ˊ	luan ˊ	cuan ˊ	lun ˋ	cun ˊ
零	粗	撸	搓	罗	崔	峦	蹿	论	村

附录 A 400 单音节

long ˊ	cong ˉ	lv ˊ	sa ˉ	lue ˋ	se ˋ	ga ˉ	si ˉ	ge ˉ	sai ˉ
龙	葱	驴	仨	略	色	嘎	司	哥	腮
gai ˉ	sao ˋ	gei ˋ	sou ˉ	gao ˉ	san ˉ	gou ˉ	sen ˉ	gan ˉ	sang ˉ
该	扫	给	艘	高	三	沟	森	干	桑
gen ˉ	seng ˉ	gang ˉ	su ˉ	geng ˉ	suo ˉ	gu ˉ	sui ˊ	gua ˉ	suan ˉ
跟	僧	钢	苏	耕	缩	姑	随	瓜	酸
guo ˉ	sun ˉ	guai ˉ	song ˉ	gui ˉ		guan ˉ		Gun ˋ	rua ˉ
锅	孙	乖	松	归		官		滚	

附录 B 曲目

A. 1 前门情思大碗茶

我爷爷小的时候，常在这里玩耍，
高高的前门，仿佛挨着我的家，
一蓬衰草，几声蛐蛐儿叫，
伴随他度过了那灰色的年华。
吃一串冰糖葫芦，就算过节。
他一日那三餐，窝头咸菜么就着一口大碗茶。
啦…… 啦……
世上的饮料有千百种，也许它最廉价，
可谁知道，谁知道，
谁知道它醇厚的香味儿，饱含着泪花，
它饱含着泪花。
如今我海外归来，又见红墙碧瓦，
高高的前门，几回梦里想着它，
岁月风雨，无情任吹打，
却见它更显得那英姿挺拔。
叫一声杏仁儿豆腐，京味真美，
我带着那童心，带着思念么再来一口大碗茶。
啦…… 啦……
世上的饮料有千百种，也许它最廉价，
可为什么，为什么，
为什么它醇厚的香味，直传到天涯，它直传到天涯。

A. 2 黛玉悲秋

哎哪大观园滴溜溜起了一阵秋风，
那位林黛玉娇姿与众不同，
哎哪生成的倾国倾城人也难比，
书卷儿破读笔法儿妙灵，
她的骨格儿清高情性儿最冷，
她的文章儿博通心性儿聪明啊。

A.3 万里春光（一声祝福）

一声爆竹一首乐章，万方乐奏万里春光，
创伟业共贺新春，年年如意逢盛世，
同庆佳节岁岁吉祥。
举杯未饮心欲醉，风送花香伴酒香，
香满乾坤山河壮，春在心头豪气扬。
啊，春雷中华飞腾鹏程远，
展宏图如灿烂明霞 春日的朝阳。

A.4 重整河山待后生

千里刀光影，仇恨燃九城，
月圆之夜人不归，花香之地无和平。
一腔无声血，万缕慈母情，
为雪国耻身先去，重整河山待后生。

A.5 两头忙

高高山上两间房，一家姓李一家姓王。
李家有位大公子，王家有位大姑娘。
正月说媒二月取，三个月生了一个小儿郎。
四月会爬五月里会跑，六个月他就会叫爹娘。
七月里送到学堂把书念，八月就会啊做文章。
九月进京去赶考，十月得中状元郎。
十一月告老还乡回家转，十二月大年三十儿见阎王。
您说这小孩命有多苦，一辈子没喝过饺子汤。
您要问这叫什么段，取名就叫两头忙。

A.6 鸿雁捎书

哎那塞北沙陀迎烈风，哎我表得是出塞的昭君盼想还宫。
哎她在心中恼恨奸贼毛延寿，将哀家的美人图献与番营。
这位昭君她玉石的琵琶怀中抱起，哎说您看她眼含痛泪出了雁门关城。

A.7 梅花的故事

谁曾见说唱的人儿口含灯，字要正腔要圆请君仔细听。
且不唱雪中寒梅花开早，唱得是高洁的花高洁的人触发了高洁的爱情。
您可知道宋朝有个诗人名叫林和靖，将梅花做妻子万古传名。
这一段罗曼史虽然已经古老，这一点情感却要比现代的诗人、更多几分执着、
更多几分朦胧。

附录 C 语篇《北风和太阳的故事》

C.1 北风和太阳的故事

有一回，北风和太阳在那儿争谁的本事大。争来争去就是分不出高低来。这时候路上来了个走道儿的，他身上穿着件厚大衣。他们俩就说好了，谁能先叫这个走道儿的脱下他的厚大衣，就算谁的本事大。北风就使劲儿地刮起来了，不过他越是刮得厉害，那个走道儿的把大衣裹得越紧。后来北风没法儿了，只好就算了。过了一会儿，太阳出来了。他火辣辣地一晒，那个走道儿的马上把那件厚大衣脱下来了。这下儿北风只好承认，他们俩当中还是太阳的本事大。

附录 D Matlab 数据归一化时用到的程序①

D.1 Matlab 数据归一化时用到的程序

```
% Excel 对数据归一化
% 表单中数据格式如下： 有表头，有空格，并且每列数据长短不齐
% F1          F2          F3
F1          F2          F3
% 803.64025    1278.946842    1523.544293    748.246469
    1222.048823    1548.887193
% 795.944482    1281.714479    1533.02869    753.791609 1217.695218
    1553.419909
% 983
% 2015-04-05
clear all
close all

N = 30; % 归一化的点数

% 打开要处理的文件夹
ExcOpenDir = uigetdir('C:\', 'Open Excel folder'); % 打开文件夹
ExcFiles = dir(fullfile(ExcOpenDir, '*.xlsx')); % 获得文件夹
中所以文件
ExcLengthFiles = length(ExcFiles);
% 选择保存归一化数据的文件夹 【!!! 最好新建一个文件夹，保存归一化
数据!!!】
SaveExcOpenDir = uigetdir('C:\', 'Save Excel folder'); % 打开文件夹
for FileCount = 1:ExcLengthFiles
    fileName = strcat(ExcOpenDir, '\', ExcFiles(FileCount).name);
    [Type Sheet Format] = xlsinfo(fileName); 1
    % 查看 Excel 文件中有多少个 sheet，循环计算
```

①本程序由北京大学中文系语音音乐律实验室的姚云博士编写。

```

for activeSheet = 1: length(Sheet)
    [num txt] = xlsread(fileName,activeSheet);
    [m n] = size(num);
    Data_Norm = zeros(N,n);
    for i =1:n
        % 找到非 inf、非 NaN 数据
        ind = find(isfinite(num(:,i)));
        if isempty(ind)
            continue;
        end
        data = num(ind,i);
        % 对数据做平滑和归一化
        data = smooth(data,10,'rloess'); % 利用移动平均法 [moving]
        [sgolay] [lowess] [rloess] [loess]对 data 做平滑处理
        Len = length(data);
        Ti = linspace(1,Len,N);
        Data_Norm(:,i)=interp1(1:Len,data,Ti,'linear');
    end
    % 写入表头和左侧四列数据
    % 写入归一化数据
    % 保存 xls 文件
    saveFileName = strcat(SaveExcOpenDir,'\','归一化',num2str(N),'点',
        ',ExcFiles(FileCount).name);
    if ~isempty(txt)
        xlswrite(saveFileName,txt,activeSheet);
    end
    if ~isempty(Data_Norm)
        xlswrite(saveFileName,Data_Norm,activeSheet,'A2');
    end
end
end
end

```

附录 E 归一化后的单元音共振峰

/a/ (正常)			/i/ (正常)			/u/ (正常)		
F1	F2	F3	F1	F2	F3	F1	F2	F3
814.9	1200.4	1549.7	367.1	3009.7	3754.8	583.8	982.1	3549.1
815.6	1200.2	1549.9	371.6	3006.0	3752.2	581.7	983.9	3544.3
816.1	1200.0	1550.1	375.9	3002.6	3749.8	579.9	986.0	3539.6
816.6	1199.7	1550.2	380.3	2999.1	3747.3	578.6	987.7	3534.3
817.2	1199.3	1550.2	384.5	2996.2	3744.9	577.7	990.6	3530.1
817.0	1198.9	1550.0	388.3	2994.3	3740.0	577.0	992.8	3528.0
815.9	1198.1	1548.3	391.8	2990.9	3731.0	577.1	994.4	3526.7
814.9	1197.3	1546.5	394.5	2985.7	3719.0	578.0	994.7	3528.7
814.8	1196.7	1544.9	395.9	2979.0	3710.5	579.0	996.9	3531.3
815.7	1196.2	1543.2	396.7	2971.5	3706.2	580.5	1000.1	3533.5
817.3	1196.0	1542.1	396.7	2964.6	3705.1	582.5	1002.6	3537.4
819.6	1196.2	1541.9	396.5	2958.3	3706.9	585.7	1003.4	3541.2
822.7	1196.8	1542.1	396.4	2954.0	3705.3	589.1	1001.1	3542.0
826.0	1197.4	1542.3	396.9	2950.0	3698.3	590.1	1001.6	3540.1
829.2	1197.8	1542.4	398.4	2944.9	3688.4	590.0	1001.6	3538.5
832.0	1197.9	1542.5	400.1	2939.2	3680.1	589.6	1002.5	3538.9
834.3	1197.7	1542.3	401.3	2934.6	3676.5	590.1	1001.5	3537.8
836.0	1197.4	1541.8	401.8	2931.7	3676.4	591.5	997.6	3535.0
837.1	1197.1	1541.2	401.6	2929.7	3678.6	592.4	993.1	3532.2
837.7	1196.9	1541.3	401.2	2926.9	3678.7	592.4	990.4	3529.2
838.2	1196.6	1542.7	402.0	2923.8	3675.2	591.3	994.1	3526.0
838.6	1196.4	1544.6	403.6	2920.9	3669.0	589.8	1005.4	3522.7
838.5	1196.0	1546.5	406.1	2920.6	3664.6	588.9	1018.9	3519.5
839.1	1195.6	1547.7	409.4	2921.9	3663.3	589.0	1032.5	3515.3
840.2	1195.1	1548.6	411.9	2920.0	3660.8	589.9	1041.7	3510.4
840.7	1194.8	1549.2	413.7	2916.1	3658.5	590.9	1049.6	3508.2
841.0	1194.5	1549.7	415.1	2911.2	3655.4	591.2	1051.2	3506.8
841.4	1194.2	1550.3	416.8	2905.6	3649.3	591.4	1046.7	3505.8
841.8	1193.8	1550.8	418.6	2899.9	3643.0	591.4	1041.4	3505.1

/a/ (含灯)			/i/ (含灯)			/u/ (含灯)		
F1	F2	F3	F1	F2	F3	F1	F2	F3
788.3	1140.5	1464.6	414.9	2415.4	2864.4	572.9	890.5	2247.6
789.4	1140.4	1463.7	419.0	2417.4	2865.7	574.0	896.5	2253.8
790.1	1140.1	1462.5	423.1	2419.4	2867.0	575.1	902.3	2259.4
790.6	1139.9	1461.5	427.2	2421.2	2868.4	575.9	907.7	2264.1
789.9	1139.7	1459.4	430.8	2422.4	2870.0	575.9	911.1	2262.4
788.0	1139.4	1456.7	432.4	2424.2	2871.4	575.9	912.8	2255.1
783.9	1138.7	1454.0	433.5	2422.5	2873.3	576.0	913.2	2243.3
776.8	1138.4	1449.9	434.1	2418.6	2874.3	576.6	915.0	2234.0
767.6	1138.0	1444.6	434.1	2416.3	2873.8	577.9	918.3	2227.9
758.6	1137.7	1439.7	434.1	2418.0	2872.3	579.5	921.6	2221.6
751.0	1138.1	1437.1	434.6	2420.3	2871.0	581.3	925.1	2211.5
747.5	1138.9	1437.6	435.7	2423.0	2869.8	582.3	924.6	2200.0
747.3	1138.8	1438.8	437.3	2424.5	2869.9	581.9	920.7	2187.4
747.7	1137.9	1438.8	439.1	2426.8	2871.1	580.8	913.8	2169.8
748.7	1137.3	1438.8	441.6	2429.6	2873.2	579.7	906.7	2156.8
750.5	1136.8	1439.0	444.6	2431.5	2875.4	578.5	900.1	2148.2
751.2	1136.7	1438.7	448.0	2434.2	2877.0	577.7	896.5	2142.9
752.0	1137.4	1439.5	451.2	2438.2	2878.1	577.5	896.2	2140.1
752.6	1137.9	1441.4	453.6	2439.6	2878.5	577.6	897.8	2140.7
752.7	1137.7	1442.3	455.0	2440.3	2878.2	578.2	901.3	2146.1
753.4	1137.0	1442.8	456.0	2443.1	2878.0	578.9	904.7	2153.1
755.6	1135.9	1444.1	457.3	2448.5	2877.6	579.2	907.0	2159.9
758.4	1134.9	1445.2	458.6	2454.3	2876.6	579.4	908.3	2157.8
761.6	1134.5	1445.8	459.0	2455.6	2874.6	579.5	908.7	2152.7
763.8	1134.2	1446.7	457.7	2450.4	2871.9	580.3	909.7	2149.2
763.8	1133.7	1446.2	455.6	2443.4	2869.3	581.3	911.7	2154.3
762.8	1133.7	1444.7	453.7	2437.5	2867.3	582.3	914.7	2166.1
760.9	1133.5	1442.4	452.6	2433.9	2865.3	583.3	917.8	2180.9
758.9	1133.2	1440.2	451.9	2431.0	2863.3	584.4	920.5	2196.0

附录 F 长时平均谱数据

F.1 《北风和太阳的故事》长时平均谱

频率	含灯 1	含灯 2	正常 1	正常 2
78.1	-19.0	-18.8	-18.4	-17.7
234.4	-7.3	-7.6	-7.4	-6.7
390.6	-2.3	-2.8	-2.6	-2.1
546.9	-2.1	-2.2	-1.8	-1.5
703.1	-0.2	-0.7	-0.9	-1.1
859.4	0.0	0.0	0.0	0.0
1015.6	-1.1	-1.8	-1.3	-1.2
1171.9	-2.1	-3.8	-2.9	-3.2
1328.1	-5.4	-6.1	-5.7	-5.0
1484.4	-7.7	-9.3	-9.3	-8.1
1640.6	-9.1	-10.6	-10.8	-10.3
1796.9	-11.6	-11.9	-11.9	-11.9
1953.1	-13.2	-14.7	-14.1	-14.1
2109.4	-15.0	-16.8	-16.3	-16.8
2265.6	-16.4	-17.7	-19.0	-19.1
2421.9	-17.4	-18.4	-19.7	-20.2
2578.1	-17.0	-18.6	-19.4	-19.3
2734.4	-17.6	-18.6	-20.3	-18.8
2890.6	-17.7	-18.3	-20.2	-20.3
3046.9	-18.6	-18.9	-22.1	-21.4
3203.1	-20.3	-21.6	-23.6	-23.0
3359.4	-23.3	-25.2	-26.0	-26.0
3515.6	-26.6	-29.4	-29.7	-29.2
3671.9	-29.8	-32.3	-32.8	-32.1
3828.1	-32.1	-33.5	-34.9	-34.0
3984.4	-34.6	-35.4	-36.1	-35.3
4140.6	-36.7	-36.9	-36.0	-35.3
4296.9	-38.7	-38.8	-35.8	-35.7
4453.1	-41.0	-41.4	-37.0	-36.7
4609.4	-43.3	-42.9	-38.5	-37.5
4765.6	-44.3	-43.4	-37.1	-36.8
4921.9	-42.4	-42.5	-36.7	-35.5
5078.1	-41.6	-41.4	-36.7	-35.2
5234.4	-40.7	-40.0	-36.0	-34.5

5390.6	-39.5	-39.0	-33.6	-33.3
5546.9	-38.3	-38.7	-32.6	-31.9
5703.1	-37.6	-37.7	-31.7	-31.5
5859.4	-36.8	-37.4	-31.1	-30.7
6015.6	-37.0	-38.4	-30.7	-30.4
6171.9	-37.5	-38.3	-30.6	-30.2
6328.1	-37.3	-38.4	-30.1	-29.8
6484.4	-37.7	-38.7	-30.1	-30.1
6640.6	-37.6	-38.7	-30.4	-30.9
6796.9	-38.2	-38.9	-31.0	-31.5
6953.1	-38.7	-38.7	-30.9	-32.1
7109.4	-38.9	-38.7	-31.6	-32.0
7265.6	-39.4	-39.1	-32.1	-32.4
7421.9	-39.6	-39.8	-32.7	-33.1
7578.1	-39.6	-39.8	-33.5	-34.0
7734.4	-39.7	-39.1	-34.3	-34.5
7890.6	-39.7	-39.0	-34.6	-34.8
8046.9	-39.8	-39.4	-34.4	-34.7
8203.1	-39.7	-39.4	-33.7	-34.7
8359.4	-39.8	-39.7	-34.1	-34.8
8515.6	-39.6	-39.7	-34.6	-34.7
8671.9	-39.1	-39.6	-34.6	-35.5
8828.1	-39.0	-39.7	-35.9	-36.0
8984.4	-39.2	-40.3	-37.4	-36.6
9140.6	-40.0	-41.1	-38.2	-37.9
9296.9	-40.5	-41.6	-38.9	-38.6
9453.1	-41.5	-42.8	-39.2	-39.7
9609.4	-42.2	-43.4	-39.2	-40.0
9765.6	-42.7	-43.8	-39.9	-39.9
9921.9	-43.5	-43.8	-40.2	-40.6

F.2 《前门情思大碗茶》长时平均谱

频率	含灯	正常
86.1	-28.3	-27.2
258.4	-15.3	-14.0
430.7	-9.6	-8.0
602.9	-8.5	-7.6
775.2	-1.5	-1.2
947.5	0.0	0.0
1119.7	-2.7	-2.3
1292.0	-4.3	-4.0

1464.3	-8.3	-7.6
1636.5	-11.0	-10.5
1808.8	-12.8	-13.0
1981.1	-14.4	-15.7
2153.3	-15.5	-16.8
2325.6	-15.5	-18.2
2497.9	-16.3	-18.7
2670.1	-16.3	-21.8
2842.4	-19.2	-23.5
3014.6	-19.4	-25.9
3186.9	-21.0	-26.7
3359.2	-19.9	-27.4
3531.4	-22.8	-31.3
3703.7	-25.7	-32.7
3876.0	-30.0	-38.6
4048.2	-37.2	-39.4
4220.5	-40.1	-41.0
4392.8	-41.9	-42.3
4565.0	-42.0	-42.9
4737.3	-44.2	-42.9
4909.6	-45.0	-42.5
5081.8	-44.5	-41.7
5254.1	-43.8	-41.0
5426.4	-43.3	-39.3
5598.6	-41.8	-38.4
5770.9	-40.5	-38.1
5943.2	-39.2	-38.6
6115.4	-39.2	-38.4
6287.7	-40.8	-38.6
6460.0	-41.0	-38.3
6632.2	-39.7	-38.6
6804.5	-39.5	-37.8
6976.8	-39.1	-38.0
7149.0	-39.8	-38.8
7321.3	-39.4	-39.3
7493.6	-39.2	-38.7
7665.8	-39.1	-38.5
7838.1	-38.0	-38.6
8010.4	-38.3	-38.2
8182.6	-37.7	-39.1
8354.9	-38.1	-39.8
8527.1	-38.5	-41.0

8699.4	-39.5	-42.5
8871.7	-41.5	-43.9
9043.9	-44.2	-45.4
9216.2	-45.9	-47.1
9388.5	-47.3	-48.1
9560.7	-48.4	-48.5
9733.0	-49.3	-48.7
9905.3	-49.4	-49.3
10077.5	-51.3	-51.2
10249.8	-54.1	-54.0
10422.1	-54.5	-54.3
10594.3	-54.5	-54.3
10766.6	-54.5	-54.3
10938.9	-54.5	-54.3

F.3 《黛玉悲秋》长时平均谱

频率	含灯	正常
86.1	-22.8	-23.5
258.4	-11.2	-11.4
430.7	-5.9	-5.2
602.9	-3.9	-4.6
775.2	-0.6	-1.2
947.5	0.0	0.0
1119.7	-0.5	-0.5
1292.0	-2.8	0.0
1464.3	-7.2	-4.2
1636.5	-12.0	-8.4
1808.8	-17.3	-12.4
1981.1	-20.5	-20.0
2153.3	-21.4	-21.3
2325.6	-22.1	-21.9
2497.9	-18.1	-21.2
2670.1	-15.9	-19.5
2842.4	-13.5	-16.6
3014.6	-15.6	-19.4
3186.9	-17.6	-22.0
3359.2	-21.8	-24.3
3531.4	-25.3	-26.3
3703.7	-25.6	-26.7
3876.0	-27.3	-32.6
4048.2	-31.6	-34.9

4220.5	-36.6	-37.3
4392.8	-40.6	-39.7
4565.0	-43.6	-40.9
4737.3	-46.7	-43.5
4909.6	-47.7	-43.9
5081.8	-48.0	-43.7
5254.1	-48.1	-41.8
5426.4	-46.4	-40.9
5598.6	-44.9	-39.4
5770.9	-44.1	-38.8
5943.2	-43.8	-39.9
6115.4	-44.1	-39.7
6287.7	-45.5	-38.5
6460.0	-45.9	-38.0
6632.2	-44.7	-37.8
6804.5	-43.4	-35.3
6976.8	-42.2	-35.2
7149.0	-41.4	-36.8
7321.3	-42.2	-36.9
7493.6	-42.3	-37.0
7665.8	-41.5	-35.7
7838.1	-40.9	-36.5
8010.4	-40.2	-37.0
8182.6	-40.6	-37.0
8354.9	-41.1	-38.8
8527.1	-41.4	-39.7
8699.4	-41.4	-40.1
8871.7	-39.8	-41.5
9043.9	-40.9	-42.0
9216.2	-42.2	-44.3
9388.5	-44.4	-45.4
9560.7	-46.5	-46.6
9733.0	-48.0	-48.7
9905.3	-48.6	-50.1
10077.5	-50.9	-51.8
10249.8	-53.7	-53.4
10422.1	-54.0	-53.5
10594.3	-54.0	-53.5
10766.6	-54.1	-53.5
10938.9	-54.1	-53.5

F.4 《两头忙》长时平均谱

频率	含灯	正常
78.1	-25.5	-26.5
234.4	-12.9	-13.8
390.6	-7.6	-8.4
546.9	-6.3	-6.9
703.1	-2.3	-3.4
859.4	0.0	-1.0
1015.6	-1.1	-1.4
1171.9	-1.4	0.0
1328.1	-4.5	-2.3
1484.4	-8.9	-5.6
1640.6	-12.2	-10.1
1796.9	-15.8	-13.4
1953.1	-17.6	-17.8
2109.4	-17.0	-17.4
2265.6	-16.1	-18.0
2421.9	-13.6	-17.0
2578.1	-14.5	-17.6
2734.4	-14.8	-18.9
2890.6	-15.9	-19.5
3046.9	-16.9	-21.4
3203.1	-17.7	-22.6
3359.4	-20.2	-23.7
3515.6	-22.2	-25.8
3671.9	-24.0	-27.2
3828.1	-26.7	-30.2
3984.4	-31.0	-32.8
4140.6	-35.4	-34.2
4296.9	-39.5	-36.6
4453.1	-42.6	-38.5
4609.4	-45.0	-40.0
4765.6	-46.5	-40.8
4921.9	-46.7	-39.8
5078.1	-46.6	-39.0
5234.4	-45.9	-37.7
5390.6	-44.9	-36.8
5546.9	-43.8	-37.0
5703.1	-42.1	-36.0
5859.4	-41.2	-35.9
6015.6	-41.6	-35.9

6171.9	-42.5	-35.6
6328.1	-43.5	-36.1
6484.4	-42.6	-35.7
6640.6	-41.5	-35.1
6796.9	-39.8	-35.1
6953.1	-38.9	-34.9
7109.4	-39.0	-35.1
7265.6	-39.2	-35.3
7421.9	-39.8	-35.2
7578.1	-40.0	-34.4
7734.4	-39.6	-33.9
7890.6	-38.8	-34.0
8046.9	-37.8	-34.7
8203.1	-38.4	-35.6
8359.4	-39.1	-36.5
8515.6	-39.1	-37.5
8671.9	-39.1	-38.1
8828.1	-39.8	-39.4
8984.4	-40.6	-40.3
9140.6	-41.3	-41.5
9296.9	-43.3	-42.4
9453.1	-45.3	-43.4
9609.4	-46.7	-44.3
9765.6	-47.5	-45.0
9921.9	-47.9	-45.3

F.5 《梅花的故事》长时平均谱

频率	含灯	正常
86.1	-27.7	-27.3
258.4	-14.1	-13.8
430.7	-8.4	-7.9
602.9	-6.0	-5.3
775.2	-1.0	-1.5
947.5	0.0	0.0
1119.7	-2.5	-2.1
1292.0	-4.3	-2.9
1464.3	-8.8	-7.1
1636.5	-13.4	-11.7
1808.8	-16.5	-16.5
1981.1	-18.5	-19.9
2153.3	-16.7	-18.4

2325.6	-17.8	-20.5
2497.9	-15.2	-19.3
2670.1	-16.9	-20.9
2842.4	-17.3	-20.8
3014.6	-19.3	-21.9
3186.9	-20.9	-24.5
3359.2	-21.9	-25.6
3531.4	-24.8	-27.5
3703.7	-25.8	-29.1
3876.0	-28.0	-32.3
4048.2	-32.1	-35.6
4220.5	-37.5	-37.4
4392.8	-41.9	-40.3
4565.0	-45.5	-42.1
4737.3	-47.5	-43.5
4909.6	-47.9	-42.8
5081.8	-47.8	-41.6
5254.1	-47.3	-40.2
5426.4	-46.4	-39.0
5598.6	-45.5	-38.6
5770.9	-44.0	-37.6
5943.2	-43.0	-37.4
6115.4	-43.1	-37.4
6287.7	-44.0	-37.7
6460.0	-43.9	-37.8
6632.2	-42.1	-37.6
6804.5	-40.6	-37.5
6976.8	-39.1	-36.5
7149.0	-39.3	-37.3
7321.3	-40.2	-37.2
7493.6	-39.7	-37.1
7665.8	-39.6	-37.5
7838.1	-38.7	-37.1
8010.4	-38.0	-37.3
8182.6	-38.0	-38.0
8354.9	-38.9	-39.1
8527.1	-39.0	-40.3
8699.4	-40.4	-40.9
8871.7	-41.6	-42.3
9043.9	-42.7	-43.3
9216.2	-43.3	-44.1
9388.5	-44.8	-44.8

9560.7	-47.3	-45.4
9733.0	-48.3	-45.8
9905.3	-48.9	-46.4
10077.5	-51.2	-49.2
10249.8	-54.5	-53.9
10422.1	-54.7	-54.3
10594.3	-54.7	-54.3
10766.6	-54.8	-54.3
10938.9	-54.8	-54.4

F.6 《四世同堂》长时平均谱

频率	含灯	正常
86.1	-26.3	-26.6
258.4	-13.0	-13.8
430.7	-8.4	-9.3
602.9	-5.8	-6.5
775.2	-0.1	-0.1
947.5	0.0	0.0
1119.7	-3.8	-2.9
1292.0	-6.0	-2.5
1464.3	-10.5	-6.9
1636.5	-14.8	-12.7
1808.8	-15.8	-14.6
1981.1	-17.9	-18.3
2153.3	-16.2	-18.5
2325.6	-16.9	-21.0
2497.9	-16.6	-20.8
2670.1	-17.9	-22.2
2842.4	-18.2	-23.4
3014.6	-18.1	-24.0
3186.9	-19.8	-25.2
3359.2	-22.4	-26.1
3531.4	-24.2	-27.8
3703.7	-26.2	-31.1
3876.0	-29.1	-34.1
4048.2	-35.0	-38.2
4220.5	-40.7	-39.5
4392.8	-43.8	-41.6
4565.0	-46.6	-41.8
4737.3	-46.9	-42.2
4909.6	-47.6	-41.6

5081.8	-46.5	-40.6
5254.1	-45.5	-40.3
5426.4	-45.4	-38.6
5598.6	-44.1	-37.4
5770.9	-43.3	-37.9
5943.2	-41.0	-37.3
6115.4	-42.4	-37.7
6287.7	-42.9	-37.6
6460.0	-41.0	-38.1
6632.2	-40.4	-38.5
6804.5	-38.8	-38.2
6976.8	-38.2	-37.3
7149.0	-37.6	-37.0
7321.3	-38.4	-37.3
7493.6	-39.9	-36.5
7665.8	-40.3	-36.8
7838.1	-39.8	-36.9
8010.4	-39.2	-37.1
8182.6	-39.3	-37.9
8354.9	-39.7	-38.6
8527.1	-40.4	-39.3
8699.4	-40.6	-40.1
8871.7	-41.0	-41.2
9043.9	-42.4	-42.5
9216.2	-43.9	-44.1
9388.5	-45.6	-45.1
9560.7	-46.7	-46.5
9733.0	-47.7	-47.2
9905.3	-49.1	-48.2
10077.5	-51.6	-50.8
10249.8	-54.7	-54.1
10422.1	-54.9	-54.4
10594.3	-54.9	-54.4
10766.6	-54.9	-54.4
10938.9	-54.9	-54.4

F.7 《一声祝福》长时平均谱

频率	含灯	正常
86.1	-26.3	-27.7
258.4	-12.5	-13.4
430.7	-6.8	-7.2

602.9	-4.0	-3.0
775.2	0.0	0.0
947.5	-0.1	-1.4
1119.7	-1.5	-2.2
1292.0	-2.2	-1.9
1464.3	-6.0	-4.3
1636.5	-9.9	-9.0
1808.8	-13.7	-15.2
1981.1	-16.4	-19.4
2153.3	-16.0	-19.1
2325.6	-15.4	-19.0
2497.9	-14.6	-17.5
2670.1	-15.2	-17.2
2842.4	-14.2	-16.7
3014.6	-16.2	-19.2
3186.9	-18.4	-21.6
3359.2	-19.5	-23.1
3531.4	-21.7	-25.5
3703.7	-24.2	-28.0
3876.0	-28.6	-31.3
4048.2	-34.8	-35.1
4220.5	-39.1	-38.1
4392.8	-40.8	-40.4
4565.0	-43.1	-42.1
4737.3	-45.3	-43.8
4909.6	-45.9	-44.0
5081.8	-45.3	-42.5
5254.1	-44.2	-41.0
5426.4	-43.0	-39.5
5598.6	-41.8	-38.1
5770.9	-40.1	-37.6
5943.2	-39.4	-37.0
6115.4	-39.9	-36.6
6287.7	-40.2	-36.6
6460.0	-39.2	-36.7
6632.2	-38.1	-36.6
6804.5	-37.5	-36.0
6976.8	-37.2	-35.4
7149.0	-37.0	-36.0
7321.3	-37.3	-36.4
7493.6	-36.8	-35.9
7665.8	-36.7	-35.4

7838.1	-36.3	-35.3
8010.4	-36.8	-35.7
8182.6	-37.2	-36.6
8354.9	-37.5	-37.8
8527.1	-38.4	-39.0
8699.4	-39.2	-39.6
8871.7	-39.8	-40.1
9043.9	-40.3	-41.5
9216.2	-42.8	-42.8
9388.5	-44.7	-43.7
9560.7	-46.2	-44.3
9733.0	-47.3	-44.2
9905.3	-47.9	-44.4
10077.5	-49.7	-47.1
10249.8	-53.1	-53.0
10422.1	-53.4	-53.7
10594.3	-53.4	-53.7
10766.6	-53.4	-53.7
10938.9	-53.4	-53.7

F.8 《昭君出塞》长时平均谱

频率	含灯	正常
78.1	-28.1	-27.4
234.4	-14.8	-14.5
390.6	-8.4	-7.5
546.9	-8.3	-7.2
703.1	-3.2	-2.6
859.4	0.0	0.0
1015.6	-1.6	-0.3
1171.9	-3.3	-1.0
1328.1	-3.9	-1.1
1484.4	-7.9	-5.1
1640.6	-11.3	-9.6
1796.9	-13.4	-12.1
1953.1	-16.4	-15.9
2109.4	-16.2	-16.2
2265.6	-17.3	-18.5
2421.9	-14.3	-16.2
2578.1	-16.1	-17.8
2734.4	-15.7	-16.9
2890.6	-17.1	-19.4

3046.9	-19.1	-22.0
3203.1	-19.4	-23.1
3359.4	-21.2	-23.3
3515.6	-22.5	-26.1
3671.9	-22.1	-27.5
3828.1	-25.2	-30.6
3984.4	-29.7	-33.5
4140.6	-33.7	-34.5
4296.9	-40.1	-37.7
4453.1	-42.8	-39.2
4609.4	-45.3	-39.8
4765.6	-47.3	-41.1
4921.9	-48.0	-41.4
5078.1	-47.7	-41.7
5234.4	-47.1	-40.0
5390.6	-46.0	-37.9
5546.9	-45.0	-38.4
5703.1	-42.8	-36.3
5859.4	-41.9	-36.8
6015.6	-42.2	-37.1
6171.9	-43.4	-37.3
6328.1	-43.9	-37.7
6484.4	-43.0	-37.7
6640.6	-42.6	-37.4
6796.9	-41.4	-36.9
6953.1	-40.1	-36.9
7109.4	-40.0	-36.5
7265.6	-39.5	-36.5
7421.9	-39.3	-36.4
7578.1	-40.1	-36.7
7734.4	-39.9	-36.0
7890.6	-40.7	-36.6
8046.9	-40.1	-36.6
8203.1	-40.3	-36.5
8359.4	-41.1	-37.5
8515.6	-40.6	-38.4
8671.9	-41.2	-39.9
8828.1	-43.1	-41.3
8984.4	-44.8	-42.2
9140.6	-45.7	-43.3
9296.9	-46.2	-44.0
9453.1	-47.7	-44.5

附录 F 长时平均谱数据

9609.4	-49.1	-45.5
9765.6	-49.6	-46.2
9921.9	-49.8	-46.1

后记

随着毕业论文的完成，我也即将硕士毕业，心中百感。

硕士三年要感谢的人很多。首先要感谢的是我的导师孔江平教授。在研究生期间，导师给了我极大的支持与鼓励。无论是前期课程的学习，还是后期论文的写作，孔老师都认真、耐心地给予我指导。作为一名学生，能够遇到如此良师心中满是感激。在语音乐律实验室期间，除了严谨的学风、科学的方法外，我学到了很多宝贵的做人、做事的道理，这将成为一种潜在的东西对我有长期的影响。

其次要感谢的是各位同门。各位同门就像兄弟姐妹一样聚集在这个大家庭中，生活上、学习上都互相照顾和帮助，这份类似亲情的感情会一直珍藏在我的心里。大家在五院二楼的实验室一起吃西瓜校稿、作为团队一起为几个会议忙碌、围坐在孔老师办公室喝茶，这些美好的记忆将一直闪亮。

还有很多其他同学也很感谢。同级的几位好友、三位室友、其他一些接触过但不熟的同学，虽未一一点名，但都谨记在心。另外还有方言调查小团队的五位同学、上海语音学研修班认识的几位其他院校的同学也一并感谢。因为学术大家走到了一起，但我看到的是一个个鲜活的灵魂，他们感染着我、激励着我。

假期几近包场的图书馆阅览室、令人平静的闭馆音乐、长长的走廊一侧的检索柜；被窗帘遮住隐隐约约的塔、绿色的湖水、满地的黄色银杏叶；二教的弧形教室、三教冻人的冷气、冬日有橘黄色阳光照进来的文史楼；三教后面夜色中的杏花、沿路的迎春花和连翘、三角地墙内的桃花；生科楼顶、遥感楼楼顶、夜晚的石坊；图书馆楼顶的乌鸦、湖里的锦鲤、各处的猫。还有什么？我都不想遗漏。

喜爱北大的原因有很多，于我自己的还有壁球馆、餐厅合胃口的饭菜。

但这些都还不够，比起从我身边走过的那些人。

校园教给了我很多，愿我不负君，满载而归。



北京大学

博士研究生学位论文

题目： 基于呼吸的汉韩韵律对比研究

姓 名： 吴 韩 娜

学 号： 10720806

院 系： 中国语言文学系

专 业： 语言学及应用语言学

研究方向： 实 验 语 音 学

导师姓名： 孔 江 平 教授

二 0 一 五 年 五 月

基于呼吸的汉韩韵律对比研究

内容提要

韵律不仅具有辨义功能,而且是影响语音表现力的主要因素,声音的轻重缓急、抑扬顿挫、不同语气和口气的传达、个人语音特征的辨别,都要依靠韵律的作用。因此,韵律研究在当前的语音研究中占有重要位置,深入理解韵律不仅是语音学理论发展的需要,也是语音合成和语音识别发展的需要。

由于语流的韵律构造是一个复杂的过程,它受到生理、心理、语法等各个方面的制约,因此研究者的角度也各不相同。而目前的研究大多集中在音系学、声学 and 心理学方面,从生理角度进行研究的并不多。呼吸运动是人类维持生命的重要功能,也为言语提供动力,满足言语交际的需求,与此同时,言语也影响着呼吸的形式和特点,因此对言语呼吸进行探索非常具有意义。

本文^①在前人研究的基础上从呼吸信号和语音信号两个方面来考察汉语和韩语的不同文体中的呼吸重置特征与相关韵律特征,从而找到

1. 语音共性和不同语言之间的呼吸与韵律特征的差异,为汉语和韩语韵律节奏的深入研究提供参照,并将研究结果运用到语言教学中;
2. 各个韵律单元的呼吸信号和声学参数之间的关系;
3. 以语义相同而句法结构不同的汉韩语语料进行对比分析,能够揭示句法结构对呼吸节奏与韵律构造的影响;
4. 探讨不同语言对表达内容的整体规划方式和呼吸节奏的调节。

该研究对言语呼吸生理机制的认识和理解具有重要意义,为言语产生呼吸生理建模提供研究基础。

本文研究结果显示,因为汉语近体诗的格律严格,形式固定,从而不管哪种语言哪种形式,朗读近体诗中整体呼吸格局有一定的一致性。

与之前的研究结果最显著的不同点是,不管是哪种语言的呼吸曲线,韵律短语边界处不仅会出现抖动,还会出现明显的小呼吸。有趣的是,这些小呼吸的出现的位置往往与声学分析中的停顿的位置和时长无关。

不同语言版本之间主要差别在于:朗读汉语近体诗时,胸、腹呼吸分工明确,胸呼吸与生理信号(语音信号)对应,腹呼吸与韵律单位对应,而韩语版朗读中,胸、腹呼吸功能比较模糊。

从五言近体诗到七言近体诗,汉语版和韩语版呼吸变化趋势也截然不同,但是,汉语近体诗格律的严格性制约不同结构语言的节奏特征,虽然呼吸配合方式不同,但韩语朗读者调整呼吸重置幅度与次数而体现汉语近体诗节奏特征。可见,此结果既可以显示汉语近体诗格律的特征,也可以表现韩语音节结构的特征“等时性”。从整体呼吸安排来说,汉语近体诗格律优先于不同语言节奏特征,不管是哪种语言,五言还是七言,朗读者按照严格的近体诗格律来安排

^①本研究得到了以下项目的资助:中国哲学社会科学重大投标项目“中国有声语言及口传文化保护与传承的数字化方法研究和基础理论研究”(批准号:10&ZD125)和国家自然科学基金面上项目“汉语普通话语音生理模型研究”(批准号:61073085)

呼吸方式，体现近体诗的情感以及近体诗格律带来的节奏感，该特征具有一定的普遍性。这还说明，呼吸安排决定于格律和朗读者对诗歌内容的不同解释而产生的情感，基本上不在于生理上的需求。

不管哪种语言，朗读新闻中整体呼吸格局有一定的一致性。朗读中胸、腹呼吸基本分成三级，腹呼吸重置的幅度、频度都高于胸呼吸的。在大多数情况下，每段话的第一句会出现大呼吸，表现形式主要有阶梯式呼吸或者一个大呼吸。

朗读汉语新闻时，胸、腹呼吸分工更加明确，就是说在开始朗读时给整个言语活动提供能量，比胸呼吸早点结束而朗读中对韵律单位的界限更敏感，韵律短语边界处常出现腹呼吸，从而呼吸次数比胸呼吸多一些；开始呼气之后保持一定时间使言语活动持续下去，因此胸呼吸的终点和言语信号终点基本吻合，胸呼吸与整个言语活动的完成有关。虽然韩语胸腹呼吸之间的分工比汉语版呼吸特征模糊，但还是可以说，这些不同功能属于朗读言语中出现的普遍性特征。

从呼吸安排与语言结构而言，汉语与韩语之间最显著的共同点是，不管朗读哪种语言，文本中出现的标点符号引发生理呼吸重置，如，句号（。）、逗号（，）和顿号（、）处都会出现呼吸重置。

值得注意的是，以往的声学研究表明，韩语的停顿单位大于汉语，但是通过生理呼吸信号分析结果发现，汉语的呼吸单位反而大于韩语的，与声学分析结果相反。汉语新闻朗读中，停顿数远远多于呼吸重置数。对汉语发音人而言，停顿对应于小语义单位，相比呼吸重置对应于大语义单位，主要以大呼吸或者中呼吸来完成言语活动，朗读汉语新闻中的呼吸节奏主要受标点符号的制约。与此相比，韩语新闻朗读中，声学分析上的停顿数与呼吸重置次数相差不大，停顿位置与呼吸重置位置基本重合。由此可见，对韩语发音人而言，停顿与呼吸重置是功能相等的节奏单位，主要以小呼吸来完成每个语义单位。不同认知方式直接影响汉语与韩语两种不同语言的呼吸重置平均幅度。

从呼吸幅度与时长的相关关系上，汉语和韩语之间存在显著性差异：新闻朗读中，不管是哪种语言，胸呼吸幅度与时长之间成正比，存在显著性相关。与此相比，腹呼吸重置与时长相关分析结果显示，汉语新闻朗读中，大呼吸与中呼吸出现的频度极高，不管是小语义单位还是整个句子，几乎无规律地出现。因而，汉语腹呼吸重置幅度与时长不存在显著正相关，只有韩语新闻朗读中的腹呼吸重置与时长呈正相关。

总之，这些结果既可以显示朗读新闻中的呼吸节奏特征，还可以发现语言的不同结构会影响整个呼吸节奏的安排和呼吸方式。另外，从生理信号与声学信号对应中发现，声学分析结果还是不能完全解释言语过程中的认知方式与过程，有待进一步深层研究。

关键词：

汉语普通话 韩语 呼吸节奏 韵律

A Comparative Study on Breathing Rhythm Properties in Chinese and Korean

Oh Hanna [Phonetics]

Directed by **Professor Kong Jiangping**

Abstract

Prosody not only has the function of meaning discrimination, but also is the main factor to affect the performance of the speech, the varieties of sound, cadence, different mood and tone, personal voice characteristics to identify all rely on the role of prosody. Therefore, currently, study of prosody occupies an important position in research of phonetics. In-depth of the rhythm is not only the need of phonetics theory development, but also the need of the development of speech synthesis and speech recognition.

Because of the flow of the language rhythm structure is a complicated process, it is under the restriction of many aspects, such as physical, psychological, grammar, so researchers point of view are also different. Now, most studies focus on phonology, acoustics and psychology, physiological research are very few. Respiration is an important function of human life. It also provides the impetus for the speech, to meet the communicative needs, at the same time, the speech also affects the forms and characteristics of breathing, so the study of speech respiration is very meaningful.

In this paper, different styles of breathing reset feature on the basis of previous research from the two aspects of respiratory signals and voice signals to examine the different style of Chinese and Korean in breathing reset features and the related prosodic features, so as to find

1. The commonality and differences of breathing and prosodic features between different languages to provide reference for further study of Chinese and Korean rhythm, and the results can applied to language teaching as well;

2. The relationship between respiratory signal and acoustic parameters of different prosodic unit;

3. This paper compares and analyzes Chinese and Korean materials with the same semantic and different syntactic structure, and can reveal syntactic structure impact on breathing rhythm and rhythm structure;

4. Investigate the expression of different languages on the contents of overall planning and the regulation of respiratory rhythm.

This research has great significance in understanding and understanding of

speech respiratory physiology mechanism, provides the basis for the study of speech production model of respiratory physiology.

The results of this study show that, because of the strict rules of Chinese modern style poetry, fixed form, so no matter what kind of language which forms, whole breathing pattern has a certain consistency in the modern style poetry.

The most significant difference with previous research results is, in both languages of respiration curve, prosodic phrase boundaries can appear not only shakes, there are obvious small breathings as well. It is interesting to note that these little breathing pauses are independent of the position and length of pauses in acoustic analysis in most cases.

Main differences between different languages is that : When Chinese speakers read Chinese modern style poetries, chest and abdominal breathing have clear division of responsibilities between them, chest breathing is corresponding to physiological signals (speech signal), abdominal breathing is corresponding to prosodic unit, and in the Korean version, chest and abdominal breathing function is less clear.

From the five characters poetry to the seven characters poetry, Chinese version and Korean version of respiratory variation trends are very different, however the Chinese modern style poetry controls different structure of language rhythm feature strictly while breathing in different ways. The Korean speakers adjust breathing reset amplitude and frequency as well as the reading speed and rhythm for reflecting the Chinese poetry rules and characteristics.

Therefore, the results show the features of Chinese modern style poetry, also can show the feature of Korean syllabic structure of "isochronous". From the overall arrangement of breathing, Chinese modern style poetry rules prior to the different characteristics of language rhythm, regardless of language, five or seven words, the readers arranges the breathing patterns in strict accordance with the modern style poetry rhythm, reflects the modern Chinese poetry emotions and bring a sense of rhythm, this feature has a certain universality. It also indicates that, breathing arrangement depends on the rhythm and the readers' different interpretation of the poem content and emotions, basically not on the physical demands.

No matter what kind of language, read the whole breathing pattern in the news have a certain consistency. When the speakers read news, the chest and the abdominal breathing basically divided into three stages, the amplitude and frequency of abdominal breathing reset are higher than that of chest breathing. In most cases, the first sentence of each paragraph will be a big breath, mainly in the form of a ladder or a big breath.

When the Chinese readers read Chinese news, division of chest and abdominal breathing is clear. The abdominal breathing provides energe to beginning of the reading, ends earlier than chest breathing. And it is more sensitive to the boundaries of prosodic unit in reading, so it often appears in prosodic phrase boundaries. Consequently, the abdominal breathing appears more often than chest breathing.

When the exhale begins, it has to keep a certain time to make continue to speech activity. The end point of chest breathing and speech signal are matched basically, so it shows that the chest breathing is associated with the completion of the whole speech activities. Although the division of between chest and abdominal breathing in Korean version is unclear than Chinese version, but still can be said, these different functions between chest and abdominal breathing belong to universal features.

For the breathing arrangement and the language structure, the most striking similarity between Chinese and Korean is, no matter which language, punctuation mark in the text cause physiological breathing reset, such as full stop(。), comma(,) and pause(、) marks, the readers reset their breathe at these marks.

It is notable that previous acoustic researches showed that the Korean pause unit bigger than Chinese one, but the physiological respiration analysis results showed that the Chinese respiratory unit is bigger than the Korean one, it's the opposite of the acoustic analysis results. When the Chinese readers read news, the number of pause are far more than the number of breathing reset. For Chinese, the pause corresponds to the small semantic unit, compared with the breathing reset corresponds to the big semantic unit, the breathing rhythm is mainly controlled by punctuation marks.

In contrast, when the Korean readers read news, the number of pauses of the acoustic analysis and the number of breathing reset are similar, places of pauses and breathings are basically equal. This shows that, for the Korean readers, pause and breathing reset are same rhythmic units with same function, mainly the small breathing completes each semantic unit. As a result, different cognitive ways affect the amplitude of breathing reset directly in Chinese and Korean.

For the correlation between the respiratory amplitude and duration, there are significant differences between Chinese and Korean breathing rhythm of reading. Regardless of language, the chest breathing amplitude is proportional to duration, there was a significant correlation. In contrast, for the abdominal breathing reset, only the abdominal breathing of Korean new reading showed amplitude and duration are positively related.

In conclusion, these results can be showed the features of respiratory rhythm in text with different style. It also can be found the different structure of language affect the whole arrangement of breathing rhythm. In addition, we found that acoustic analysis results still cannot fully explain the cognitive way of speech in the process of different languages, it is still needed further research.

Key words

Chinese Korean Breathing Rhythm Respiration Prosody

目录

第一章 绪论	1
1.1 研究对象的定位	2
1.1.1 语言的韵律和节奏	2
1.1.2 语言的停顿和呼吸	3
1.2 汉语和韩语呼吸韵律研究	4
1.2.1 汉语韵律研究	4
1.2.2 韩语韵律研究	13
1.3 研究目的和意义	17
1.4 章节安排	18
第二章 研究方法	19
2.1 实验方法	20
2.1.1 言语呼吸节奏的生理机制	20
2.1.2 研究设备	23
2.1.3 参数定义和程序介绍	24
2.1.4 参数提取和处理	27
2.2 研究内容	29
2.2.1 实验语料	29
2.2.2 发音人	29
第三章 汉语和韩语诗歌的呼吸与韵律特征对比	30
3.1 汉语诗歌的呼吸与韵律特征	32
3.1.1 汉语诗歌的呼吸节奏特征	32
3.2 韩语诗歌的呼吸与韵律特征	42
3.2.1 韩语汉字音版诗歌的呼吸节奏特征	42
3.2.2 韩语翻译版诗歌的呼吸节奏特征	53
3.3 结果	63
第四章 汉语和韩语新闻的呼吸与韵律特征对比	65
4.1 汉语新闻的呼吸与韵律特征	66
4.2 韩语新闻的呼吸与韵律特征	73
4.3 结果	80

第五章 结论	81
5.1 研究结果	81
5.2 创新之处	83
5.3 问题与展望	84
参考文献	85
附录 1 录音材料（汉语版）	91
附录 2 录音材料（韩语版）	99

第一章 绪论

语言是人与人之间进行信息交互的重要手段，也是最直接、最自然、最方便的方法。使用语言进行交流时，相互传递的不仅仅是语言的文字信息，还包含语言表达的情感信息。

语言中声音的高低、轻重、长短、快慢、间歇和音色造成语言的韵律。说话人的语气、态度、感情色彩、个人特点就在话语的韵律特征中都有所体现，因此韵律信息在语言表达的自然度和理解程度方面扮演了非常重要的角色。它不仅将长句以韵律短语的形式分割为一个个较短的小单元，以更便于人们理解或者机器处理，各韵律短语之间的间隔也为消除句法歧义提供了非常重要的信息。由于在语言中的重要地位，韵律特征研究受到语言学、心理学、声学以及计算机科学、人工智能等诸多领域的重视。

语言教学领域也一直特别关注韵律特征研究。由于汉语有声调，汉语和其他语言进行对比时，韵律成分可以说比音段成分更重要，所谓的“洋腔洋调”在很大程度上也取决于声调和比声调更高的语音层次，即韵律的掌握程度（林焘，1996；郑明叔，2007）。

近十几年以来，汉韩语言对比研究日益增加，而语篇节奏对比研究很少见。汉语和韩语是不同语系的两种语言，包括语音和句法结构等语言特征差别很大，如，从语言类型学的角度看，汉语属于孤立语，韩语则属于黏着语。从语序而言，汉语属于将动词放在宾语前的 SVO 型语言，也属于声调语言，以音步（foot）为基本节奏单位。韩语则属于把动词放在句尾的 SOV 型语言，也是一种以重音短语（accentual phrase）为节奏单位的语调语言。由于两种语言具有不同语言特征，如果对语篇韵律特征进行对比研究的话，不仅使汉语和韩语两种语言的韵律特征更加明确，而且还能够从多角度显示语言韵律结构特征的共性和对韵律特征影响的种种因素。

在韵律研究中，韵律单位的切分和层级构造一直是研究的热点。由于语流的韵律构造是一个复杂的过程，它受到生理、心理、语法等各个方面的制约，因此研究者的角度也各不相同。而目前的研究大多集中在音系学、声学和心理声学方面，从生理角度进行研究的并不多。

在生理方面，根据 Liberman (1967) 提出的呼吸群（breath-group）理论，谭晶晶（2006，2007）利用机电和呼吸带传感器记录了呼吸信号，发现呼吸结构可以反映人们对表达内容的整体认知规划和生理机能的制约。其中生理机制主要指的是呼吸调节，呼吸是划分韵律层级的重要线索（郑秋豫，2006）。即呼吸运动是人类维持生命的重要功能，也为言语提供动力，满足言语交际的需求，与此同时，言语也影响着呼吸的形式和特点，因此对言语呼吸进行探索非常具有意义（王毓钧，2012）。

鉴于以上几点，本文打算从呼吸节奏这一新的角度着手，以对比研究的方法，探讨汉韩两种语言的不同文体及语流中的呼吸节奏的类型和特点，并试图寻找呼吸信号和声学信号怎么对应、呼吸节奏和韵律结构的关系以及不同句法结构对呼吸节奏的影响。

1.1 研究对象的定位

1.1.1 语言的韵律和节奏

韵律 (Prosody) 是关于语言中所有“超音段特征 (suprasegmental features)”的一个总体的概念。Hadumod Bussmann (1996) 在《语言与语言学词典》中指出, 韵律是指“重音 (stress)、语调 (intonation)、音长 (quantity) 和停顿 (pauses) 等超音段语言特征”, 也包括“语速 (speech tempo) 和节奏 (rhythm), 叫超音段特征”。

从语言学的角度, 韵律指音段层次以上的音系组织。是将各种语言学单位组织成话语或话语中关联组块的系统组织。从物理学角度, 是指基频、时长、和强度等声学参数。从感知层面来讲, 语言的韵律是音调和节奏的变化所产生的感知模式 (Nooteboom, 1997)。音调在细分为各种特定的语调要素, 节奏再细分为不同层次的韵律单元, 这些韵律要素之间有密切的关系, 他们在语流中的交叉作用形成了一个语言的韵律特征。

韵律的实现涉及语音的音段和超音段特征 (即韵律特征), 它不但能够传递语言学信息, 而且能够传递副语言学和非语言学信息 (包括传达焦点、强调、短语分组等), 另外还有区分词汇、表达语气态度等功能。根据韵律特征的不同又可以划分出不同的韵律单元, 如音步、韵律词、次要韵律短语、语调短语、韵律组/语调组等。

关于韵律概念的界定, 赵元任 (1980) 先生早在五十年代末就曾经讲过“上加成素” (suprasegmental elements), 包括长短音、重音、断续度、语调、速度、嗓音等。罗常培和王均先生 (1981) 也采用韵律概念进行描述, “语言中声音的高低、轻重、长短、快慢、间歇和音色造成语言的韵律。语言的韵律是人们为了准确的转达词句的内容, 表现说话人的思想感情, 以求达到相互了解的必要手段。语言中的句子自然而然可以分成许多节落, 这就是意群; 每一个节落又可以分成许多小节落, 就是所谓节拍群。节落和节落之间有长短不等的间歇, 句子里的词或词里的音节又各有不同的高低轻重 (声调、重音) 和快慢不等的速度, 有时咱们还利用若干句子的某些地位上的相同音色的反复再现 (双声、叠韵、韵脚、腹韵等), 来造成语言的某种气氛和情调, 这一切都构成了语言的节律, 或者说是语言的节律感。所以语言的节律就是音和音的相对关系和组合关系。”二位先生对语言节律的界定, 显然还是把重点放在了超音段因素上。人们经常提到的声调 (tone)、连调 (tone sandhi)、轻重音、停顿、语调和节奏等, 都是某些超音段特征的表现, 都可以被看作是韵律的一种。从这个意义上讲, “节奏”是“韵律”的下位概念 (叶军, 2008)。

其实“节奏 (rhythm)”本身指音乐中, 交替出现的有规律的强弱、长短的现象。节奏寓于音乐之中, 对音乐具有重要的意义。因为音乐有三大要素, 节奏就是其中之一。即, 和声 (harmony)、旋律 (melody) 和节奏 (rhythm)。可以说, 有了节奏音乐才有了音节的长短和强弱的交替变化, 才有了赏心悦耳。《礼记·乐记》也说“文采节奏, 声之饰也”。

然而节奏不仅在音乐之中, 还存在于世界的每个角落。在自然界的角度, “节奏”是指一种有规律的、反复出现的运动形式。如, 季节的变化、昼夜交替、脚步声和脉搏等等。在生活中的这些节奏无处不在, 使我们的日常生活能

够变得更加和谐平衡。某种意义上说,节奏就是实物运动规律的外在表现。从语言学的角度,最有规律的形式,诗歌就以“格律(meter)”这一术语来说明根据诗歌每行所用的音步(foot)的数目和种类而配置的不同方式(语音学与音系学词典)。“音步”就是英语的节奏单位,一个音步由一个或多个音节组成,第一个音节必须是重音,而其后的音节均为非重音(Halliday,1970)。《现代语言学字典》(2007)中的定义是“按其一般含义用于音系学,指言语中可感知的单位突显的规律性,这种规律性可用重读对非重读音节、音节长度(长对短)或音高(到对低)等型式或这些变项的组合形式来说明”。

朱光潜先生还从语流的角度论述了节奏,他在《诗论》中提出“声音的段落”的概念,认为通过声音段落的断续、起伏和交替,可以形成节奏。他认为节奏的产生有三个条件:一是要有时间段落来打破时间延绵的直线;二是时间段落间差别不能过大;三是时间段落后要形成起伏规律;三者结合,也就是“节奏是声音大致相等的时间段落里所产生的起伏。”

语音有四大要素:音色、音长、音强、音高。王力(2002)和王洪君(2008)都认为音长、音强和音高三个要素都可能和节奏发生关系。语言节奏的关键是这些超音段要素倾向于等距离的小间隔的回复。由此,我们可以概括地说,语言节奏的要素是在言语中以“等时性(isochrony)”为基准某种单位的反复出现,即节奏形成的必要条件是“周期性”和“反复性”。

Allen(1975)还用节奏的“普遍性(universal)”和“制约性(constraints)”来阐述语言的节奏和人类行为之间的关系。“语言是人类产生,也是人类认知的。因此,和人类所用的其他肌肉的运动和认知行为一样,语言也会被节奏制约”。他认为语言是人类行为之一,所以所有人类行为和节奏有关系的话,语言也不能例外。可以说,所有人类的语言都具有节奏。他所提出的制约性指的是人类生理条件本身的限制性。因此,他强调节奏结构要简单,主要出现的形式限于连续(succession)和交替,交替一般在每两个或三个音节出现一次”。但是,从语言学意义的角度看,节奏并不是生理学特征的单纯反映,如果真是那样,每个语言的节奏模式应该是一样。而大多数学者都认为,不同语言具有不同节奏形式。Dauer(1983)也指出,节奏的特征的不同在于语言结构的不同。这意味着语言的节奏也只要学习才能习得,还告诉我们不同结构语言之间的节奏特征对比的重要性以及生理学和语言学等交叉学科研究的必要性。在语言单位形成节奏时,能够反映人的生理学限制的是“停顿”和“间歇”。下面简单介绍停顿的定义和在言语中的表现。

1.1.2 语言的停顿和呼吸

林焘先生和王理嘉先生(1992)在《语音学教程》中说:“在比较长的句子里,有的音节关系比较近,结合比较紧密,有的则比较疏远,形成长短不等的音节组合。”林焘先生(1996)在《汉语韵律特征和语音教学》一文中还说:“各音节之间往往有或长或短的停顿。”可见,停顿存在于音节与音节之间,并且大多存在于较长的句子里。

随着各种韵律特征研究的深入,停顿也越来越得到人们的重视,郑秋豫(2001)认为,韵律现象的主要微信不只是语流单位,停顿也是一个主要微信,而且其地位不亚于各韵律单位。

吴宗济、林茂灿(1989)把停顿定义为“人们在说一段话时,中间总是会

有一些间歇，这些间歇就叫做停顿”。许毅（1986）通过对汉语普通话的连续语流所作的音联研究，发现汉语普通话里至少存在着四个不同等级的音联：（1）闭音联（2）音节音联（3）节奏音联（4）停顿音联。他也认为语流中的短暂间歇就是停顿。

关于停顿，郭锦桴（1993）定义为“语句中以及语句之间的间歇”，认为停顿产生是出于三种需要：一是语法上的需要，二是语义表达的需要，三是说话时换气的生理需要。当肺里的气流用完（太弱）的时候，必定要从外部吸入新气，以补充肺里气流的不足，然后才能呼气发声。就在这个吸气的瞬间，声音自然产生了停顿。停顿也可以分为语法停顿和心理停顿，语法停顿以标点符号为书面标志，心理停顿则是说话者为了表达的需要而有意安排的停顿，时间比较自由。

吴洁敏（1990）也从汉语普通话产生停延原因的角度将停延分为 1. 生理停延。2. 逻辑停延。即语法停延，是为清楚地表达言语结构层次而出现的停延。逻辑停延出现的频率最高。3. 强调停延。为了强调某一事物或突出表达某种思想感情而改变语法停延位置或时值，在语流中出现突破逻辑停延常规的一种变式停延。她认为生理停延不起表意作用，但它借助逻辑停延和强调停延获得吸气机会。逻辑停延必须服从强调停延。

良好的气息控制和节奏停顿把握是朗读的前提和基础，研究朗读气流和呼吸的变化可以揭示出朗读气息调节的规律。要研究朗读的呼吸规律我们首先要对呼吸机制有所了解。

1.2 汉语和韩语呼吸韵律研究

本章将要对汉语和韩语的呼吸节奏特征进行对比研究，探究二者之间存在的差异及原因，因此对于汉语和韩语的了解是本文的出发点。下面从声调语调的研究、重音的研究、停顿、韵律层级及边界研究、韵律特征预测建模等方面介绍了与节奏和停顿有关的汉语和韩语韵律研究情况。

1.2.1 汉语韵律研究

（1）声调语调研究

汉语是一种声调语言，孤立音节的声调一般不能任意改变；如果声调改变，那么音节的意义有可能就改变。但是在汉语连续语音中，受上下文音节声调、韵律信息如停顿层级、重音层级等因素的影响，声调的模式和孤立音节时相比发生了很大的变化。

普通话声调跟元音、辅音同样具有区别字义的作用，所以声调问题向来为语言学家以及有关言语工程学家所重视。对于声调的研究，集中在 F0、振幅及时长的研究上。本世纪二十年代，刘复用浪纹计测量了字音基频 F0，研究了其变化规律（刘复，1924）。林茂灿先生用音高显示器对普通话声调作了实验，看

到字音 F0 曲线上不仅有“调型段”，而且还可能有“弯头段”和“降尾段”（林茂灿，1965）。Howie(1972)利用声码器研究普通话声调的辨认问题，指出 F0 曲线是四声辨认的充分征兆。他还认为普通话声调定义并不是在音节的所有带音部分上，而是在主要元音和韵尾上。

在深入的研究中，吴宗济先生（1982，1985，1997，2002）自 80 年代以来就着手于普通话的变调研究，分别对二、三、四字组变调进行了调形分析，总结了短语中的连读变调规律。吴先生认为，普通话中的二字组、三字组、四字组是句子中的基本调群单元，语调就是由这些声调连读的基本单元组成的，在此基础上按照变调规则，“从句子中的最小单元变起，然后分层跟邻调发生作用，最终构成表层调型”。

林茂灿先生（1980，1984，1988）用实验方法对 F0 和时长进行测量，得出普通话中凡属同一调类的字，不管其声韵结构如何，单念时的调型段 F0 模式大体相同；在不同调类之间，调型段 F0 模式大体互有区别。他对自然语音做声学分析看到，普通话字音四声的调型段基频有不同的变化模式：字音时长在四声中没有固定的相对关系，但绝大多数的上声时长比其他三个调的长；上声振幅曲线并不都是双峰型的，其强度只是多数或大多数为最小。

沈晓楠（1992）通过对普通话四声和轻声组合的声学分析，指出协同发音不限于音段，它也出现在声调的产生中，他对普通话声调协同发音进行研究，得到声调协同发音的几个特性：

- 1、普通话声调受到顺向和逆向的两种协同发音的作用，这种双向作用是对称的。
- 2、普通话不但声调的始点和终点受到协同发音的影响，而且整个声调曲线受前后声调的作用而上下移动。
- 3、声调协同发音只影响 F0 的高度，不影响其方向。但是前后声调的 F0 方向影响 F0 高度的变化。当声调终点的音阶相同时，上升终点后面的声调比下降终点后面的声调高。
- 4、去声前面的去声半降是由于声调协同发音而不是由于变调引起的。
- 5、四个单字调对其前后声调施加主动作用，轻声则施加被动作用。
- 6、协同发音作用不能跨过声调，声调协同发音（陈娟文，2004）

对于声调和语调的关系，赵元任先生早在 1933 年就用两个经典的比喻——“大波浪和小波浪”和“橡皮条”理论——生动形象的说明了字调和语调的相互影响，指出了汉语的字调和语调的区别。

吴宗济先生从 80 年代开始，就一直从事并关注韵律的研究。吴先生认为，普通话二字组、三字组、四字组是句子中的基本调群单元，语调就是这些小单位为基础，合理运用变调规则，就能够生成符合情感的句子。

沈炯先生通过进行大量的实验，提出了“语调构造”理论，认为“语调是由一连串声调音域组织起来的音高调节形式，声调是在声调音域中滑动的曲拱”。

林茂灿（2004）认为汉语语调体现于两个变量：音高重调和边界调。只有边界调起到区分疑问语气和陈述语气的作用。疑问的边界调调型（音高模式），保持其本调形式，其调型并没有因为语调的作用而发生改变。汉语语调表现于音高重调和边界调的音高曲线的位置，而声调由其音高曲拱表征；语调和声调有不同的声学表现。

许毅（1999）也一直致力于研究汉语的声调和语调。他在叠加观和线性观的基础上，提出了汉语声调语调的平行观，言语中的音高有多种表意功能，这

些功能是相互独立的，它们在基频中同时、平行存在，各自有相应的有效的表现方式，认为声调是通过与音节同步的音高目标来实现的。

当然，在言语工程领域，有关汉语声调的一些声调合成或者声调识别研究或者建模方法也有很多。曹阳，黄泰翼，徐波（2004）提出采用决策树的数据驱动方法，从大规模语料中统计学习出连续语音中声调模式的分布。由于声调主要体现在基频轮廓曲线上（吴宗济、林茂灿，1989），顾良，刘润生（1998）引入声调差别的方法，有效提高了数码语音中“2”和“8”的识别率。声调识别的前提是建立声调模型，目前的声调模型有模糊模型（徐士林，1996）、HMM（隐马尔可夫模型）模型和人工神经网络模型（孙放，胡光锐，1997），其中HMM模型应用较多，主要是因为许多语音识别系统的声学模型就是HMM模型。人工神经网络(Artificial Neural Networks, ANN)具有高维性、自组织性、模糊性、冗余性、自学习能力等优良特性和较强的非线性区分模式边界的能力，因此特别适合于声调识别，其中的BP网络应用最为广泛（邵鹏飞，2009）。

（2）重音研究

重音分布是语句韵律特征的一个方面，以音高重音(accent)、强调(emphasis)和焦点(focus)等的语音手段，在语句组织和语义表达上起到十分重要的作用。同一个语句，由于重音位置的不同，往往会表达不同的语义。重音是一种语音现象，指的是在多音节的字词中将某个音节念得特别明显。

汉语是声调语言，但是汉语也同样具有重音，而且汉语重音也是韵律表现的一个重要的手段，因此是韵律特征的一个重要方面。

重音可以从不同的角度加以定义。从说话的角度看，重音是语言中重读的音，是说话时把在表情达意方面比较重要的词语说得用力一点的现象，在发音生理上表现为声道运动离中性位置远，发音特别到位甚至越位；从听觉感知的角度看，重音是语流中听起来相对响亮和凸显的成分，通常是指在词、词组、短语乃至句子中某些成分或某个音节突出于其他相邻音节的知觉印象；从声学语音学的角度、或者从客观的物理声学表现来看，重音是语音四要素综合增强的效应，往往是音长显著地加长，音高变化特别突出，具体表现为音阶或音域的抬高或下沉，音域动程增大，音强相对提高，音色也更加典型（曹剑芬，1994）。重音是一个相对量，它的存在是非常重要的。但是，对于重音特征的描述，各家都持有自己的观点：

赵元任先生（1980）将汉语的重音等级划分为三级：(1)特强的；(2)普通的；(3)轻的。他是在音位层次上对重音进行的划分。特强重音就是一般说的对比重音；普通重音是既不是轻声，又不是特强重音的重音；弱重音即一般说的轻声或轻音。他曾经提出，双音节词的重音分布是“中重”型的。这说明，从语音学的角度看，普通重音词的内部还可能有轻重之分。对于汉语重音的特征，赵先生指出重音表现为音域加宽、音程加大，其次是气流加强。

沈炯先生（1985）关于重音表现的论述是在对语势重音的研究中指出的。他指出，“在听辨语势重音时，音长的作用并不明显，音高的作用却很重要”。

林焘先生（1983）曾经用合成的方法探讨过北京话双音一节词中轻音的性质，发现对于分辨轻重音来说，音强起的作用很小，音长则是最重要的辨别因素，而音高在听辨轻音音节时所起的作用远没有音长重要，主要体现在音节起点的高低。

郭锦择(1993)经过实验分析得出关于语句重音的声学特征,“普通话语句重音的声学征兆主要表现在时长和音高的变化上,时长和音高并行变化,然而,重音时长的变化更具普遍性。几乎所有重音的时长都较长。”关于重音与音强的关系,郭锦择认为,“语句重音与音强关系不明显,但在强调重音中,音强有明显加强。”

林茂灿等(1984)从感知的角度研究了孤立的正常重音(即上文的普通重音)双音节韵律词(二字组)的重音分布,对各种声调组合构成的双字词的声学分析发现,普通话两字组中的轻重音分布主要是后音节较重:普通话两字组的轻重音在声学上的表现是具有较长的长度和较完整的音高模式,而不是具有较大的强度。

冯胜利(1997)认为节奏中的“轻重”是一种相对关系,是一种结构,即“相对轻重原则”。他明确的指出,汉语重音“所谓的‘句后’实质上是指句重音要落在句中最后一个韵律单位的最后一个成分上(除了一些特殊情况以外)。

端木三(1999)认为,有重音是所有语言的共性,汉语也不例外。他认为汉语的音步是左重的,是重轻交替的左重音步。另外,在辅重原则的支配下,重音得以在各级语言单位中体现出来。汉语由于“音调已经用来区别词义,不能随便改变”,就“听不出哪个词更重”(端木三,2000),但听不出不代表没有重音,重音是“绝对”存在的。

吴洁敏(2001)经过分析总结出了不同语境中汉语语句重音的几种表现形式:一是扩大音域(增加音高上限)和加大时长;二是用加大音长和音强来表示语义焦点;三是用增加音长、减弱音量来表示语义焦点;四是用增加音高上限和强气流加停顿来表示强调重音;五是用较短的音长加上停顿来表示重音。

王韞佳等(2003)研究将重音分成节奏重音和语义重音进行分析。节奏重音指语句中那些听感上有一定的重度、但并不特别突出、只是为了形成语句在节律上的轻重循环,使语音信号产生节奏感的重音,这种重音对语句意义的表达基本没有贡献。语义重音是在听感上有明显的重度突显,从而使得语句中各信息单元的重要程度得以确定的重音。节奏重音和语义重音能够很明确的感知到,前者倾向于出现在较大的韵律单元内的最末韵律词的末音节上,并且同时有停延,后者的分布则与语句的韵律结构关系不大。

韵律词重音是音系结构中最低层次的重音,它的作用单位是韵律词中的某一个音节。韵律词重音的知觉是听者感觉到一个音节重于另一个音节。关于汉语韵律词重音的声学参数,仲晓波等(2001)研究证明,按照贡献大小的次序,影响普通话韵律词重音感知的声学参数是:时长、音高、频谱倾斜和平均能量。同时也研究了非声学参数因素对普通话韵律词重音知觉的影响,实验结果表明,听者关于普通话韵律词重音方面的语言学知识对韵律词重音的知觉有显著影响。

高明明(1993)对普通话语句中强调重音的声学表现做了深入的研究,她的结论是:(1)基频升高是强调重音的重要声学表现。(2)强调重音的时长普遍加长。(3)音节强调重读时对音强没有明显的影响。

许洁萍(1999)在对一个广播数据库进行了研究之后得到结论,语句重音对音高和时长这两个韵律特征参数都有重要影响,说明语句重音主要是通过音高和音长这两个特征参数的变化来体现出来。在一般情况下,音高和时长是彼此相互独立的,但是在某些情况下,音高和时长表现为正相关。

王韞佳等(2001)对语句层面上的双音节韵律词的重音感知特点进行了研究。实验结果表明,连续话语中双音节词的重音分布受到词所在的韵律边界的显著

影响,和孤立词的重音分布特点有所不同,词内两音节在感知中的轻重对比度与音节的高音点差表现出正相关,与音高差的相关高于与时长的相关。统计分析的结果还表明,音节间的轻重对比度受到调型组合的显著影响。研究还表明,音节的重音感知受到调型的显著影响。

刘亚斌(2003)对比了自然口语和朗读话语的区别,得到了重音在自然口语中的表现。他的研究表明,不同级别重音的音高表现不同,并且同一个音高特征对不同级别重音的表征力不同,同一级别重音在不同位置时的音高表现也不同。重音音节时长的拉长广泛存在于各种声调之中,并且和声母相比,韵母时长的拉长程度更大。对音强的分析表明,音强的均值、范围和高点都是随重音级别增加而增大,而语调短语重音与其它级别重音的差别尤其大;各级重音的音强均值、高点、低点从句首到句尾都呈下降趋势,因此音强的变化受位置影响较大。

王蓓(2002)对汉语重读音节的时长变化模式进行了研究,发现重读程度越高,时长越长;并且,语调短语首的音节重读时时长延长的程度最大,语调短语末的音节重读时时长延长最小;音节在语调短语中的位置对重音的表现有影响。提出了对高音点、低音点的研究,并且得出结论是高音点的提高对重音知觉的作用更为明显。

(3) 停顿研究

停顿是语言进行中的间歇,它也是一种超音段音位,是有声语言表达语意不可少的一种重要的手段。它有调节气息、显示语气、突出重点等作用。合理的停顿,可使话语表意显豁,增加语言的节奏。

曹剑芬(2001)依据句法结构确定出一至四级停顿的出现规律:即一级停顿绝大多数出现在主语和谓语之间,这在比较平衡的句子中尤为典型;二级停顿的分布相对分散;三级停顿跟二级停顿的分布存在一定的交叉;四级停顿的分布范围很广。她认为,语句内部各级停顿的分布与句法结构和词性之间没有始终一致的相关关系。

林茂灿(2002)指出,语句中有两种人们可以感觉到的间断:无声波间断和有声波间断。无声波间断是由停顿(pause)引起的;而有声波间断不是由停顿引起的,而是由间断前的音节时长拉长等所致。无声间断和有声间断之后音节F0(基频)相对于其前面音节都发生了程度不同的跃变。

郑玉玲(2008)的研究表明,韵律词边界闭塞和停顿的区别在于,停顿表现在元音韵尾无过渡音且时长延长,辅音无声段时间较长,则闭塞则不同。语调短语的韵律词边界间是停顿而非闭塞,不存在协同发音现象。

吴洁敏(2001)则提出用“停延”来代替“停顿”的说法,认为语流中音节的疏密度不仅用“停顿”表示,还需用不等音长的延连表示。她把汉语停延分为内部停延和外部停延,外部停延是指由音长、音空形成的语速,内部停延以句子为界分为“内三层”和“外三层”,内三层为:音节<音步<气群<句子,外三层为:句子<句群<段落<篇章。汉语停延时值往往是随着逐级递升的语音层次而不断增长的,句外的语音切分和语言结构一致,句中的语音切分和语法结构不完全对应。同时她也指出,在口语中由于语速较快,各级停延的差异不会像朗读时那么明显。

周明强(2002)指出,句子的停延和句法结构有一定的联系,二者有时会

表现出一致性。但是句子的停延不仅受到句法结构的影响，还会受语义、节律和语境的影响，使句子的停延和句法结构表现出不一致。

（4）韵律层级及边界研究

Selkirk(1984)提出韵律结构包括语调短语(Intonational Phrase)、音系短语(Phonological phrase)、韵律词(Prosodic word)、音步(foot)、和音节(syllable)这样几级单位。Nespor and Vogel(1986)提出韵律结构包括语调短语、音系短语、附加组(clitic group)、韵律词、音步、音节这样几级单位。她的研究是现代韵律结构的基础，在国内外有重要的影响（王茂林，2003）。

王洪君（2000）将韵律单位与层级分为五级，从大到小依次为：语调段、调群段、音步、音节、摩拉。她认为，汉语普通话以双音节音步为主，其次是单音节和三音节。“韵律”指音流中音质之外的音高、音强、音长结构，韵律单位与层级“是指那些由不同的韵律标记定义的大小不同的单位。”语法边界与韵律边界，有时一致，有时不一致，并将不一致的情况形象地定义为“跨界”。调群段由若干音步组成，其韵律标记是“较大的停延和段内声调音域的逐渐收缩和下降”（声调音域指语流中特定段落内声调最高限和最低限的绝对差值，是音高的上限、下限两条线的展敛变化）；语调群的韵律标记也是停顿和声调音域的变化，但停顿的时间比调群长，声调音域的变化模式也与调群不同（沈炯，1985）。

冯胜利（2000）提出了一门新的语言学分支—韵律构词学，将韵律词定义为“一个最小的语流片段/单位”，并将其韵律层级分为：韵律词、音步、音节、韵素(mora)。在此基础上，他根据观察进一步得出了汉语自然音步构成的六条基本规则。同时，他也指出，这些规则只是音步的自然属性，并不意味着音步不受句法、词汇跟语义等因素的限制，在日常口语中也会出现一些违反这些规则的“非自然音步”，这也是很正常的。

曹剑芬（2001）认为连续话语可以切分出韵律词、韵律短语和语调短语三级韵律单位。她对韵律词的界定是：三音节以下的语法词或词组以及其他凡是属于一个音步的结构；跨度为 1-3 个音节；具有类似语法词的连调模式和词重音模式。对韵律短语的界定标准是：由一个或几个韵律词组成，即“可能多音步”；跨度为 7 ± 2 个音节；内部各个韵律词间可能出现韵律上的节奏分界，主要表现为韵律词末尾音节的延长(或伴有较短的无声停顿)和韵律词之间的音高重设(包括音阈和音域)；具有相对稳定的短语语调模式，即音阶的 declination 及 resetting；具有相对稳定的短语重音配置模式，即与句法结构相关的常规重音模式。

林茂灿（2000，2001）认为语句包含韵律词和韵律短语，并指出普通话语句中有两种人们可以感觉到的间断：无声间断和有声间断。韵律短语有大小之分，韵律大短语主要由大的无声停顿和边界前音节时长拉长引起，韵律小短语主要由小的有声停顿和/或无声停顿引起。韵律词有简单韵律词、复杂韵律词和复合韵律词三种类型，简单韵律词指两音节和三音节的词汇词，以及强调念得单音节词如副词、介词等，复杂韵律词有两个以上词汇词组成，简单韵律词和复杂韵律词可以称为“F0 变化组”，复合韵律词则多数是由两个以上简单韵律词和/或复杂韵律词组成。

李爱军（2001）将汉语的韵律结构分为五级，从小到大依次为：音步、韵

律词、次要韵律短语、主要韵律短语和语调组/韵律组。她认为音步是语流中趋于等长的节拍单位，普通话以双音节音步为主，其次是单音节和三音节音步；韵律词一般包括一个或一个以上的语法词，词内音节之间没有停顿，具有一个词重音（韵律上的凸显成分），词后感知到的间断很小，一般没有无声停顿出现，她同意 ShiwookLee 等的观点，“韵律词可以对应一个超音步(suprafoot)，由一个音步和其相邻音步组成，特别是一些未组合的单音节”；次要韵律短语由一个或多个韵律词构成，具有一个短语重音，其后感知到的间断比韵律词大，但比主要韵律短语小，可以是较短的无声停顿或有声停顿；主要韵律短语由一个或多个次要韵律短语构成，具有一个短语重音，其后可以感知到明显的停顿，从声学上看，主要韵律短语后有明显的无声段，与后接短语之间有明显的音高重设，且音高曲线有明显的下倾（declination）；语调组是指由一个或多个主要韵律短语构成且有一个语句重音的语句，其后可感知到的停顿最大。

Tseng（2005）通过对大批口语语料进行处理得到的证据，提出阶层式多短语语流韵律架构（HPG 模型）说，“阶层式多短语语流韵律架构”理论明确说明跨短语之间的韵律关系。每一层韵律单位对于最终的语流韵律的表现都有贡献，而语流韵律是各层级韵律单元贡献度的总和。不同于传统语言学对语流或句调的看法，这个韵律架构把句群（即意义连贯的语段）中的短语或语句视为彼此之间具有亲属关系的韵律成份，而不是孤立、不相关的韵律单元；而阶层式架构表示上层管辖下层，同一层级的韵律单元则是姊妹成分。韵律短语在进入语段甚至语篇中的特定位置时，势必将有所调整以达到语流的最终韵律。对韵律组织而言，阶层性韵律单元和对应的韵律边界及边界信息十分重要，并有其相对应的声学韵律特征。

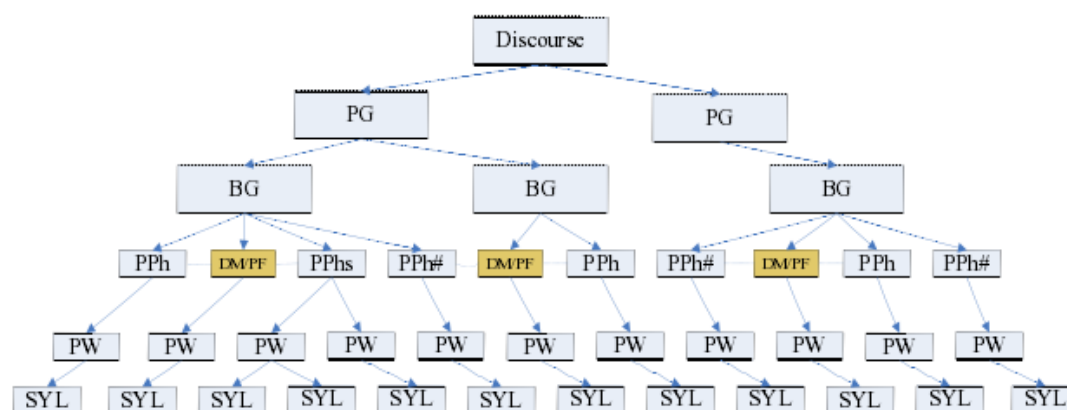


图 1.1 “阶层式多短语韵律句群 HPG” 韵律单位架构图，由下而上分别为音节（SYL），韵律词（PW），韵律短语（PPh），呼吸组（BG），韵律句组（PG）及篇章（Discourse）

为了更科学准确的界定韵律单元，许多学者在寻找韵律边界的声学相关物方面做了大量工作。

冯勇强等（2001）对大规模朗读语料库进行了时长统计，研究了韵律边界、音节的声、韵、调结构以及音节在词中的位置对音节时长的影响，发现边界对音节时长有明显的延长效应；钱瑶（2001）等统计了 116 组语音测试样本，对时长和音高归一化后发现，中间短语和韵律短语边界处的前音节有音延和静音段，韵律短语边界前后音节的音高动态范围有错位现象，静音段长度与前音节

时长是相互补偿的,而静音段长度与后音节音高上、下限线的重置程度是相互增益的;王蓓(2001)等通过实验得出,低音线较好的反映了韵律层级结构;韵律词边界的声学线索是低音线的不连续性和边界前音节的延长,在韵律短语和语调短语的边界处,低音线出现重置;无声段随韵律等级的增加呈对数增长,等级越高增长越快;胡伟湘(2001)等以带有韵律标注的朗读语料库 ASCCD 为基础,研究韵律短语间断模式的时长、基频和音强特征,并提出了一系列声学特征参数,建立了识别分类的决策树模型;熊子瑜(2001)考察了朗读话语中语流间断处的各种韵律表现,发现无声停顿的长短、语音拉长的幅度、基频重设的大小以及边界音调的高低等特征是引起听感上语流间断的知觉线索,并提出了划分各级语流间断的几个标准;Chao Wang(2001)对 JUPITER 电话语料库进行词重音建模实验,分别利用能量(积分、最大值、均值)、基频(最大值、均值、斜率)和时长信息对词重音进行预测,结果发现能量信息的效果最好,时长次之,最后是基频信息。

韵律边界的声学征兆是连续话语的韵律研究中的一个基本问题,是在语流中划分韵律结构的依据。对此,许多学者基于朗读语料进行了深入的研究和探索,得出了一些有用的结论。Colin W. Wightman(1992)研究发现韵律单元末尾音节的拉长是标志韵律边界的一个重要特征,据此可以区分出至少 4 个级别的边界。之后还有不少中外学者也都得出了相似的结论,总的来说,目前公认可以作为朗读话语中边界征兆的因素有停顿、边界前音节拉长、边界处 F0 重设等,并且各因素之间并非互不相干,而是互相影响的。(刘亚斌,2003)。

(5) 韵律模型预测研究

近年来,语音识别、合成技术发展很快,语音合成技术正在影响着现代社会的方方面面。这其中建立韵律预测模型一直都是一个研究重点,韵律结构的预测在 TTS(Text to Speech)系统中发挥的作用越来越大,合成语音的自然度和可懂度与韵律节奏有很大的关联。其中包括对音系层的韵律事件(prosody events)的预测和对声学层的韵律参数(prosody parameters)的预测(初敏,2004)。语音合成首先对于输入的文本,依据外部的词典、规则等,在语言层、语法层和语义层上进行分析,得到文本的层次化信息(包括音标的转写,词组、短语、句子的划分等);然后在层次化信息的基础上在语音层上进行韵律分析,韵律分析的任务就是从输入文本中预测出韵律边界的位置和类型,再送入语音合成部分。得出语音层面上的韵律信息(例如基频、时长、能量参数的预测);然后根据生成的韵律特征,利用参数合成器生成或者从语音库中挑选单元来合成最终的语音数据(凌震华,2008)。而本文的韵律研究也是韵律分析研究工作的一个组成部分。

目前,汉语的韵律结构预测已提出不少方法。应宏、蔡莲红(1999)认为韵律短语扮演着重要的角色,总结了结构助词与韵律短语的关系。其他相关研究有采用基于反馈神经网络(RNN)的学习方法来预测韵律短语(YING,2001),使用基于二元词性文法的统计方法预测汉语的韵律词边界(YAO 等,2001),牛振雨等(2001)中采用三元词性文法做韵律短语预测。除此之外,决策树算法和 Transformation-based learning(TBL)转换规则学习算法等已被普遍应用到韵律结构预测中,主要是通过研究句子语法结构和韵律短语结构之间的关系,从中学习归纳出规则。董宏辉(2007)提出了基于语法约束和长度约束的韵律短语

预测模型，利用隐马尔可夫模型(HMM)对语句中韵律短语的长度规划进行建模(邵鹏飞，2009)。

(6) 呼吸节奏研究 - 生理信号研究

从生理学的角度研究汉语韵律的比较少见。Lieberman (1967) 在对英语进行研究时提出了一个“呼吸群(breath-group)”的理论，认为人们是通过呼吸群来产生和感知语调的，呼吸群分为有标记的和无标记的两种。在无标记的呼吸群中(如陈述句)，随着人们说话时肺部气流的呼出，声门下压力会逐渐下降，这就导致了音高的逐渐下降。在有标记的呼吸群中(如疑问句)，声带肌的紧张度提高，这就抵消了声门下压力逐渐下降所造成的音高下降(谭晶晶，2008)。

陈少松先生(2002)根据自身吟诵体会得出结论，腹式呼吸法已用于吟诵婉约、阴柔的诗文，豪放、激昂慷慨的诗文适宜用胸腹联合式呼吸方法。陈先生还认为在吟诵的过程中要不断的补充气息，不仅要及时“换气”还要适当“偷气”。把“换气”安排在拖音较长的节奏点(平声字)后面。“偷气”主要安排在较高音的节奏单元之后(杨锋，2011)。

谭晶晶(2006, 2007, 2008)对朗读近体诗、词、新闻等文体时的呼吸特点进行了研究。运用呼吸带传感器记录了发音人朗读时的呼吸信号，并自动提取了呼吸信号的重置时长和重置幅度的变化，通过分析发音人朗读不同文体的语料时的呼吸节奏变化，发现呼吸信号的重置幅度比重置时长更能反映不同文体的差别，也更能反映不同呼吸级别的大小。近体诗、词等韵文的呼吸节奏比较简单，其呼吸重置的大小大致可以分成两级。

小说、散文、新闻等非韵文的呼吸节奏较复杂，其呼吸重置可以大致分成三级，如大呼吸群对应于自然段，中呼吸群对应于复句，小呼吸群对应于分句或句子成分。

通过上面的分析，对发音人朗读不同文体的语料时的呼吸节奏变化有了初步的了解，并得到了以下结论：

(1) 从宏观上看不同文体的呼吸重置幅度有一定的一致性：呼吸信号的重置时长和重置幅度之间显著正相关。一级呼吸重置的分布一致性较高，二级呼吸重置和三级呼吸重置之间的界线不是很清楚。

(2) 不同文体的呼吸节奏有较明显的差异。宏观上，近体诗、词等韵文的呼吸节奏比较简单，其呼吸重置的大小大致可以分成两级。小说、散文、新闻等非韵文的呼吸节奏较复杂，其呼吸重置可以大致分成三级。具体到韵文、非韵文内部，不同文体朗读时的呼吸重置时长和幅度也有细微的不同。

(3) 从数量上看，小说、散文、新闻的大呼吸重置(我们将其称之为一级呼吸重置)最少，小呼吸重置(三级呼吸重置)最多。

(4) 呼吸信号的重置幅度比重置时长更能反映不同文体的差别，也更能反映不同呼吸级别的大小(谭晶晶，2008)。

她的研究表明，呼吸节奏可以反映人们对表达内容的整体认知规划，同时它也反映人们生理机能的制约。

张锦玉(2010)研究了10名被试在语篇朗读时的停延率和呼吸特征，语篇朗读时的停延起伏线呈现出不规整的现象，但各级韵律层级边界处的停顿和边界前末音节的延连则有规律可循，停延率也随着韵律边界等级的不同而不同；其

次，语篇朗读时的呼吸参数在不同韵律层级下存在显著差异，且与相应位置的停延呈现出较高的相关性，说明停延和呼吸均受韵律等级的制约，二者呈现出交错对应的关系。

尹基德（2010）对汉语韵律的嗓音发声特点进行了研究。研究表明诗歌朗读中基频和开商为正相关，而基频和速度商为负相关。语流中出现高音调嗓音、挤喉音、气泡音等发声类型，并分析了不同发声类型组合和转变形式跟汉语不同声调之间的关系。在呼吸群边界上基频下降、开商上升、速度商下降，这跟嗓音的基本特性不一致，充分反映出呼吸节奏的生理机理。呼吸群边界的嗓音特征体现为SPLH-SP 值的降低，反映出高频能量的减少。这表明呼吸节奏的嗓音特征不但体现为基频的降低，还体现为若干的嗓音气化现象。

杨锋（2012）专门对中国传统诗歌的吟诵进行了呼吸节奏研究，呼吸与韵律之间的关系。实验结果表明：言语以胸腹联合式呼吸为主，腹呼吸重置时间早于胸呼吸重置时间和语音起始时间，呼气相时长约等于语音时长。胸呼吸主要作用是保证足够的气息量，在发音时胸腔保持扩张状态至发音结束。腹呼吸主要作用是，通过腹肌和膈肌稳健收缩，控制气流持续释放，以获得连续的语音。韵律句起始处对应一个胸腹呼吸重置，韵律短语边界对应胸腹呼吸间断。该研究对言语呼吸生理机制的认识和理解具有重要意义，为言语产生呼吸生理建模提供研究基础（杨锋 等，2013）。

1.2.2 韩语韵律研究

（1）声调和语调研究

众所周知，汉语是声调语言（tone language），即汉语的声调具有区别意义的功能。然而，从韵律研究的角度，有时候在音高曲线（tone contour）中出现的音高特征也被称为声调。虽然这个概念没有区别意义的功能，但是我们可以把它理解为提高言语自然度的韵律特征之一，也是在语调研究中的主要研究对象。因此，我们应该把两种声调区别开：前者为词汇性声调（lexical tone），后者为语音性声调（phonetic tone）或非词汇性声调（non-lexical tone）（安秉燮，2010）。在韩语的声调就属于非词汇性声调或语音性声调，不能辨别词汇的意义。非词汇性声调有时能为形成韵律结构做出贡献，也对形成语调群或整个言语的语调起一定的作用。

在韩语语调研究中，有关语调的研究可以说是比较薄弱的环节。与元辅音以及其他韵律成分相比，关于语调的专门论著也显得较少。对韩语语调问题，许雄（1985）、崔铉培（1971）等早期的学者在他们的著作中做过简单的说明，主要是在句子末尾记上几个符号表示陈述、疑问、祈使等功能。

李炫馥（1965；1974；1976；1982；1984）发表了一系列有关韩语语调的文章，他认为，为了分析韩语语调的调型和功能，应该考虑一下四个因素：（1）音高的高低，（2）音高曲线的变化，（3）音高变化的速度，（4）重音的位置与节奏。他以单音节作为实现语调的最小单位，并把单音节的语调分为音节里没有高低变化的“静态（static）语调”和有高低变化的“动态（kinetic）语调”两种类型。此外，根据音高的高低和音高曲线的变化方向，再把语调分成 20 个小类：音高有高、中、低之分，音高曲线有单向型、双向型和三向型，例如，

双向型有升降调型和降升调型，三向型有降升降调型和升降升调型。李炫馥教授对于语调的分析只包括语调的形态和它所表达的语气，并不包括它的语法功能，而且对句子整体的语调并没有作出充分的描写和解释，但是与早期学者对语调的简单描写相比，他的研究确实使语调研究前进了大一步（宋炫宣，1999）。

在语调类型的研究基础上，李浩英（1996）采用了语调组（intonation group）的概念。他认为韩语句子有一个以上的语调组组成，而语调组则由一个以上的节奏单元（rhythm unit）来组成。节奏单元指的是一个重音音节或者一个在它的前边或后边带一个以上非重音音节的语音单位，这样的语音单位的前边和后边一定会出现句中或句末停顿。节奏单元也可以说是语调单元（intonation unit），语调组可以分为两个部分：一个是核调（nuclear tone）；另一个是短语调（phrasal tone）。核调是语调的核心部分，指的是在一个语调单元的最后音节上出现的音高变化。如果在一个句子中有两个以上的语调组的话，前一个语调组的最后音节上也会出现核调，但它所能传达的语调信息有限。短语调指的是在一个语调组里除了最后音节之外的其他音节中的音高变化。

他对核调和短语调的音高变化进行了分析并归纳语调类型，具体如下。核调有 9 种类型：低平调（Low Level）、中平调（Mid Level）、高平调（High Level）、高降调（High Fall）、低降调（Low Fall）、全升调（Full Rise）、低升调（Low Rise）、降升调（Fall-Rise）和升降调（Rise-Fall）；短语调有 4 种类型：升调（Rising）、平调（Level）、降调（Falling）和升降调（Rise-Falling）。上述的核调类型反映的是句末位置的核调的情况，而语句中的核调只有 5 个类型：低平调、低降调、中平调、低声调及降升调。

根据李浩英的研究结果来看，韩语语调中核调的作用比较大，它是语调信息的最主要载体。例如，陈述句的主要核调有低平调、低降调和升降调，其中低平调没有感情色彩，低降调表达亲切，升降调表达不愉快的情绪。是非问句的主要核调是高平调，当提问的人对某一事实或情况有了某种估计，但又不能完全肯定时，就分别出现低降调和高平调。祈使句的语调中，表示命令时句末用低平调，低降调或升降调，而在表示请求的句式用低声调或降升调。该结果与李淑香（1985）的实验结果也比较一致，韩语语调中最重要的功能表现于句末语调，陈述语调呈现音高线的下降，疑问语调呈现音高线的升高，祈使语调呈现音高线的下降或升降。

Jun（1993，2000）以实验语音学的方式归纳韩语语调类型，现在学界最被广泛接受。韩语的语调可以分为重音短语（Accentual Phrase），语调短语（Intonation Phrase）。重音短语只局限于表示说话者的感情、态度、个人口音，Jun（2000）设定韩语的韵律短语的基本类型是 LHLH，而重音短语的开头音素是送气音（ph、th、kh）、紧音（p*、t*、k*）、擦音（s、s*、h）的时候，韵律短语是 HHLH，不过因为韩语的韵律短语有把基本类型 LHLH 给予每个音节的特点，按音节的个数，实现的类型很多样，在实际上会出现的类型是共 14 个：LH、LHH、LLH、LHLH、HH、HLH、HHLH、LL、HL、LHL、HHL、HLL、LHLL、HHLL。

Jun（2000）还把韩语边界语调短语分为 9 个类型，L%：陈述句；H%：是否疑问句；LH%：疑问句，生气，愤怒，怀疑；HL%：陈述句，疑问词疑问句，在新闻上使用；LHL%：说服，主张，确认，生气的陈述句；HLH%：是 HL%和 H%的组合，确信，期待听者的同意；LHLH%：是 LHL%和 H%的组合；HLHL%：

是 HLH%和 L%的组合,是 HL%的强调,确信自己的意见,偶尔跟 LHL%相似,央求或说服对方; LHLHL%:跟 LHL%相似,强调愤怒。Kim (2009) 也分析两个人的自由对话(spontaneous speech)的 335 个边界语调短语,其中,104 个在句中出现,231 个在句末出现。出现的类型只是 5 个类型: L%、HL%、LHL%、H%、LH%,上升的语调(H%、LH%)出现率是 50.16%,虽然上升的边界语调是疑问句的,但是在两个人的自由对话,陈述句的边界语调也是上升的边界语调。我们可见韩语的边界语调短语也不是固定的,实际对话中会发生变化。

因此, Jung (2002) 指出韩语学习者该学的语调是重音短语的 LHLH 类型和边界语调短语的 L%、H%、LH%、HL%, 4 个基本类型。

Kim (2009) 分析了两个人的自然对话(spontaneous speech),从 541 个韵律短语调查出现最多的重音短语的类型,得出的结果是 LH、LHL、LT(LL、HH)、HL,这 5 个类型的出现率是一共约 64%,另外的各个类型的出现率比较相近。不过,韩语的重音短语不会影响说话者的句子的意思,韩语的边界语调短语也没有区分意思的功能,但是按什么样的边界语调短语,决定句子的类型和表示说话者的感情、态度。

(2) 重音研究

对韩语的重音问题,过去在学者之间存在较大的分歧,大致分为三种不同看法:李崇宁(1960)提出的“强弱重音”、“高低重音”(郑仁燮,1965)和“复合重音”(李炫馥,1973)。李崇宁认为韩语重音取决于音节的强弱,而音节的强弱主要取决于辅音的发音方法。例如,如果音节的首位辅音是紧辅音,该音节就是重音音节。这是主要靠听觉印象判断重音的。后来,当语音学者采用仪器进行实验分析之后,他的主张就全部被否定了。

郑仁燮认为不管单音节词或双音节词都有起辨义作用的高低重音,并且每个词都有固定的重音音节。他几乎把韩语说成是用不同音高区别词义的声调语言。但在首尔标准语中并没有不同音高所产生的最小对立体。郑仁燮虽然用频谱仪做过实验,但他并没有考虑元音固有的声学因素,也没有进行听辨实验,而是紧靠仪器分析的数据下结论。这些因素都使他的实验结果失去了可靠性。

李炫馥认为,韩语重音属于音节的强弱、长短和音色同时起作用的复合重音,其中音长的作用较大。他还把重音和音节结构联系起来用以说明双音节或三音节词的重音规则。例如,如果第一音节结构是(C)VC,并且第二音节的首位辅音不是/h/的话,重音在第一音节。之后,学者们对重音和音节结构之间的关系,作了进一步的研究,并获得了不少成果。

目前,韩国学者们普遍接受的是复合重音说,而后来李炫馥又提出了新的重音规则,主要内容为:1)多音节词里如果出现有长元音音节的话,重音出现在该音节上;2)如果多音节词中没有长元音音节,而第一个音节的结构为(C)V+(C)V(C)的话,重音落在第二音节上;3)如果多音节词中没有长元音音节,而第二个音节的结构为(C)VC+CV(C)的话,重音一般落在第一个音节上。

成哲在(1991,1994)对韩语的正常重音进行了听辨实验和声学分析。研究表明韩语词不管有多少个音节,它的重音都落在第一或第二个音节,他将每个音节分为两种:重音节(Heavy Syllable,如/mal/)和轻音节(Light Syllable,如/ma/),如果第一和第二个音节的结构相同(HH,LL),重音落在第一个音节,

如果其音节结构不同（HL，LH）的话，重音大部分落在重音节上。根据声学分析结果，重音音节的时长比非重音音节更长，强度也会更大。至于基频，所有的实验材料中，都是第二个音节的基频最高。可以说，基频对重音不起决定性作用。该研究在一定程度上揭示了韩语重音的声学特征，但在试验中采用的材料是无意义的音节组合，因而与实际语言有一定的距离（宋炫宣，1999）。

（3）停顿研究

在词典上停顿是“说话时语音上的间歇”。韩国学者对停顿的理解是：说话或者朗读时语音上的间歇的现象，就意味着熟练掌握语法关系以及充分理解文章内容和结构。

李浩英（1996）和金成硅（1999）研究表明，韩语的停顿单位平均包括13-14个音节，或者有四个韵律短语的话，在第三个开始处会出现一个停顿，而汉语的停顿平均包括10个音节（谭晶晶，2008），可见在声学研究中，韩语的停顿单位大于汉语的。

根据李浩英（1991,1996）的韩语语调解释方法，探讨韩语语调。他将语音的节奏单位层级分为语片、语块和句子。自然语流当中出现停延段，最小的停延段是语片，语片是韩语节奏单位（rhythm unit），而且是最小的停延单位，边界段末音节较长。说话当中，由于生理上的原因或者语片内的发声方法不同，有强弱的现象。因此，一个音节重的话，其他音节会弱，有强弱模型和弱强模型。说话的句子有一个或一个以上的意义和信息单位，而且伴随停延显示。

韩语书写的最大特点是每个词语（韵律短语）之间都会出现一个“空格”，根据说话者的意图，比较自由地选择停顿的位置。但是，安秉燮（2007）的研究表明，朗读韩语时，在一般的情况下，在韵律句边界处会出现停顿，而韵律短语边界处无停顿。可以说，韩语的停顿单位为韵律句。

（4）韵律层级及韵律单位

自从1970年以来，韩语韵律层级研究一直由语音韵律学者从不同的角度进行韵律研究。最早从韵律音系学的角度研究韵律结构，后来盛行基于节奏的韵律结构研究。最近学界最认可的模型是，有JUN（1996；2000）提出的基于语调的韩语韵律结构模型，图1.2所示。

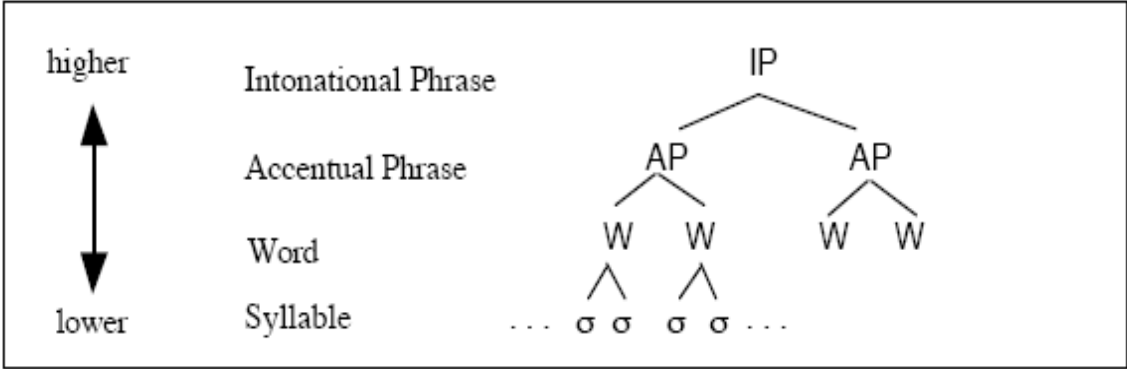


图 1.2 韩语韵律结构（Jun，2000）

随着 80 年代以来美国韵律音系学的语调标记体系 ToBI 的发展,有些学者运用 ToBI 研究韩语语调的特征,如上图 1.2 所示。虽然研究者之间有若干的意见差异,K-ToBI(韩语语调标记体系)定义的韩语韵律单位有重音短语(Accentual Phrase)与语调短语(Intonation Phrase)。重音短语(Accentual Phrase)是以一个以上的音系短语(Phonological word)组成的单位,语调短语(Intonation Phrase)是以一个以上的重音短语(Accentual Phrase)组成的单位。关键是每个语调短语(Intonation Phrase)的边界上出现 6 个边界声调(这并不是说词汇层级的声调,而是说特定的语调模式),则是 L%、H%、LH%、HL%和 LHL%、HLH%(JUN, 2003; 尹基德、孔江平, 2007)。

(5) 呼吸节奏研究

在韩语韵律研究中,从生理学的角度考察语言呼吸节奏的研究几乎没有。只有尹基德和孔江平(2007)曾经对进行初探韩语新闻和散文朗读时呼吸节奏与语调群的关系。结果表明,韩语新闻朗读的呼吸单位准确地对应于每个停顿的地方,且有呼吸下倾和呼吸上升两个常见的呼吸结构模式。

1.3 研究目的和意义

韵律不仅具有辨义功能,而且是影响语音表现力的主要因素,声音的轻重缓急、抑扬顿挫、不同语气和口气的传达、个人语音特征的辨别,都要依靠韵律的作用。因此,韵律研究在当前的语音研究中占有重要位置,深入理解韵律不仅是语音学理论发展的需要,也是语音合成和语音识别发展的需要。

呼吸信号作为韵律研究的新方法和新手段,无论是汉语还是韩语研究,从生理学角度的韵律研究非常少见,汉韩对比研究几乎没有。因此,这是能够为语音学的韵律研究开辟出新的研究课题的,呼吸信号引入语音学的研究对韵律风格和个人风格的研究都会有很大推动(孔江平, 2007),进一步了解结构不同的两种语言的不同呼吸节奏方式和不同认知安排。

谭晶晶(2008)曾使用呼吸信号对不同文体的朗读语料进行了分析,发现不同文体朗读时的呼吸节奏确有区别,不同发音人朗读时的呼吸节奏也有区别,说明呼吸信号是一种反映不同风格的手段。但是由于新闻语料比较复杂,语料的语法结构不确定,难以对其进行定量的声学分析,也就无法揭示呼吸的参数怎样和声学参数联系起来。

本文在前人研究的基础上从呼吸信号和语音信号两个方面来考察汉语和韩语的不同文体中的呼吸重置特征与相关韵律特征,从而探讨

1. 语音共性和不同语言之间的呼吸与韵律特征的差异,为汉语和韩语韵律节奏的深入研究提供参照,并将研究结果运用到语言教学中;

2. 各个韵律单元的呼吸信号和声学参数之间的关系;

3. 以语义相同而句法结构不同的汉韩语料进行对比分析,能够揭示句法结构对呼吸节奏与韵律构造的影响;

4. 不同语言对表达内容的整体规划方式和呼吸节奏的调节。

该研究对言语呼吸生理机制的认识和理解具有重要意义,为言语产生呼吸

生理建模提供研究基础。

1.4 章节安排

本文分为五个章节：

第一章首先介绍了论文的选题背景，概述了研究内容、方法和论文结构。对汉语与韩语不同韵律要素的研究概况，对和现状进行综述，同时分析比较了的异同点。

第二章主要简述了研究方法和研究内容。言语呼吸节奏的生理机制和研究设备以及本实验所使用的程序和参数处理方法。

第三章主要从呼吸和韵律特征的角度，对汉语和韩语诗歌的呼吸重置以及韵律特征进行对比。对得到的数据进行统计分析，并与 计算结果进行比较。

第四章主要对汉语和韩语新闻朗读中的呼吸节奏特征进行分析，探讨二者之间的异同以及语言结构和呼吸节奏之间的关系

第五章主要对本文整个实验研究的结果进行总结，对本研究中存在的问题进行分析，并说明可以继续研究的方向和展望。

第二章 研究方法

本文从语音多模态研究的理念出发，语音多模态研究主要是针对于语音的产生是肺部气流产生动力，气流冲击声带引起震动，经声道和口腔调制而成，是人体多个器官共同作用的结果（杨锋，2012）。下图2.1显示的“言语链（Speech Chain）”也说明整个言语过程是不是分开的，从说话者的认知活动到听话者的感知，包括语言学、生理学、语音学的一系列不同平面的复杂活动。

言语信息的传递从说话人的头脑开始，说话人要做的第一件事就是整理自己的思想决定要说的内容，表达想说的内容。接着，这种语言形式就以神经冲动的形式沿着运动神经传到各个发音器官的肌肉，指挥这些肌肉活动，发音器官的肌肉活动扰动周围空气，就形成了通常所说的声波。声波在说话者和听话者之间传播，进入听话者的耳朵以后，就唤起听话者的听觉机制并引起神经活动，沿着听觉神经向大脑，以某种我们尚未十分了解的方式被听话者感知为相应的语言形式，至此，说和听的言语过程就基本完成了（Denes，1993）。

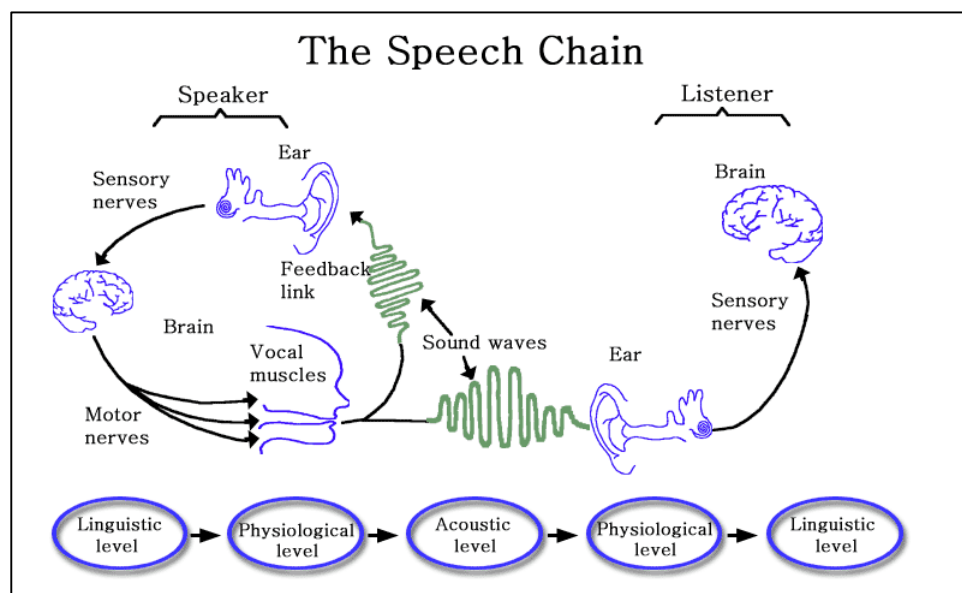


图 2.1 言语链（Speech Chain）示意图（引自 Denes（1993））

因此对语音的研究并不能局限于单纯的语音研究，而是从语音产生的过程：认知-呼吸-发声-调音，多角度多维度进行研究极其重要。其中言语呼吸研究是生理语音学的重要组成部分，同时也为语音多模态研究奠定了基础。

针对语音产生过程的三个环节，呼吸、发声、调音，我们同步采集这三个环节所对应的信号。第一路是朗读者的语音信号，针对语音产生的调音环节。第二路是EGG 嗓音信号，针对语音产生的嗓音发声环节。第三路是胸呼吸信号，第四路是腹呼吸信号，针对语音产生的呼吸环节。采录音频信号主要使用的设备是话筒、调音台，采集第二路嗓音信号所使用的设备是电子声门仪（EGG），采集第三、第四路呼吸信号所使用的是系在胸部和腹部的两根呼吸带，此四路信号通过十六通道采集器输入计算机。通过编写Matlab 程序建立言语呼吸韵律分析平台，提取语音、嗓音、呼吸三方面的参数，试图探讨不同语言的呼吸特

征之间的异同以及不同认知方式对呼吸信号和语音信号的影响。

2.1 实验方法

2.1.1 言语呼吸节奏的生理机制

一般来讲，一直以来呼吸研究大多数与医学和音乐有关，而和语言学的关系不是特别近。良好的气息控制和节奏停顿把握是朗读的前提和基础，研究朗读气流和呼吸的变化可以揭示出朗读气息调节的规律。要研究呼吸规律我们首先要对呼吸机制有所了解。

从剖析角度看，肺与外界环境进行气体交换的过程，叫做肺通气。肺通气是通过呼吸运动实现的。呼吸运动是指胸廓有节律地扩大或缩小。它包括吸气运动和呼气运动两个过程。下图 2.2 给出的是与呼吸相关的肌肉部位。

如上图 2.2 显示，完成呼吸和吸气运动的肌肉是不完全是一致的：

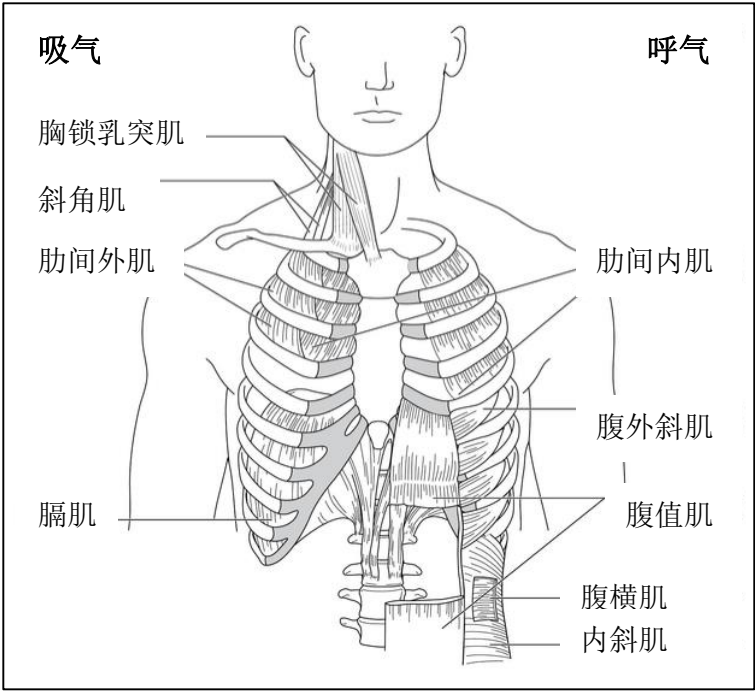


图 2.2 与呼吸相关的肌肉部位

呼吸肌主要包括肋间肌和膈肌。吸气时，除了膈肌、肋间外肌的收缩，胸锁乳突肌、背部肌群、胸部肌群等发生收缩，参与扩张胸廓。呼气时，除了膈肌、肋间外肌的舒张，肋间内肌、腹肌等发生收缩，参与收缩胸廓。

(1) 吸气运动

人在平静呼吸状态下，肋间外肌收缩，肋骨上提，胸骨向上向外移动，使胸廓前后径和左右径都增大；同时，膈肌收缩，膈顶部下降，使胸廓的上下径增大。此时，胸廓扩大，肺扩张，肺的容积增大，肺内气压下降，外界空气通过呼吸道进入肺。完成吸气动作。

(2) 呼气运动

当肋间外肌舒张时，肋骨因重力作用而下降，使胸廓的前后径和左右径都缩小；同时，膈肌舒张，膈顶部回升，使胸廓的上下径缩小。此时，胸廓缩小，肺也随着回缩，肺容积偏小，肺内气压升高，迫使肺泡内的部分气体通过呼吸道排出体外，完成呼气动作，气流呼出的体积流速度和气息量，直接决定着声带振动频率的高低和振幅的大小，继而形成不同的语音基频和能量（高道兴，2007；杨锋 等，2013）。

Baken et al. (1979) 的研究表明，胸腔和横隔膜的变化都对呼吸有贡献，但是它们的运动方式并不完全一样。胸腔和腹腔并不是同时开始运动的，通常腹腔比胸腔提前 100ms 运动，他认为这种“发声前的不同步性（prephonatory asynchrony）”也许和胸腔壁为适应语言需要而作出的调整的方式有关。

现代声乐理论中把呼吸分为下列三种：胸式呼吸、腹式呼吸和胸腹式联合呼吸。歌唱实践的经验表明胸腹式联合呼吸是较完善的呼吸方法，因为它充分的利用了胸部、腹部的体腔容积，灵活应用横隔膜和两肋肌肉的控制能力，所以它吸气和蓄气的量都大，控制力也强，运用气息也较灵活，因此它是合乎歌唱发声要求的科学发声方法（杨锋，2011）。

冯葆富等的研究（1981）表明，安静时的呼吸和说话时的呼吸方式并不相同，安静时呼吸的呼气相和吸气相大致相同，而说话时的呼气相远远大于吸气相；呼吸频率方面，安静时每分钟呼吸 16-20 次，说话时每分钟呼吸 8-10 次；另外，这两种呼吸运用到的肌群也不一样，具体差别如下表所示：

呼吸种类	安静呼吸	言语呼吸
呼吸目的	气体交换	谈话
呼吸控制	非意识完成	受人意识控制
呼吸比值	1: 1.2	1: 5 ~ 1: 8
呼吸次数	16~20 次/分钟	8~10 次/分钟
呼吸量	500~600ml	1000~1500ml
呼吸路径	主要经鼻	主要经口
肌肉动作	吸气时胸部吸气肌群用力，隔略微下移； 呼气时胸部吸气肌群放松，腹部肌微用力帮助膈复位	吸气肌群之中提高肋骨、胸骨、固定锁骨以及板直胸椎的各肌收缩，膈下降较明显； 呼气时，胸部吸气肌群放松，呼气肌群收缩，腹部各肌一起用力，膈上升

表 2.1 安静呼吸与言语呼吸的差别（引自冯葆富等，1981）

Hixon (1987) 的研究表明，胸腔壁（chest wall）是一个由两部分组成的系统，各部分运动的自由度不同，第一部分是胸廓（rib cage），另一部分则是横隔膜（diaphragm），两部分独立运动。

胸廓参与呼吸运动，吸气时胸廓各径均增大，其前后径和横径增大是肋骨和胸骨运动的结果，垂直径的增大是膈肌收缩、膈穹下降的结果。肋骨的运动，吸气时肋颈沿自身长轴向后旋转，肋体上提，并将其前端的胸骨推向前上，肋骨两侧外翻，所以胸廓的前后径、左右径均加大，呼气时做相反方向的运动，

使胸腔容积减少。吸气过程是肋间外肌提起肋骨，导致胸腔增大，膈肌收缩下移，腹肌扩张，胸腔内压力变小，气体进入肺部。呼气过程是肋间内肌收缩及肺的弹性回缩力使得胸腔变小，膈肌上移，腹肌收缩，声门下压力增大，气流被呼出。气流呼出的体积流速度和气息量，直接决定着声带振动频率的高低和振幅的大小，继而形成不同的语音基频和能量（杨锋 等，2013）。

如果把胸腔看作一个容器，那么胸廓是容器壁，横隔膜就是容器的底。呼吸过程中横隔膜的运动如图 2.3 所示，初始状态的横隔膜是穹窿形的，它收缩的时候会变平，这样胸腔的高度就会增加，体积就会增大。同时，横隔膜的变平会导致紧贴在它下方的内脏移位，由于腹腔内没有空气，所以移位的内脏就会导致腹腔壁前方的凸起。横隔膜不能凭它本身的力量恢复到收缩前的穹窿状态，所以必须有其他的力量把它顶回去。这个力量来自于被顶起来的腹腔前壁的肌肉。它们的收缩把内脏推向横隔膜，使其恢复原来的穹窿形状。这样就会再次缩小胸腔的体积。当胸腔扩大的时候，内部气压会减小。如果此时气流通道没有碰到阻碍，那么外部的气流就会由于胸腔内外气压差为负而被吸进肺里。同理，胸腔缩小的时候胸部的气流就会由于胸腔内外气压差为正而被挤出去。

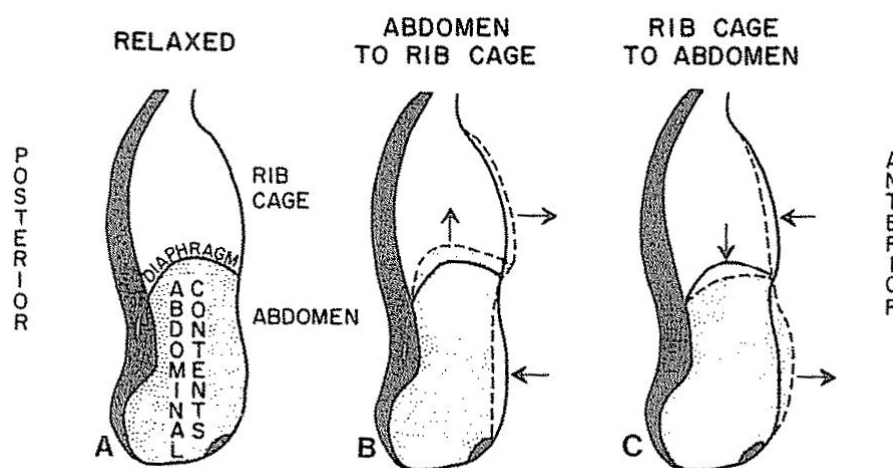


图 2.3 呼吸时胸腔和腹腔的变化（引自 Hixon (1987)）

胸腔体积可以通过胸廓的运动而增大也可以通过横隔膜的运动来增大。由于这两部分是共同作用的，所以肺部体积的变化应该是这两部分体积变化的和。如果胸腔扩大而横隔膜收紧，那么它们都会导致吸气。但是，如果胸腔体积缩小，而同时横隔膜收紧，那么是呼气还是吸气就取决于胸腔缩小的体积是否大于腹腔扩大的体积。因此我们必须明白不能单凭胸腔或腹腔的变化知道此时是吸气还是呼气，在判断是呼气还是吸气之前，我们必须首先了解这两个部位的变化以及它们对于胸腔体积变化的贡献度。

Baken et al. (1979) 的研究表明，胸腔和横隔膜的变化都对呼吸有贡献，但是它们的运动方式并不完全一样。胸腔和腹腔并不是同时开始运动的，通常腹腔比胸腔提前 100ms 运动，他认为这种“发声前的不同步性（prephonatory asynchrony）”也许和胸腔壁为适应语言需要而作出的调整的方式有关。他还认为人们是通过调整腹腔来适应表达词重音的需要的。

2.1.2 研究设备

随着科学技术的进步, 语音学科的前沿已逐渐从语音产生的声学 and 生理平面接近语音产生的生物机制平面, 包括言语产生, 语音病理, 嗓音类型和口鼻腔调音的机理, 大脑感知和情感分析等方面, 即“语音多模态(multi-modal of speech)”的研究。语音多模态研究的发展也反映出了当前科学研究和社会实际应用发展的需要, 已经成为当前语音学研究的热门课题。在理论上, 注重语音产生理论的研究和语音共性的研究。在方法上, 尽可能采用声学、生理、心理学的研究方法, 采集各种声学、生理信号和心理信号。

呼吸信号最早是用于医学病理研究, 近几年引入言语生理研究, 以往对呼吸的研究主要是用气流气压计研究比较微观的语音现象, 如辅音、嗓音等, 北京大学孔江平教授率先采用呼吸绑带采集胸围或腹围的变化来研究汉语普通话语音的韵律特性(谭晶晶, 2007)。

本实验的录音工作于 2011-2012 年间在北京大学中文系的隔音室内进行, 实验中主要使用话筒、调音台、呼吸带、十六通道采集器等设备, 分别录制不同参数, 第一路是发音人的语音信号, 针对语音产生的调音环节。第二路是 EGG 嗓音信号, 针对语音产生的嗓音发声环节。第三路是胸呼吸信号, 第四路是腹呼吸信号, 针对语音产生的呼吸环节。通过 PowerLab 十六通道采集器和 Chart 软件采集为*.adicht 格式的四通道语音文件, 最后使用言语韵律呼吸分析系统进行分析, 提取参数。

首先, 本次录音使用的话筒(麦克风)是 Sony 公司生产的 Sony Electret Condenser Microphone (驻极体电容传声器)。该传声器具有体积小、频率响应比较均匀以及接近全指向性的优点。调音台是德国 Behringer 公司生产的 XENYX502 型调音台, 5 声道, 只使用第 1 声道, Sony 麦克风信号输入, BNC 接口输出, 接入 PowerLab 十六通道采集器第一通道。

本次实验中胸呼吸和腹呼吸的信号通过两条呼吸带传感器和十六通道采集器录入。呼吸带是澳大利亚 AD Instrument 公司生产的 MLT1132 压电呼吸带传感器 (MLT1132 Piezo Respiratory Belt Transducer), 如 2.4 右图所示。呼吸带和该公司生产的生物电信号采集器 (PowerLab) 连接, 与 Chart 软件配合使用采集信号。十六通道采集器型号是 16/30, 如下左图所示, 前置面板是 16 个 BNC 输入端口, 输出端口是 USB 接口, 和电脑连接(杨锋, 2012)。

录音使用的主要设备是澳大利亚 AD Instrument 公司生产的呼吸带传感器, 如下图所示:



图 2.4 十六通道采集器

呼吸带传感器

呼吸带传感器可以测量呼吸导致的胸腹部收缩扩张的变化，本次实验是将呼吸带传感器系在发音人的胸部和腹部由压电设备检测出发音时呼吸带长度的变化，从而获得呼吸节奏的信号，无损伤检测人或动物的呼吸运动波形。这种呼吸带传感器使用方便，可靠，灵敏度高。

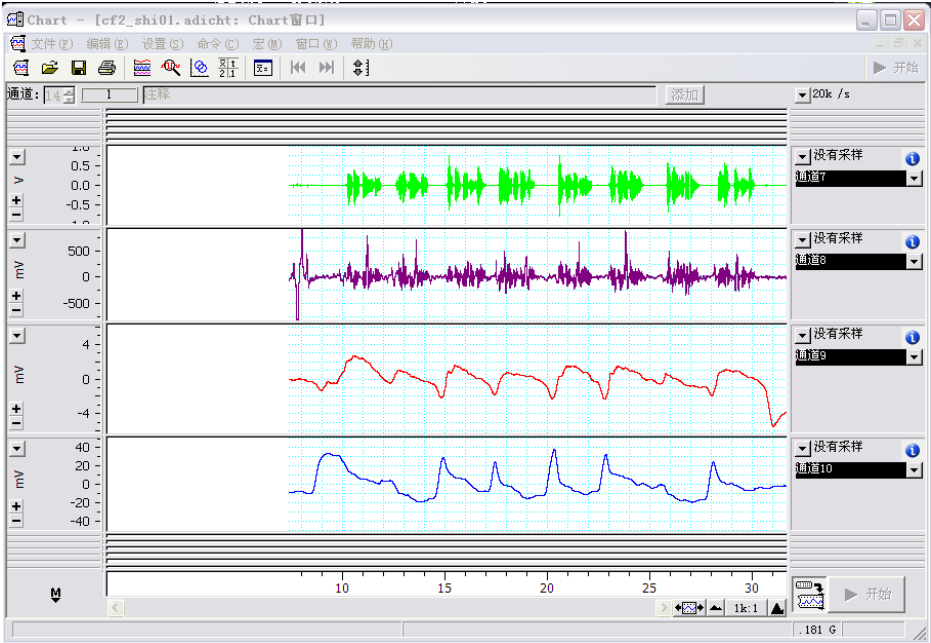


图 2.5 用 Chart5 采集的四通道信号

如图 2.5 所示，通过十六通道采集器和 Chart 软件，把由呼吸导致的电压值变化反映在二维图谱上，横轴是坐标，纵轴是振幅，振幅的变化对应呼吸的变化。呼吸曲线上升表示吸气，下降表示呼气（杨锋，2012）。

实验用十六通道采集器的采集软件 Chart5 采集了 4 个通道的信号：1 通道为用麦克风采集的语音信号，2 通道为用电子声门仪（EGG）采集的噪音信号，3 通道为通过胸呼吸带传感器采集的呼吸信号，4 通道为通过腹呼吸带传感器采集的呼吸信号，采样频率是 20kHz。

2.1.3 参数定义和程序介绍

运用 Chart 5 采集的四通道信号是*.adcht 格式，我们首先将它们转成*.wav 文件，为了便于分析，又将采样频率降为 10KHz。接下来我们运用实验室自行编写的程序自动提取呼吸信号的重置时长和重置幅度。具体办法如下：

(1) 参数定义

我们将呼吸信号的变化反映在二维图谱上，横轴是时间，纵轴是由呼吸导致的电压变化（我们称之为“呼吸曲线”），呼吸曲线上升表示吸气，下降表示呼气。

本实验提取的参数主要是呼吸重置的时长和重置幅度，其中重置时长指的是从开始吸气到开始呼气之间经过的时间，即图 2.6 中波谷和波峰之间的水平距离 T ，重置幅度指的是一次吸气过程中呼吸信号数值的变化幅度，即图 2.6 中波谷和波峰之间的垂直距离 A 。

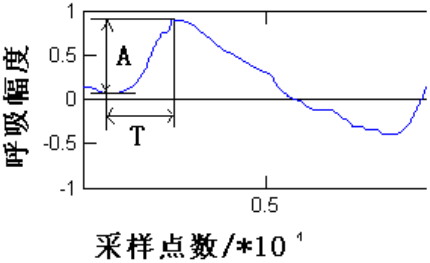


图 2.6 呼吸重置

呼吸信号参数设置我们认为人在朗读之前呼吸均匀，气流顺畅，体内气压与外界大气压一致的状态，信号为零，在读第一句话时会有有一个深的吸气过程，信号从零点攀升到信号峰值点。在朗读新闻的过程中，随着体内气流的减少，气压的降低，信号随之也缓慢下降，直到一个短语或者小句读完，信号降到最谷点，完成第一个呼吸过程。然后重复吸气过程准备后面内容的朗读，直到整篇新闻读完，同时结束整个呼吸过程。每一个完整的呼吸过程，我们称之为每个呼吸基元（Breath Cell），每个呼吸基元都包括一个吸气单元（Inspiration Cell）和一个呼气单元（Expiration Cell）呼吸信号曲线上升一般是吸气过程，一般对应于语音信号的静音段。信号下降表示呼气过程，一般对应于语音信号的语音段将呼吸信号的变化反映在二维图谱上，横轴是时间，纵轴是由呼吸导致的电压变化，呼吸曲线上升表示吸气，下降表示呼气其中重置时长指的是从开始吸气到开始呼气之间经过的时间，即图 2.6 中波谷和波峰之间的水平距离 T ，重置幅度指的是一次吸气过程中呼吸信号数值的变化幅度，即图 2.6 中波谷和波峰之间的垂直距离 A （李永宏 等，2010）。

呼吸信号：由呼吸带传感器所采集到的呼吸变化幅度信号。呼吸带传感器内含压力传感器，可测量呼吸导致的胸腹部收缩扩张的变化。将两条呼吸带传感器分别系在发音人的胸部和腹部，由压电设备检测出发音时呼吸带长度的变化，从而获得胸部和腹部呼吸信号。把由呼吸导致的电压值变化反映在二维图谱上，横轴是时间，纵轴是振幅，振幅的变化对应呼吸的变化。呼吸曲线上升表示吸气，下降表示呼气。

呼吸重置时长：指从开始吸气到开始呼气之间经过的时间。

呼吸重置幅度：指一次吸气过程中呼吸信号数值的变化幅度。

吸气相时间：图 2-7 中从红色虚线到红色实线的时间。

吸气相斜率：图 2-7 中从红色虚线到红色实线区域中蓝色呼吸曲线的斜率，为正值。

呼气相时间：图 2-7 中从红色实线到黑色虚线的时间。

呼气相斜率：图 2-7 中从红色实线到黑色虚线区域中蓝色呼吸曲线的斜率，为负值。

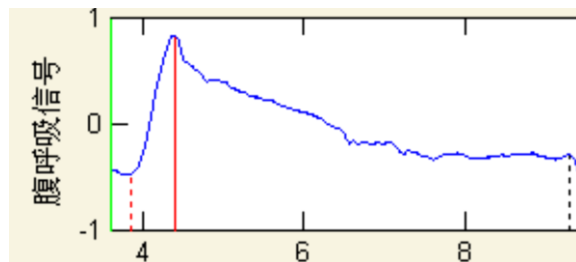


图 2.7 呼吸信号实例

汉语和韩语呼吸信号研究的主要目的在于对二者进行对比，探索呼吸规律，以及不同文体呼吸规则的差异性。换言之也就是研究朗读中不同韵律层级上呼吸重置时长和重置幅度的具体表现和规律。

(2) 程序说明

本次实验采用的“言语呼吸韵律分析系统”用Matlab编程实现。本文以语音多模态的理念进行研究，采录的信号包括语音、EGG信号（本文未涉及分析）、胸呼吸和腹呼吸。但是面临的一个棘手的问题是如何整理分析这些数据。目前的语音分析软件大多只针对语音信号和噪音信号进行分析，而不具备对语音、噪音、胸呼吸、腹呼吸进行同步分析的功能。因此，杨锋（2012）建立了一个能够同步分析语音、噪音、呼吸的软件平台，提取相关参数，以便对吟诵进行多角度深入研究。该平台界面如下图所示：

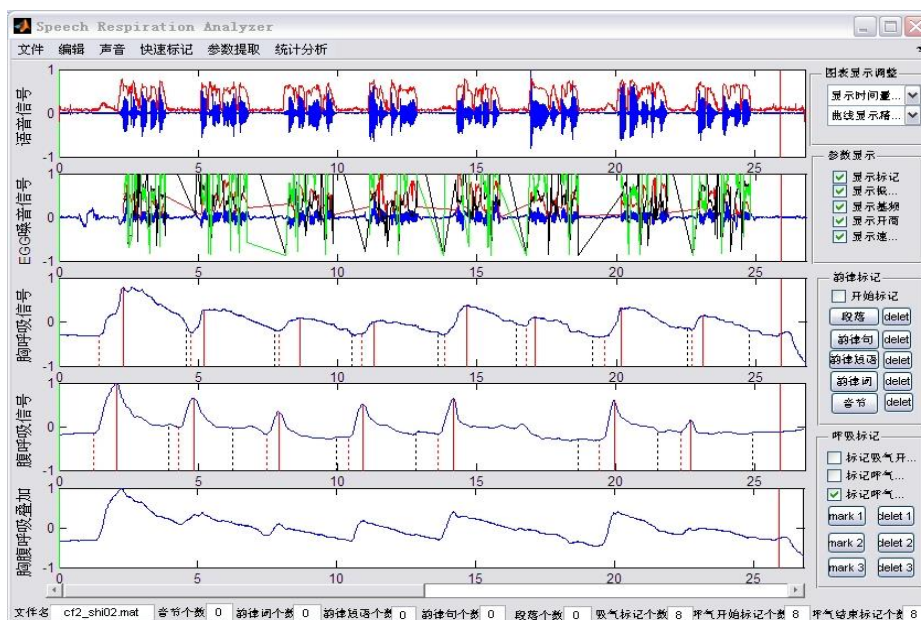


图 2.8 呼吸信号处理平台

本软件为同步研究分析语音、噪音、胸呼吸、腹呼吸四路信号而设计。对韵律层级进行标记，提取音节、音步、韵律短语、韵律句等韵律单元的时长、基频、振幅和韵律边界时长等参数，输出到 Excel 表格中。将各音节归一化后

的基频、开商和速度商参数输出到 Excel 表格中，同时可输出整篇诗词或某一段嗓音信号的原始基频、开商和速度商参数。标记出吸气开始点、呼气开始和结束时间点，由此计算出胸腹呼吸的吸气相和呼气相的重置幅度、时长、斜率、面积等参数，输出到 Excel 中。分析呼吸特征、嗓音发声方式和韵律层级模式，以及三者间的关系。同时也可用于言语工程、音乐、语言教学等领域，同步分析朗读、吟诵、歌唱等语音声学信号、嗓音信号和胸腹呼吸信号，提取各项参数进行统计分析（杨锋，2012）。

第一窗口显示语音信号，第二窗口显示嗓音信号（本文未涉及分析），第三窗口显示胸呼吸信号，第四窗口显示腹呼吸信号，第五窗口是对胸腹呼吸经过计算后的信号。红色虚线是吸气开始时间点，红色实线是呼气开始时间点，黑色虚线是呼气结束时间点。程序首先需要标记吸气开始点、呼气开始和结束时间点，由此计算出胸腹呼吸的吸气相和呼气相的重置幅度、时长、斜率、面积等参数，输出到 Excel 中。

韵律标记显示区用于韵律层级标注，首先选择“开始标记”，然后用鼠标左右键选中某段语音信号，再点击“音节”或其他按键即可显示韵律标记线条。若要删除韵律标记，先鼠标选中要删除的线条，点击对应的“delet”按键即可删除。呼吸标记区用于对胸腹呼吸信号进行呼吸重置标注，“标记呼吸”和“标记吸气”互斥，同一时刻只能选择一种操作。首先用鼠标左键点击呼吸重置所在的位置，点击对应的“mark”按键，系统自动检测在 0.1 秒区域内最大点或最小点所在的位置，然后显示呼吸重置线条并记录下所在位置。若删除则先用鼠标选中线条，然后点击对应的“delet”按键即可删除。

在第一窗口上，我们对语音信号首先进行音节标记，根据音节间停顿时长的聚类分析结果划分音步、韵律短语、韵律句等韵律单元，输出各韵律单元时长、基频、振幅和韵律边界时长等参数到 Excel 表格。如图 2-8 所示，蓝色是语音波形，红色曲线是振幅。

在第二窗口上，通过对 EGG 嗓音信号的分析探析吟诵和朗读在发声方式上的差异性，用到的主要参数是基频、开商和速度商。开商、速度商这两个参数概念是声源模型中的两个重要概念，希望在后续研究中可以进一步涉及到嗓音和呼吸韵律之间的关系汉语和韩语之间的差异。

在第三，四和五窗口上，程序首先需要标记吸气开始点、呼气开始和结束时间点，由此计算出胸腹呼吸的吸气相和呼气相的重置幅度、时长、斜率、面积等参数，输出到 Excel 中。

2.1.4 参数提取和处理

通过 PowerLab 采集器和 Chart 软件采集为*.adicht 格式的文件，然后转换为*.wav 格式的语音文件，最后读入到言语呼吸韵律分析系统中，对韵律层级、胸腹呼吸重置进行标记，提取韵律层级参数、胸腹呼吸参数和嗓音参数，进行统计分析。对第一通道吟诵语音信号提取各韵律单元的时长、强度、无声段时长、基频等参数，建立不同文体吟诵的韵律层级模型和特征库。对第三和第四

通道的呼吸信号提取出呼吸重置级数、幅度、时长等参数。

1. 考虑到朗读不同文体时呼吸信号幅度和时长的可比性，首先对第 3 通道的呼吸信号进行归一化处理。

2. 对归一化后的呼吸信号进行平滑滤波。

由于采集的原始呼吸信号带有很多细微的高频和部分干扰信号，所以对归一化后的呼吸信号进行了低通平滑滤波，滤波器采用零相位数字滤波 `filtfilt(b,a,x)`，通过将输入数据前向和反向处理，以完成零相位数字滤波。它先将数据按顺序滤波，然后将所得结果逆转后反向通过滤波器，这样得到的序列为精确零相位失真，并使滤波器的阶数加倍。`Filtfilt` 通过与初始条件相匹配。可使起始和结束阶段的暂态过渡过程最短。

3. 自动标记波峰和波谷。对平滑后的呼吸信号，利用局部最大法，综合考虑呼吸重置峰值点之间的距离和幅度差、谷值点之间的距离和幅度差，多次迭代循环，尽量消除小的类似于呼吸重置的峰和谷，最终获得能满足要求的呼吸重置的峰值和谷值。程序中还设置了手动修改功能，如果标记位置有误，可进行手工调整。

4. 手动标记波峰和波谷。经过对呼吸信号的大量研究，我们大致根据呼吸重置的幅度大小和时长，确定出正常人说话，一般有 3 级呼吸，程序中可以分别对每一级别的呼吸重置和呼吸单元进行标记，也可以作为自动标注的补充手段。

5. 参数保存：标记打完后，可以按实验室自己的格式保存为带参数的四通道 wav 文件，所有的标记参数都保留在原始 wav 文件的后面，参数按国际 RIFF 标准定义。

6. 要提取呼吸参数的 wav 文件进行批处理，程序中针对手动和自动标记分别做了两套自动提取参数的程序，能够直接把文件夹下面的所有 wav 文件的标记数据直接读取到 xls 文件中。

根据自动提取的呼吸重置时长和重置幅度，我们可以得到不同文体朗读时的呼吸节奏的差别。接下来我们将具体探讨每种文体的呼吸曲线的特点，由于时间限制，本次实验将主要探讨新闻语料的呼吸特点。在讨论呼吸特点的时候，我们主要考查的是呼吸单元的划分及其与语义完整性的关系，呼吸重置出现的位置，呼吸曲线的构造特点等等。

最后，我们将 4 通道的“*.wav”文件转为单通道“*.wav”文件，运用 Praat 对语音信号进行声学分析。我们先根据呼吸重置幅度的频度分布，确定了判断各级呼吸边界的办法，然后测量各个呼吸边界的声学相关物。我们测量的参数包括呼吸边界前、后音节的基频高音点和低音点的值，呼吸边界前后音节的音节时长以及呼吸边界无声段的时长。得到各种声学参数之后，我们将运用 SPSS 软件对这些数据进行统计分析（谭晶晶，2008）。

2.2 研究内容

2.2.1 实验语料

本研究采用语料时，我们首先考虑到韵文与非韵文之分。不同问题可能会引起朗读风格上的差异，因此我们选择诗歌与新闻两类。

诗歌部分没有选择白话诗，而是选择了格式严整，又能充分体现中国特色的古典诗歌，即近体诗和词。近体诗即格律诗，典型的格律诗有律诗（四联八小句）和绝句（两联四小句）两种，每种又根据小句字数的多少分为五言和七言，本语料库共选近体诗20首，五言绝句、五言律诗、七言绝句、七言律诗各5首。

新闻由于其写作具有一定的模式，不带明显的感情色彩，新闻语音文本材料容易获得等原因，在以往韵律研究中采用的也很多。本语料库包含了 20 篇独立的新闻，每段新闻的字数也在 200-250 字左右。

至于诗歌中的韩语汉字版语料，我们先要了解韩国的汉字使用情况。

谈到韩国的文字生活，必然要说到汉字。众所周知，在 1443 年韩国固有的文字“韩文（한글, Hangeul）”创制之前，韩国就从中国借用了汉字，汉字曾经在一段时间里成为记录韩语的手段，在很长时间里被用做韩国的文字语言（李翊燮 等，2008）。汉字词作为一种借词，大量融入到韩语中。时至今日，它在韩语词汇中仍占有很大的比重，尤其是在书面形式里。

在韩国所使用的汉字借字标记法中，汉语诗歌采用的是“口诀”。它是指朗读汉文时，为了了解意思或便于诵读，而在每个句子下方记录的语法要素的统称，可以写为诸如“隐（은,는）”，“伊（이）”之类的汉字。即汉文段落中，在汉字的音训后加助词，只是汉文加助词的口诀文，也是韩国特有的文体。如果除去口诀文字，就会成为完整的汉语文章，在本文中的则是汉语诗歌（黄京洙，2009）

汉文和汉字在韩语体系和文化中占有特殊的地位，迄今为止韩国学生在课堂上仍然以口诀的方式来第一次尝试阅读汉语的诗歌。因此，本文为了将不同语言结构和呼吸韵律之间的关系加以明确，并考察汉语声调和汉字版音调之间有无关系，而添加了汉字版的语料进行对比。具体语料内容参见附录。

2.2.2 发音人

本实验汉韩发音人各选男女四名，共选择了八名年龄范围为 25-35 岁的发音人，他们语音纯正，普通话标准，有较丰富的播音或朗读经验，在朗读速度以及调整情感方面上也非常稳定，便于除去韵律的相对性和可变性特征。一般来讲，女声的材料比较适合分析韵律特征。因为相对而言女声的音高变化比较活跃，音高曲线（pitch contour）的起伏变化比较明显。因此，本文主要以女声的数据结果来考察汉语和韩语的呼吸节奏特征之间的关系。

下面我们对诗歌和新闻等不同形式和汉韩不同语言进行了生理及声学对比分析，试图探讨以不同语言朗读不同形式的文体时呼吸节奏特征的异同，进一步寻找不同语言和生理信号的关系，下面我们按文体来逐一介绍。

第三章 汉语和韩语诗歌的呼吸与韵律特征对比

本章我们将首先通过汉语和韩语诗歌的呼吸信号和语音信号，分析出在呼吸节奏特征上，汉语的诗歌、韩语汉字音诗歌以及翻译成韩语的诗歌之间有什么异同，从而探讨不同语言结构和呼吸韵律特征之间的关系。

综合中国诸多学者的研究观点，他们认为诗句中两字三字是汉语诗歌的最小协奏单位，相当于“音步”。汉语的节奏模式不同于古希腊诗歌的长短型节奏模式，也不同于英语中的中阴界节奏模式。葛晓音（2006）认为：“所谓五言句节奏，指能够形成二三诵读节奏的句式，这是自明清到当代研究者公认的五言诗句的典型节奏。也就是说，一个典型的五言句应当是由一个双音节词加一个三音节词构成。”。下图 3.1 是初二学生的《汉文》教科书中的汉语诗歌选读部分，图中两首近体诗以及里边的笔记也告诉我们汉语近体诗的格律特征。

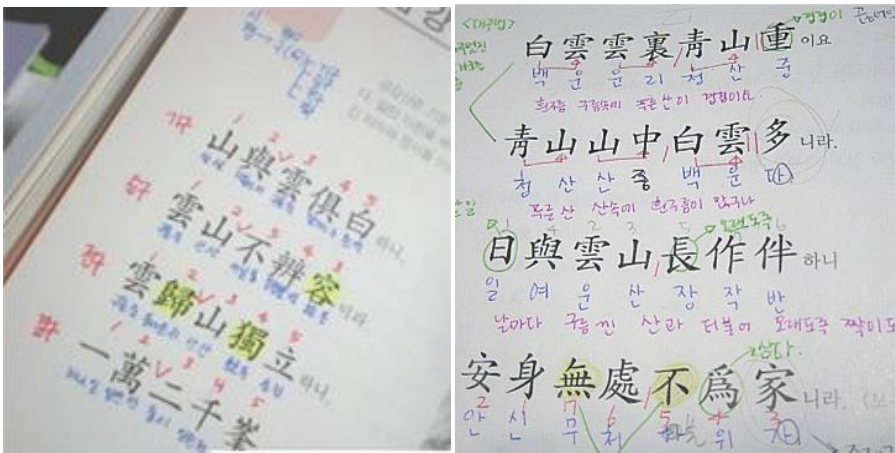


图 3.1 《汉文》教科书中的汉语诗歌选读部分

我们根据前人的研究暂且把五言近体诗的韵律层级划分为诗行和顿两级，把七言近体诗的韵律层级划分为诗行、逗、顿三级，分别对应到语言学韵律层级中的韵律句、韵律短语、音步，如表 3.1 所示。

	节奏单位	韵律单位	长度（字数）	韵律边界
五言近体诗	顿	音步	两字或三字	短时停顿
	诗行	韵律句	五字	长时停顿
七言近体诗	顿	音步	两字或三字	短时停顿
	逗	韵律短语	四字或句末三字	较长停顿
	诗行	韵律句	七字	长时停顿

表 3.1 近体诗节奏和韵律层级划分（杨锋，2012）

格律是根据民族的先验文化体系形成的，即母语者按照自己民族的呼吸以及生物节奏形成的格律来朗读诗歌，由此得到格律美的快感。因此以每个国家的格律体系来研究自己的诗歌，不能忽略不同诗歌固有的特征和特殊性。

韩国诗歌的“格律”也不同于汉语诗歌的特征。格律形式主要分为四大类：音声律、音数律、音步律和音位律。

其中在韩语诗歌中普遍采用的是“音步律”，这是一种倾向于以呼吸群为一个单位来形成格律的形式。音数律的问题被提出之后，韩国学术界广泛接受“音步律”。“音步律”的形式大致分为两类：一种是看每行由几个音步（foot）组成；另一种是由音节时间的等长性（equal length）来形成一个单位，这种形式比较适合像韩国现代诗一样每行音节数不规则的情况（赵东日，1982）。例如，

얇은 사(紗) / 하이얀 고깔은 // 고이 접어서 / 나빌레라.
파르라니 / 깎은 머리 // 박사(薄紗) 고깔에 / 감추오고,
두 볼에 / 흐르는 빛이 // 정작으로 고와서 / 서러워라.
- 조지훈, '승무'에서

这首诗只看音节数难以判断格律特征，而以朗读时间的等长性来看，我们可以知道这首诗是每行有四个音步的音步律。根据韩语诗歌的韵律结构下面把韩语近体诗的韵律层级划分为两级，分别对应为音步和韵律句。韩语诗歌中的“音步”一般由 3-4 个音节组成的，但是由于韩语本身的结构，其界限比较模糊（崔贤识，2013）。汉语音步形式是“二常规、一三可容、四受限”（初敏，2003、2004；王洪君，2010），以音节长度来划分。而韩语则根据空格单位和意义来划分韵律单位。其原因是汉语一个音节几乎有一个意义，单音节是音义结合体，但韩语的单音节不一定是音义结合体。根据韩语诗歌的格律，将韩语诗歌的韵律层级划分成音步和韵律句二级，如下表 3.2 所示。

	韵律单位	字数	韵律边界
韩语 诗歌	音步	四个字 (2-7 个字)	短时停顿
	韵律句	八个字 (6-10 个字)	长时停顿

表 3.2 韩语诗歌节奏和韵律层级划分（赵东日，1996）

本文从已有的汉语和韩语近体诗韵律研究的基础上，将通过汉语近体诗呼吸信号和语音信号的分析，探讨不同语言的呼吸节奏与近体诗韵律特征之间的关系。

3.1 汉语诗歌的呼吸与韵律特征

3.1.1 汉语诗歌的呼吸节奏特征

本研究选取了五言绝句、七言绝句、五言律诗和七言律诗各五首总共二十首，由于近体诗的格律严格，结构简单，即和其他文体相比，近体诗的呼吸信号类型比较简单（谭晶晶，2008；杨锋，2012）。

我们从下图 3.2 显示的两种呼吸形式可以大致了解到汉语诗歌朗读中的胸腹呼吸特征：如图所示，红色曲线是胸呼吸，蓝色曲线是腹呼吸，灰色是语音信号，黑色虚线指的是每小句语音信号的开始点和结束点。在大多数情况下，每一联开始处都有一个较大的呼吸重置，每小句开始处也有一个呼吸重置，小句开始处的呼吸重置比每一联开始处的呼吸重置小。即大呼吸和小呼吸信号交替出现。胸腹呼吸早于语音开始时间，而结束时间比较一致。可见，汉语近体诗朗读中出现的基本呼吸类型与谭晶晶（2008）和杨锋（2012）等的分析结果基本一致。

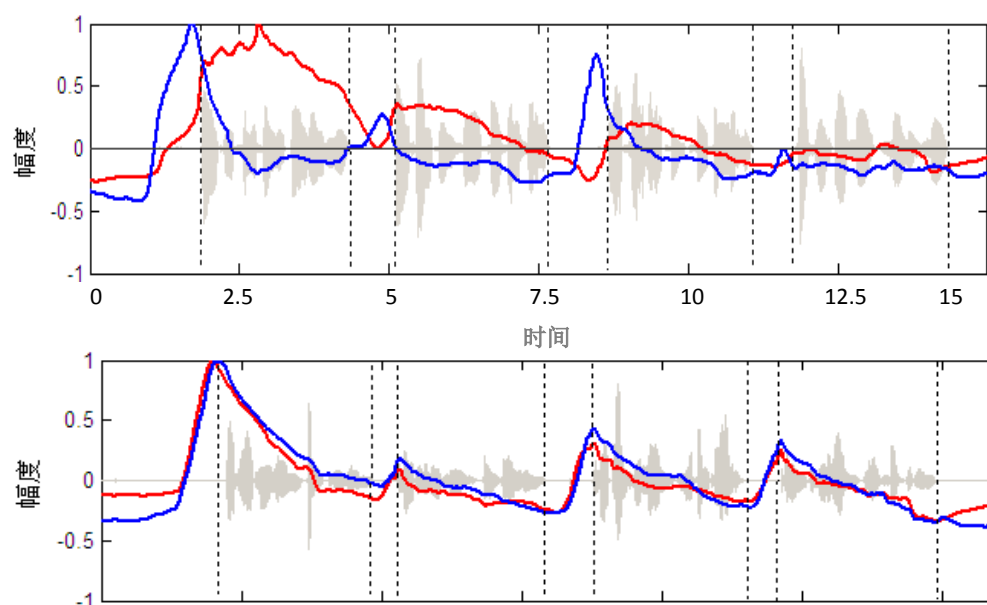


图 3.2 汉语诗歌朗读中的胸腹呼吸与语音波形实例

不同的是，朗读时出现的胸腹呼吸类型不仅仅是一种固定形式。上图显示的是诗歌朗读中出现的两种比较典型的呼吸方式：

一种是如图 3.2 的上图所示，腹呼吸比胸呼吸提前开始，为了发出声音准备足够强的力量进行短促而极强的呼吸，因而其呼吸斜率也比较急促，波峰也较尖。此时，胸呼吸跟随腹呼吸在胸腔吸气到最大值开始呼气的时候胸呼吸曲线才到最高点，接着腹呼吸到了波谷，胸腔保持一段时间的扩张状态，然后缓慢地呼气直到语音结束。

另一种是胸腹呼吸几乎同时开始，如图 3.2 中的下图的胸腹呼吸曲线所示，在朗读的整个过程中呼气和吸气的表现比较相似，两个曲线比较一致。胸腹呼吸几乎同时到达波峰，其方式跟第一种呼吸类型比较一致，而此类型在呼气时

胸腹腔都较缓慢地收缩进行呼气直到语音结束。由此我们可以知道，呼吸方式的个人差异有可能在于胸呼吸的不同，而腹呼吸方式相对比较固定，变化程度不大。从上述的两种形式可以看出，虽然二者的胸腹呼吸配合方式不同，但是他们的整体节奏和呼吸重置次数是一致的。即朗读者基于近体诗的基本格律，按照个人呼吸方式和文体的特征而调整胸腹呼吸的幅度和长度以及配合方式。下面从这几个方面来详细介绍汉语近体诗的呼吸节奏特征。

(1) 汉语五言近体诗呼吸特征

通过上文图 3. 的描写，我们大体了解汉语近体诗的呼吸形式。为了探讨大小不同的呼吸重置幅度的频度分布，我们提取 20 首汉语近体诗中所有的呼吸重置幅度，在 SPSS 中将他们画出直方图，如下图 3.3 所示。

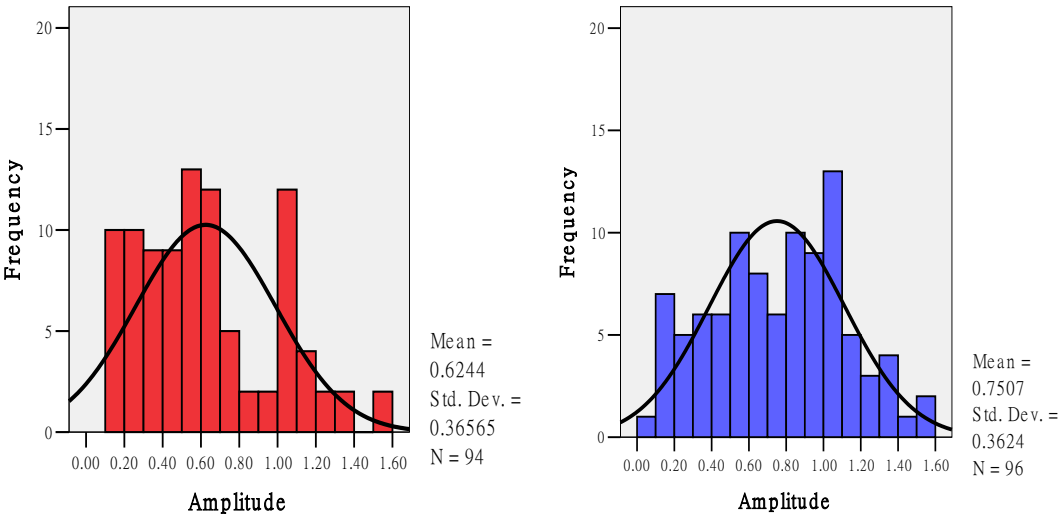


图 3.3 汉语五言近体诗呼吸重置分布图
(左图：胸呼吸，右图：腹呼吸)

汉语五言近体诗的胸腹呼吸重置频度分布基本符合正态分布，腹呼吸重置幅度的平均值比胸呼吸的稍微大一些，胸呼吸重置幅度的平均值为 0.62，腹呼吸平均值为 0.75，胸腹呼吸重置次数分别为胸呼吸 94 次（每首律诗平均呼吸次数为 6.5 次、绝句平均为 3 次）、腹呼吸 96 次（每首律诗平均为 6.4 次、绝句平均为 3.3 次），就是说汉语五言近体诗的每诗行基本出现一次呼吸，胸腹呼吸出现的次数基本一致，腹呼吸稍微多一点。这说明朗读汉语五言近体诗时，如上图发音人用腹呼吸时的重置幅度大，以给整个言语活动力量，紧接着胸呼吸要开始，用这股力量持续一段时间而能够保持言语活动。

在次数上二者基本没有差别，只是在言语活动中的功能不同，即在朗读五言近体诗时胸腹呼吸的配合是有一定规则的，与之前的研究结果（杨锋，2011 等）比较一致。

从上图 3.3.看，无论胸呼吸还是腹呼吸重置幅度在 1.0-1.1 和 0.5-0.6 处分别出现了两个比较明显的峰值，这说明汉语五言近体诗的呼吸重置大致可以分成两级。不同的是胸呼吸的重置幅度主要集中在小峰值周围，两个频度最高的峰值的频度差不多，与它相比，腹呼吸的重置幅度更接近典型的正态分布，大峰

的频度最多而大峰和小峰周围的峰值也比较均匀。

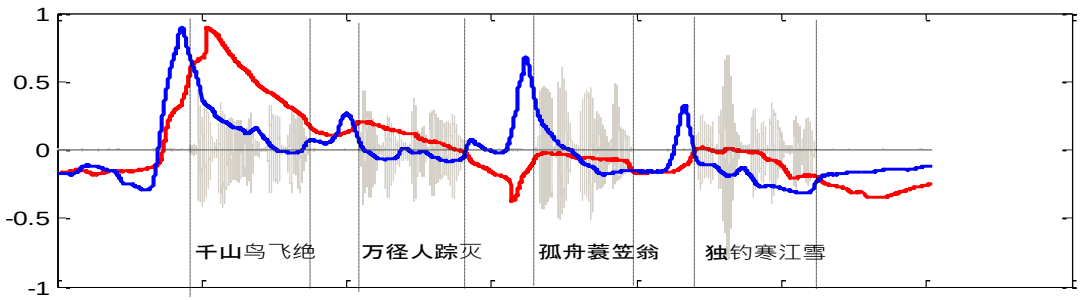


图 3.4 汉语五言绝句朗读中胸、腹呼吸与语音波形

江雪

千山鸟飞绝，万径人踪灭。
孤舟蓑笠翁，独钓寒江雪。

上图 3.4 显示，汉语五言近体诗朗读中胸、腹呼吸重置一般分为两级。上图中的红色曲线显示胸呼吸重置变化，一个大呼吸后出现几个小呼吸，下图显示的蓝色曲线则是腹呼吸变化曲线，一般大呼吸小呼吸交替出现，黑虚线是每小句语音信号的始点和终点。

从方式上看，朗读中在大多数情况下，每联开始处具有一个大呼吸重置，每小句的开始处具有一个二级呼吸重置，该特征和前研究（谭晶晶，2008；杨锋，2012）比较一致。不同的是胸呼吸重置方式比腹呼吸的自由一点，除了第一联开始处出现大呼吸之外，其他地方出现的胸呼吸重置幅度比较自由。

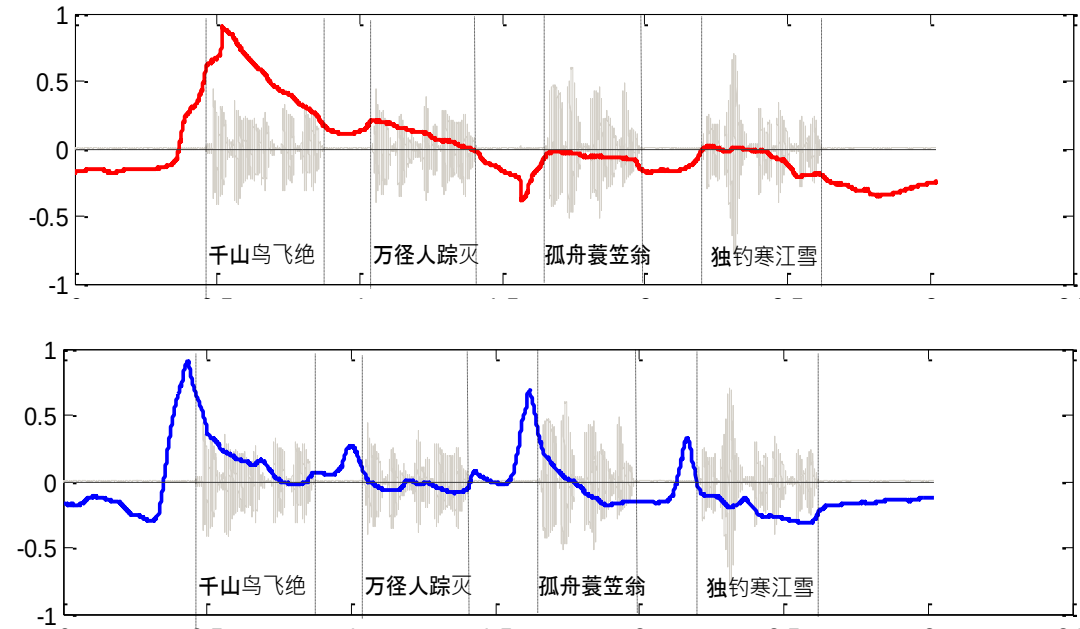


图 3.5 汉语五言绝句朗读中胸、腹呼吸与语音波形
(上图：胸呼吸与语音波形，下图：腹呼吸与语音波形)

如上图 3.5 所示，汉语五言近体诗朗读中胸呼吸重置分为两级，通常首句具有一个较大的呼吸重置，而第三句开始处并不一定每次出现大呼吸，第二、

四句具有一个较小的胸呼吸重置。从图中可看出胸呼吸的呼气开始点与诗句语音开始时间吻合，同时在诗句中显示持平或者缓慢下降的趋势，在诗句语音结束位置胸呼吸曲线会下降。

汉语五言近体诗朗读中腹呼吸重置时间点与峰值所在位置均早于语音开始时间，每联具有一个大呼吸重置，第二与第四句具有小呼吸重置。如图 3- 所示，“千山鸟飞绝”和“孤舟蓑笠翁”之前各具有一个大呼吸重置，“万径人踪灭”和“独钓寒江雪”之前各具有一个小呼吸重置。同时在诗句中间韵律短语边界处腹呼吸曲线具有一个抖动或者小呼吸。如“千山”与“鸟飞绝”之间腹呼吸曲线有一个明显抖动或小呼吸，其他三句也具有此特征。

从数据上看，朗读中胸呼吸的一级大呼吸重置幅度平均为 1.1，斜率为-0.3，二级小呼吸重置幅度平均为 0.4，斜率为-0.2，第一联开始处气息强，而其他小句开始处气息则弱，也较为舒缓。腹呼吸的一级大呼吸重置幅度平均为 1.0，斜率为-0.3，二级小呼吸重置幅度平均为 0.4，斜率为-0.2，如上图 3.4 所示，每联开始处气息强，呼吸曲线又高又尖，流速快，而每小句开始处气息则弱，高峰依然很尖。

我们运用 SPSS 对呼吸重置时长和重置幅度值作了相关分析，得到如下结果其显著性水平为 0.01，也就是说重置时长和重置幅度之间的正相关性非常显著。

Correlations

		胸呼吸重置幅度 重置时长	腹呼吸重置幅度 重置时长
胸呼吸重置幅度 重置时长	Pearson Correlation	1	.504**
	Sig. (2-tailed)		.000
	N	94	94
腹呼吸重置幅度 重置时长	Pearson Correlation	.504**	1
	Sig. (2-tailed)	.000	
	N	94	94

** . Correlation is significant at the 0.01 level (2-tailed).

表 3.3 汉语五言近体诗呼吸幅度与时长相关

从功能上看，腹呼吸比胸呼吸来得早振幅也大，但是比胸呼吸结束得快，在语音信号结束前已经开始准备下一个呼吸。可以说，腹呼吸先给言语活动提供较大的能量，另外腹呼吸也会出现在音步和音步之间，可以说对韵律单位比较敏感。相比来说，胸呼吸比腹呼吸来得晚，而语音信号结束点和胸呼吸呼气结束点比较一致，即胸呼吸是保持言语活动的能源。

下面显示的聚类分析（Cluster Analysis）结果更加明确地告诉我们胸腹呼吸重置幅度表现的不同。“聚类分析”是一组将研究对象分为相对同质的群组的统计分析技术。

因为只观察呼吸重置波形图和分布图而将不同大小的呼吸分别归为不同级别还是有较大的难度，而且必然带有很大的主观性。因此，本文对每个呼吸重置进行聚类分析而分成若干类，以保证数据分析的客观性并解释朗读不同形式文体中的不同呼吸级别特征。

开始位置，二级呼吸处于每长句的开始位置，因此，二者的呼吸重置比较有规律地交替出现，其频度也比较接近。与此相比，胸呼吸一级呼吸处于每首诗的开始位置，不一定每联都出现，经常由二级呼吸重置替代。结果，二级呼吸重置次数比一级呼吸的多一倍以上。可见，朗读汉语五言近体诗中腹呼吸重置更有规律，胸呼吸重置方式除了每联开始部分之外，比较自由地调节呼吸重置幅度。韵律边界处出现呼吸中段抖动，腹呼吸早于胸呼吸。

(2) 汉语七言近体诗呼吸特征

从下图 3. 6 看出，汉语七言近体诗朗读中胸呼吸分为两级，而腹呼吸重置幅度比较复杂。下图中的红色曲线显示胸呼吸重置变化，蓝色曲线则是腹呼吸变化曲线，黑色虚线是每小句语音信号的始点和终点。

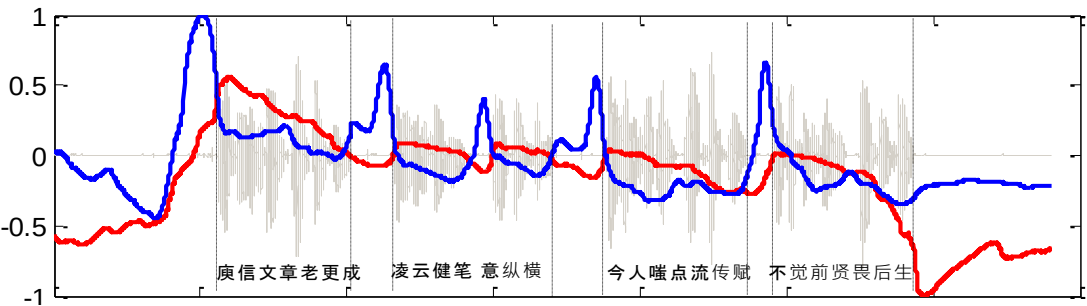


图 3.6 汉语七言绝句朗读中胸、腹呼吸与语音波形

从方式上看，朗读中在大多数情况下，每小句开始处具有一个大呼吸重置，随后每小句的开始处具有一个二级呼吸重置，该特征和前研究（谭晶晶，2008；杨锋，2012）比较一致。不同的是胸呼吸重置方式比腹呼吸的自由一点，除了第一联开始处出现大呼吸之外，其他地方出现的胸呼吸重置幅度比较自由。

一般大呼吸小呼吸呼吸交替出现，与五言近体诗的呼吸方式不同的是小呼吸出现的地方还会出现中呼吸。

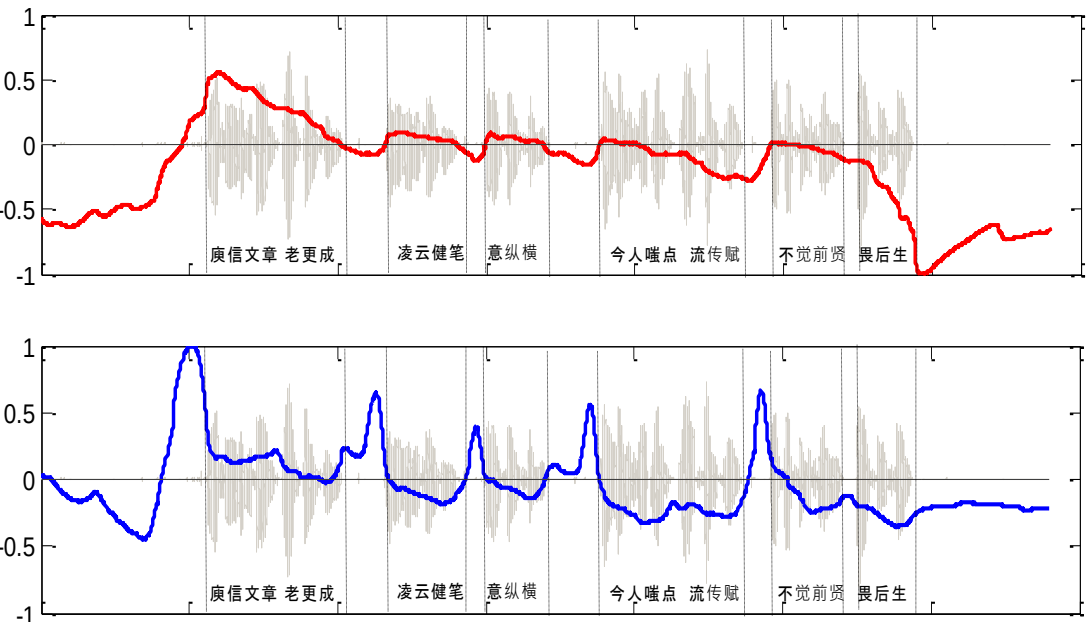


图 3.7 汉语五言绝句朗读中胸、腹呼吸与语音波形
(上图：胸呼吸与语音波形，下图：腹呼吸与语音波形)

戏为六绝句（一）

庾信文章老更成，凌云健笔意纵横。
今人嗤点流传赋，不觉前贤畏后生。

如上图 3.7 所示，汉语七言近体诗朗读中胸呼吸重置分为两级，与五言近体诗的胸呼吸方式比较一致，一个大呼吸后出现几个不同幅度的小呼吸，通常首句具有一个极大的呼吸重置，第二、三、四句具有一个较小的胸呼吸重置，也会出现在韵律短语之间。如图 3.7 所示，除了每句开始处外，“凌云健笔”和“意纵横”之间有明显的停顿和呼吸重置。从图中可看出，胸呼吸开始点比语音开始点早，而在大多数情况下呼气时间点与诗句语音开始和结束点吻合。

汉语七言近体诗朗读中腹呼吸重置时间点早于语音开始时间，开始点和峰值所在位置均早于语音开始端，腹呼吸重置结束点也比语音结束时间早，结束后接着准备下一个呼吸重置。如图 3.7 所示，“凌云健笔”的曲线中在“健”处腹呼吸一结束，紧接着在“笔”处开始下一个呼吸重置。每句开始处各具有一个大呼吸重置，第二句韵律短语之间出现了一个中呼吸，和胸呼吸不同的是，第三，第四句的韵律短语边界处也具有一个小呼吸。如“今人嗤点”与“流传赋”之间腹呼吸曲线有一个明显的小呼吸，在七言近体诗朗读中此特征比较普遍。

朗读汉语七言近体诗时，音节数增加了之后，结果显示增加了腹呼吸重置的倾向。而五言近体诗的胸腹呼吸配合方式比较稳定，二者的功能比较均匀。

结合上文所显示的结果，从方法和功能上来说，五言诗和七言诗之间最大的不同点是腹呼吸的表现。例如，

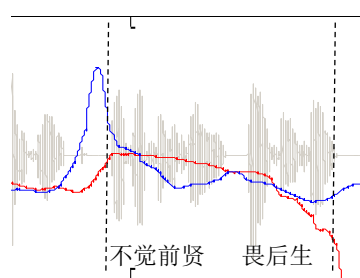


图 3.8 朗读中出现的腹呼吸实例

图 3.8 所示为七言近体诗《戏为六绝句（一）》朗读中的最后一句呼吸信号，从呼吸信号和语音波形可以看出，“不觉前贤”和“畏后生”两个韵律短语之间出现较明显的腹呼吸，与此相比胸呼吸曲线没有多大的变化，保持着扩张的状态。

值得注意的是，声学分析上的停顿与呼吸信号之间往往没有对应关系。从腹呼吸变体发现，朗读中不管语音波形上有无间隙，腹呼吸重置（抖动）还会出现在两个韵律短语之间，如下图 3.9 所示。小呼吸重置或者抖动的峰值与语音起点重合的话，其呼吸信号属于抖动，而它的峰值比语音起点早的话，那个属于小呼吸，如下图 3.9 两个例子。但是为了证实，进一步要进行气流信号的

分析。

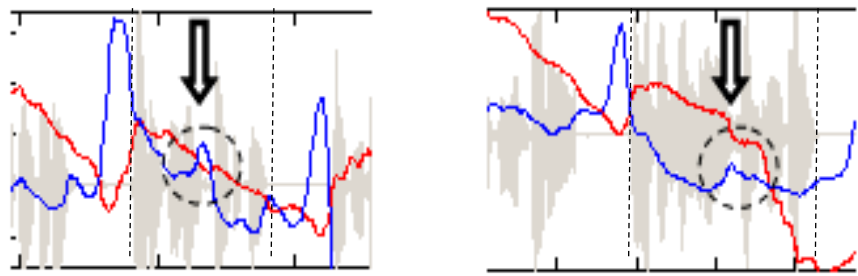


图 3.9 腹呼吸的特殊表现实例

从数据上看，朗读中胸呼吸的一级大呼吸重置幅度平均为 1.0，斜率为-0.3，二级小呼吸重置幅度平均为 0.3，斜率为-0.1，第一联开始处气息强，时长为 3.2 秒，而其他小句开始处气息则弱，而时长为 2.8 秒，与一级呼吸时长相差不大。因此，如下表 3.5 相关分析也显示，重置时长和重置幅度之间的相关系数为 0.156，说明汉语五言近体诗重置和时长之间不存在相关关系，与五言近体诗的情况不一样，与之前研究结果（谭晶晶，2008；杨锋，2012）也有所不同。

腹呼吸的一级大呼吸重置幅度平均为 1.1，斜率为-0.3，二级小呼吸重置幅度平均为 0.7，斜率为-0.2，三级小呼吸重置幅度平均为 0.2，斜率为-0.1，每句开始处气息强，开始腹呼吸曲线又高又尖，流速快，到语音开始处就变得舒缓一点，韵律短语开始处气息较弱，也较为舒缓。

Correlations

		∞Ói/ÆßÑÔ ½üìlâÊ«ÐØ °ðîü·ù¶Ê	∞Ói/ÆßÑÔ ½üìlâÊ«ÐØ °ðîüÊ±³α
∞Ói/ÆßÑÔ ½üìlâÊ«ÐØ °ðîü·ù¶Ê	Pearson Correlation	1	.156
	Sig. (2-tailed)		.101
	N	112	112
∞Ói/ÆßÑÔ ½üìlâÊ«ÐØ °ðîüÊ±³α	Pearson Correlation	.156	1
	Sig. (2-tailed)	.101	
	N	112	112

表 3.5 汉语七言近体诗呼吸幅度与时长相关

从功能上看，腹呼吸比胸呼吸来得早，而比胸呼吸结束得快。腹呼吸先给言语活动提供较大的能量，另外腹呼吸也会出现在韵律短语边界处，此时胸呼吸会持平，可以说腹呼吸对韵律单位边界比较敏感。相比，胸呼吸比腹呼吸来得晚，而语音信号结束点和胸呼吸呼气结束点比较一致，可见胸呼吸与言语活动的开始、保持和结束更有关系。

下图 3.10 给出的则是朗读汉语七言近体诗中的胸腹呼吸重置分布，左图为胸呼吸重置分布图，右图则是腹呼吸重置分布图。汉语七言近体诗的胸腹呼吸重置频度分布基本符合正态分布，就像五言近体诗的分布情况一样，胸呼吸幅度的平均值还是比腹呼吸的小一些，因而胸呼吸重置分布的整个曲线偏左一点，胸呼吸的重置幅度主要集中在小峰值周围；与它相比，腹呼吸的重置幅度更接近典型的正态分布，中间大峰的频度最多而大峰和小峰周围的峰值也比较均匀。

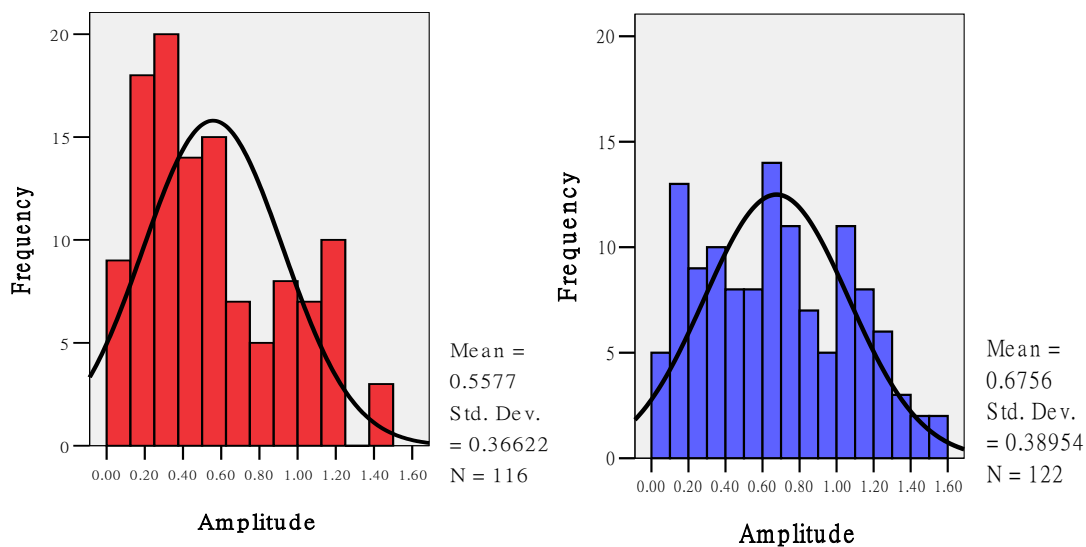


图 3.10 汉语七言近体诗呼吸重置分布图
(左图：胸呼吸，右图：腹呼吸)

有趣的是，与五言近体诗相比，朗读汉语七言近体诗的时候，胸腹呼吸的重置幅度都会比五言诗的小，而重置次数会变多。如图 3.10 所示，胸呼吸重置幅度的平均值从 0.62 调到 0.56，腹呼吸平均值从 0.75 调到 0.67，而胸腹呼吸重置次数分别为胸呼吸 94 次变成 116 次（每首律诗平均呼吸次数为 7.7 次、绝句平均为 3.9 次）、腹呼吸 96 次变成 122 次（每首律诗平均为 8.1 次、绝句平均为 4.4 次），无论是重置幅度还是次数，胸腹呼吸的变化程度比较一致，胸腹呼吸的幅度都降低了 10%，相反呼吸次数两个都增加了 25%，很有规律，无个人差异。

这说明朗读不同文体的时候朗读者会调整重置幅度和次数，根据近体诗的分析结果，从朗读者安排整体呼吸的方式来看，朗读中呼吸重置幅度调小了，就增加重置次数，保持每个人具有的整体呼吸能量的平衡（总量）。也就是说，每个朗读者所使用的呼吸能量是一样的，只是按照不同文体或不同结构的文章，由朗读者来提前安排呼吸重置幅度和次数。在此，我们只能说这是朗读汉语近体诗中的呼吸重置特征，至于韩语中的呼吸情况，下一节继续探讨。

从上图 3.10 上看，我们比较容易发现，七言近体诗的胸腹呼吸分布与五言近体诗有两个明显不同的地方：

第一，在朗读七言近体诗中，腹呼吸重置幅度会分成三个级别，在 0.2 和 0.6-0.8 和 1.0-1.3 处分别出现了三个比较明显的峰值，该特征也没有个人差异，两个发音人的趋势比较一致，如下图 3.11；

第二，和五言近体诗的分布图相比，七言近体诗的胸腹呼吸二，三级呼吸次数比一级呼吸的次数多。上文所分析的五言近体诗的呼吸重置幅度分析结果看，在一级和二级呼吸重置次数之间没有多大差异，与此相比，七言近体诗的呼吸重置分析结果显示，无论是胸呼吸还是腹呼吸，一级呼吸的次数比二级呼吸和三级呼吸的次数少。尤其从胸呼吸的情况来看，二级呼吸重置次数比一级的两倍以上。这说明从呼吸方式上看，在朗读七言近体诗中大小呼吸的交替出现的比率比五言近体诗要少很多，尤其在胸呼吸的表现中比较明显，参看下图 3.11 和表 3.6 聚类分析结果。

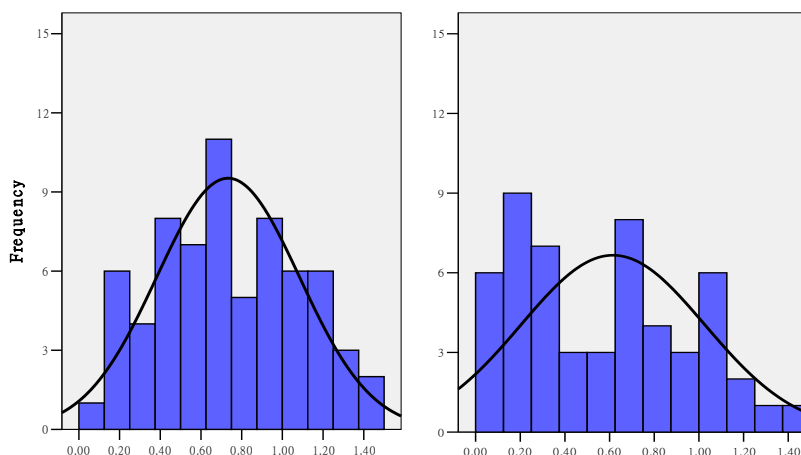


图 3.11 汉语七言近体诗腹呼吸重置分布图
(左图：中国发音人 1, 右图：中国发音人 2)

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2		
°Ói/ÆβÑÔ½üäÊ«ÐØ°ôüÖØÖÄ·ù¶È	1.04	.35	Cluster 1	35.000
			Cluster 2	81.000
			Valid	116.000
			Missing	.000

Final Cluster Centers				Number of Cases in each Cluster	
	Cluster			Cluster	
	1	2	3		
°Ói/ÆβÑÔ½üäÊ«¹º	1.16	.66	.23	Cluster 1	37.000
îüÖØÖÄ·ù¶È				Cluster 2	46.000
				Cluster 3	39.000
				Valid	122.000
				Missing	.000

表 3.6 汉语七言近体诗呼吸重置聚类分析结果

如上表 3.6 所示，胸呼吸重置聚类分析结果显示胸呼吸重置分成两级，随着平均值的减少，七言近体诗不同级别呼吸重置的中心值也比五言近体诗的稍微低一点，一级和二级呼吸的中心值分别为 1.04 和 0.35，相反重置次数分别为 35 次和 81 次，如上文分析结果已显示，比五言近体诗的重置次数增加了一些。有趣的是，虽然重置次数增加了一些，而五言和七言近体诗的一级和二级胸呼吸重置次数之间的比率比较相似，1: 2.5 左右。这说明在朗读汉语近体诗中，按照近体诗的不同类型胸呼吸重置不同阶级的幅度值会有变化，而一二级幅度之间的比率会保持一致，就是说发音人分开安排不同级别的胸呼吸重置的方式颇有规律。

结合上文给出的直方图分析来看，与五言近体诗表现不同的是七言的胸呼吸重置分布没有五言的集中，尤其在一级呼吸重置分布中没有像五言分布图中的一级胸呼吸那样频度极高的部分。由此可见，由于五言近体诗的结构比七言诗的简单，使五言近体诗的胸呼吸表现更有规则，更为稳定。

与此相比，二级胸呼吸的次数比一级胸呼吸的多两倍以上，分布范围比较广，而且除了在 0.2-0.4 处最突出的峰值之外，其他不同幅度段的频度也比较高，均高达 15。这说明在朗读汉语七言近体诗中，在诗歌开始处出现一级呼吸之后，二级呼吸出现时会比较自由地调整呼吸重置幅度。因为该特征与朗读五言近体

诗中的胸呼吸特征比较一致，所以可以说是朗读汉语近体诗中的胸呼吸特征之一。

腹呼吸呼吸重置分布和胸呼吸的情况有所不同。如上图 3.10 和表 3.6 所示，腹呼吸重置幅度分成三级，其中心值分别为 1.16、0.66 和 0.23，与图 3.10 显示的三个显著的峰值比较一致。从一级到三级不同级别的呼吸重置次数分别为 37、46 和 39 次，一级呼吸重置次数最少，二级呼吸的最多，而三者的差异不是特别明显。此结果与朗读五言近体诗中的腹呼吸重置特征比较一致，由此我们可以说，汉语近体诗腹呼吸重置的主要特征是：无论由两个级别呼吸组成的五言诗还是由三个级别呼吸组成的七言诗，不同级别的腹呼吸重置出现得比较均匀，即大呼吸小呼吸出现得更有规律。与五言近体诗的腹呼吸重置不同的是，由于七言近体诗的音节数比五言诗的多，因此整个呼吸重置分成三个不同级别，从而有效体现汉语七言近体诗的节奏感。

从上图 3.11 所示的波形图和聚类分析的结果来看，汉语七言近体诗的每诗行基本出现一次呼吸，和五言近体诗的呼吸重置不同的是，在七言诗中在韵律短语之间会出现明显的呼吸重置。总呼吸次数上，腹呼吸比胸呼吸稍微多一点，但可以说胸腹呼吸出现的次数基本一致。这说明朗读汉语七言近体诗时，和五言近体诗的胸腹呼吸功能一样，发音人用腹呼吸时的重置幅度大，以给整个言语活动提供主力量，紧接着胸呼吸要开始，坚持一段时间以保持言语活动，正好胸呼吸的终点和语音信号的终点比较一致。在次数上二者基本没有差别，只是在言语活动中的功能不同，即在朗读五言近体诗时胸腹呼吸的配合是有一定规则的，这一点与之前的研究结果（杨锋，2011 等）比较一致。

由此可以总结，朗读汉语近体诗中的呼吸重置幅度特点是：

1. 腹呼吸早于胸呼吸，重置幅度高于胸呼吸，在开始朗读时给整个言语活动提供能量，比胸呼吸早点结束而朗读中对韵律单位的界限的反映更敏感，因此呼吸次数比胸呼吸多一些；一开始朗读，胸呼吸晚于腹呼吸，呼气之后保持一定时间使言语活动持续下去，因此胸呼吸的终点和言语信号终点比较一致。
2. 不管朗读五言近体诗还是七言近体诗，胸呼吸存在二级呼吸重置，一级呼吸处于每首诗的的始点，二级呼吸出现的位置比较自由一点。
3. 腹呼吸存在三级呼吸重置，一级呼吸处于每首诗的每联开始位置，二级呼吸处于每长句的开始位置，三级呼吸处于长句末尾小句。每长句对应一个胸呼吸重置，韵律边界处出现呼吸中段抖动，腹呼吸早于胸呼吸。

总之，在朗读汉语近体诗中，汉语近体诗的不同类型对胸腹呼吸的重置幅度（谭晶晶,2008）和次数以及腹呼吸的级别有影响。

3.2 韩语诗歌的呼吸与韵律特征

3.2.1 韩语汉字音版诗歌的呼吸节奏特征

本次实验总共有 20 首汉语近体诗的韩语汉字音版（五言和七言各 10 首，绝句和律诗各 10 首），与汉语版不同的是，每句的句尾要加一到两个音节的语法词（助词），表示不同句子结构成分，结果每句音节数为 6-7 个。在韵律结

构上，韩语汉字音版与汉语版基本一致，那不同语言的呼吸韵律特征是否一致？

(1) 韩语汉字音版五言近体诗呼吸节奏特征

下图 3.12 显示的是韩语汉字音五言近体诗《静夜思》朗读中胸、腹呼吸和语音波形。下图中的红色曲线显示胸呼吸重置变化，一个大呼吸后出现几个小呼吸，蓝色曲线则是腹呼吸变化曲线，整体趋势与胸呼吸的比较相似，一般大呼吸后会有几个小呼吸出现，黑虚线是每小句语音信号的始点和终点。

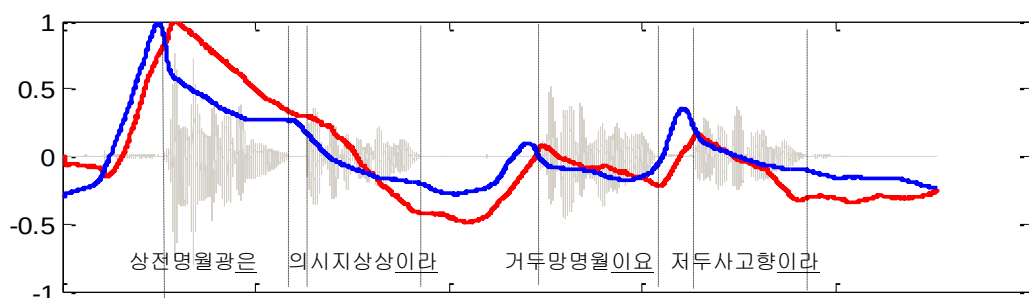


图 3.12 韩语汉字音版五言绝句朗读中胸、腹呼吸与语音波形

静夜思

床前明月光、疑是地上霜。

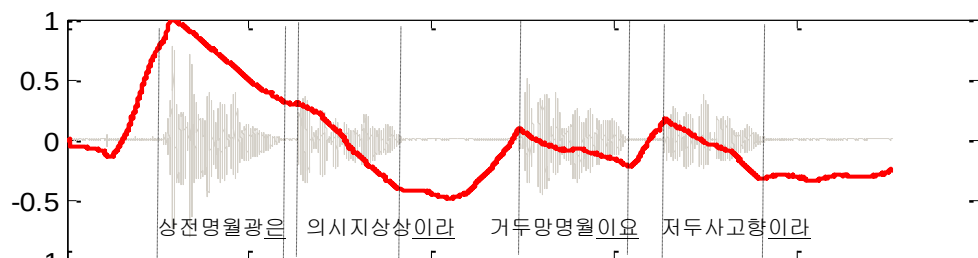
상전명월광은, 의시지상상이라.

举头望明月，低头思故乡。

거두망명월이요, 저두사고향이라

从方式上看，不管绝句还是律诗，第一联开始处具有一个极大的呼吸重置，不同的是，朗读绝句中每联开始处不一定会有大呼吸重置，而朗读律诗中，在大多数情况下，每联开始处会出现大呼吸，每小句开始处具有一个小呼吸，即大小呼吸会交替出现。

与朗读汉语近体诗中的呼吸方式相比，相同的是在大多数情况下，腹呼吸开始点早于胸呼吸和语音信号开始点，腹呼吸呼气斜率会比胸呼吸急促一些。最大的不同点是朗读韩语汉字音版近体诗中胸腹呼吸之间的重置方式和次数比较一致，朗读律诗中呼吸重置方式更有规律。



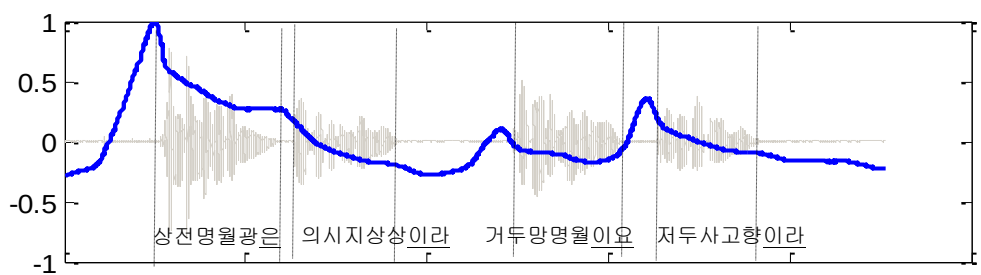


图 3.13 韩语五言绝句朗读中胸、腹呼吸与语音波形
(上图：胸呼吸与语音波形，下图：腹呼吸与语音波形)

如上图 3.13 所示，韩语汉字音版的五言近体诗朗读中胸呼吸重置分为两级，通常第一联开始处具有一个又大又长的呼吸重置，而第二联开始处不一定会出现大呼吸，第三、四句具有一个较小的胸呼吸重置。如图 3- 所示，第一联“상전명월광은, 의시지상상이란. (床前明月光、疑是地上霜。)”之前具有一个非常大的呼吸重置，“거두망명월이요 (举头望明月)”和“저두사고향이 (低头思故乡)”之前各具有一个小呼吸重置。

从图中可以看出，在大多数情况下，胸呼吸的呼气开始点与诗句语音开始和结束时间吻合，以上列出的胸呼吸特征与朗读汉语近体诗中的胸呼吸特征比较一致，不同的是在胸呼吸信号中很少发现持平的曲线，几乎都一直下降到语音结束。此外，朗读韩语汉字音版律诗中胸呼吸除了第一联具有的超大呼吸之外，从第二联开始大小呼吸比较有规律地出现。

朗读韩语汉字音版五言近体诗中腹呼吸信号呼吸重置时间与峰值所在位置均早于语音开始点，腹呼吸重置结束时间点也要早于语音结束点，该特征与汉语近体诗朗读中的腹呼吸特征比较一致。如上图 3. 所示，“거 (举)”和“저 (低)”之前先是蓝色胸呼吸曲线，然后才是红色腹呼吸曲线。

腹呼吸重置特征与胸呼吸一样，第一联具有又大又长的呼吸重置，其后的每联开始处不一定出现大呼吸，在五言绝句的呼吸重置中小呼吸比较多，相比在五言律诗的呼吸中大小呼吸交替出现的情况比较普遍。总之，在方法上朗读韩语汉字音版近体诗中胸呼吸和腹呼吸的特征显得比较相似。

从数据而言，朗读中胸呼吸的一级大呼吸重置幅度平均为 1.0，斜率为-0.3，二级小呼吸重置幅度平均为 0.3，斜率为-0.2，第一联开始处胸呼吸气息强，而其他小句开始处气息则弱，也较为舒缓，大小呼吸时长分别为 2.3 和 1.4 秒，不同级别重置之间的时长差异比较大，如下表 3.7 所示，幅度和时长的相关分析结果显示，二者之间具有显著的相关性，与汉语版五言诗呼吸特征一致。

腹呼吸的一级大呼吸重置幅度平均为 1.0，斜率为-0.5，二级小呼吸重置幅度平均为 0.4，斜率为-0.3，胸腹呼吸重置的平均幅度比较相似，而一二级的平均斜率有所差别，腹呼吸的一二级重置之间斜率的差别相对较明显。如图 3. 所示，一级腹呼吸气息强，呼吸曲线又高又尖，流速快，小呼吸气息则弱，与胸呼吸不同的是，无论是大呼吸还是小呼吸，与胸呼吸相比，腹呼吸高峰很尖，呼气一开始流速非常快，其后气息变得舒缓。

Correlations

		°«Ôï°×Ô°æ ĩãÑÔ½ûĩãÊ« ĐØ°đĩũ·ù¶Ê	°«Ôï°×Ô°æ ĩãÑÔ½ûĩãÊ« ĐØ°đĩũ·ù¶Ê
°«Ôï°×Ô°æ ĩãÑÔ½ûĩãÊ« ĐØ°đĩũ·ù¶Ê	Pearson Correlation	1	.675**
	Sig. (2-tailed)		.000
	N	105	105
°«Ôï°×Ô°æ ĩãÑÔ½ûĩãÊ« ĐØ°đĩũ·ù¶Ê	Pearson Correlation	.675**	1
	Sig. (2-tailed)	.000	
	N	105	105

** . Correlation is significant at the 0.01 level (2-tailed).

表 3.7 韩语汉字版五言近体诗呼吸幅度与时长相关

从功能上说，腹呼吸比胸呼吸来得早，在大多情况下比胸呼吸结束得也早，在语音信号结束前已经开始慢慢吸气，准备下一次言语活动。相比，胸呼吸呼气结束点和语音信号结束点比较一致，由此我们可以知道虽然胸腹呼吸几乎同时开始，但胸呼吸和言语活动更有密切的关系，腹呼吸先给整个言语活动提供能量之后，提前结束并准备下次言语活动所需的呼吸能量。

以上几个特征和朗读汉语近体诗中的呼吸特征比较一致，不同的是，朗读韩语汉字音版近体诗中，虽然腹呼吸开始点早于胸呼吸，而胸呼吸也紧接着开始，不能说只有腹呼吸先提供能量。另外，不管是腹呼吸还是胸呼吸，除了小句以上的韵律单位之外，在两种呼吸表现中，在较小的韵律单位之间几乎都不会出现小呼吸。因此，不能说哪种呼吸对韵律单位或者语言结构更为敏感，就是说二者的功能没有明显的区别。

下面给出的是韩语汉字音版五言近体诗的呼吸重置直方图和对胸腹呼吸重置幅度分别做的聚类分析结果。从此我们可以探讨韩语汉字音近体诗的呼吸重置分布趋势以及和汉语近体诗呼吸重置分布特征之间的异同。

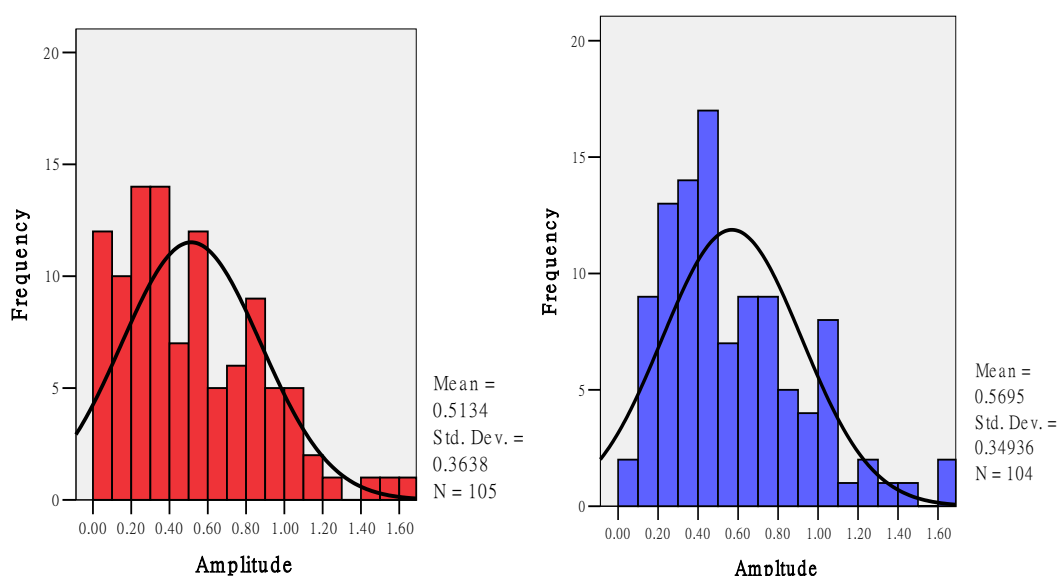


图 3.14 韩语汉字音五言近体诗呼吸重置分布图
(左图：胸呼吸，右图：腹呼吸)

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2		
°«Ôï°×ÔÔôîâÑÔ½ûîâÊ« ĐØ°ôîûÔñ·ù·ù¶Ê	.95	.30	1	35.000
			2	70.000
			Valid	105.000
			Missing	.000

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2		
°«Ôï°×ÔÔôîâÑÔ½ûîâÊ« °ôôîûÔñ·ù·ù¶Ê	.99	.37	1	33.000
			2	71.000
			Valid	104.000
			Missing	.000

表 3.8 韩语汉字音版五言近体诗胸呼吸重置聚类分析结果

从上图 3.14 和表 3.8 看出, 韩语汉字音版近体诗朗读中胸腹呼吸重置明显分为两级, 胸腹呼吸重置的平均值分别为胸呼吸 0.52、腹呼吸 0.57, 腹呼吸的重置幅度比胸呼吸的稍微高一些, 这些特征与汉语五言近体诗朗读中的呼吸信号特征比较一致。

韩语汉字音版五言近体诗朗读中胸呼吸重置聚类分析结果显示一级和二级呼吸的中心值分别为 0.95 和 0.30, 重置次数分别为 35 次和 70 次, 可见无论是胸呼吸还是腹呼吸, 都是二级呼吸次数比一级呼吸的多一倍。如图 3.14 直方图给出的分布所示, 胸呼吸重置中一级呼吸比较稳定, 幅度集中在了 1.0-1.1 上。与此相比, 二级呼吸的次数比一级呼吸的两倍以上, 从分布范围而言不同幅度段出现的频度都比较高, 该特征与汉语胸呼吸特征比较相似, 在朗读韩语汉字音五言近体诗中, 二级呼吸幅度相对来说比较自由。

腹呼吸呼吸重置分布和胸呼吸的情况比较相似, 二者的呼吸次数也基本一致, 胸呼吸次数为 105 次, 腹呼吸的为 104 次, 在每小句里平均出现 0.85 次的呼吸(绝句 0.8 次, 律诗 0.9 次)。如表 3.8 所示, 朗读韩语汉字音版五言近体诗中腹呼吸重置幅度也分为两级, 一级和二级腹呼吸次数分别为 33 和 71 次, 二级腹呼吸次数比一级的多两倍以上, 能够反映上文给出的实际呼吸曲线趋势。一级和二级呼吸的中心值分别为 0.99 和 0.37, 比胸呼吸的平均幅度稍微高一点, 也可以说比较接近。

与汉语版呼吸特征最大的不同点是, 读汉语五言近体诗中腹呼吸的一级和二级呼吸的次数相对比较均匀, 也就是说大呼吸小呼吸会有规律地交替出现, 而韩语汉字音版五言近体诗的胸、腹呼吸中二级呼吸比较多, 一二级呼吸比例大致为 1:2, 因此呼吸分布曲线均呈高峰偏左的, 胸腹呼吸表现比较一致。

由此可见, 朗读韩语汉字音版五言近体诗中的呼吸信号有如下特点:

韩语汉字音版韵律结构比较特殊, 因为韩语的音步由 2-7 个音节组成, 朗读韩语汉字音版五言近体诗中, 安排的呼吸幅度比汉语版的低, 也足以完成一小句, 体现出不同语言结构的不同节奏特征。

从呼吸方式看, 胸腹呼吸都存在二级呼吸重置, 二者的呼吸方式和特征比较相似, 一级呼吸处于每首诗的开始位置, 往往把首联都以一次长呼吸来完成, 第二、三、四联开始处不一定会出现一级呼吸, 常由二级呼吸重置替代。由于

二级呼吸出现在每小句开始处，因而二级呼吸重置次数比一级呼吸的多。

汉语五言近体诗朗读中胸腹呼吸二者具有不同功能，相比韩语汉字音版五言近体诗朗读中胸腹呼吸的功能相对比较相似，二者的呼吸大致曲线基本重合。虽然其功能不太显著，但是腹呼吸仍然具有给言语活动提供能量的功能，胸呼吸主要与韵律边界处有关。

(2) 韩语汉字音版七言近体诗呼吸节奏特征

下图 3.15 给出的是汉语七言近体诗《绝句》的韩语汉字音版朗读中的胸、腹呼吸与语音波形。在一般的情况下，韩语汉字音版七言近体诗朗读中呼吸重置明显分为两级。图上的红色曲线显示胸呼吸重置变化，大小呼吸交替出现，蓝色曲线则是腹呼吸变化曲线，与胸呼吸的情况比较一致，也是大小呼吸出现方式较有规律，黑虚线是每小句语音信号的始点和终点。

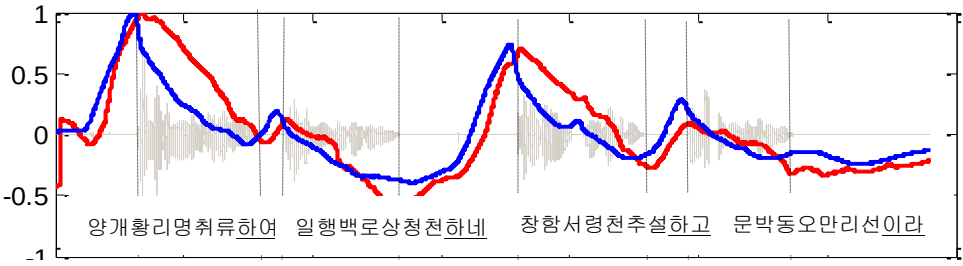


图 3.15 韩语汉字音版七言近体诗朗读中胸、腹呼吸与语音波形

绝句

两个黄鹂鸣翠柳，一行白鹭上青天。

양개황리명취류 하여, 일행백로상청천 하네.

窗含西岭千秋雪，门泊东吴万里船。

창함서령천추설 하고, 문박동오만리선 이라.

从方式上看，朗读中在大多数情况下，每联开始处具有一个大呼吸重置，随后每小句的开始处具有一个二级呼吸重置，该特征和前研究比较一致。不同的是，朗读汉语七言近体诗中胸呼吸重置方式比腹呼吸的自由一点，而韩语汉字音版近体诗朗读中胸、腹呼吸的趋势大体比较相似。+

如上图 3.15 所示，比韩语汉字音版五言近体诗朗读中的呼吸方式显得更合格律，每小句都会具有一次呼吸。该特征与韩语诗歌格律有一定关系，因为在韩语汉字音版的七言近体诗每小句里平均有 9 个音节，按照韩语诗歌的节奏单位，9 个音节大致能包括两个音步，正是一小句的长度。可以说，虽然汉语近体诗的格律主要支配韩语汉字音版的整体呼吸安排，但是在遵守格律的范围内，韩语近体诗本身的节奏特征也会影响到“小呼吸（二级呼吸或三级呼吸）”的出现与否。如，汉语基本节奏单位为音步，在汉语七言近体诗的音步或者韵律短语之间常出现小呼吸信号。

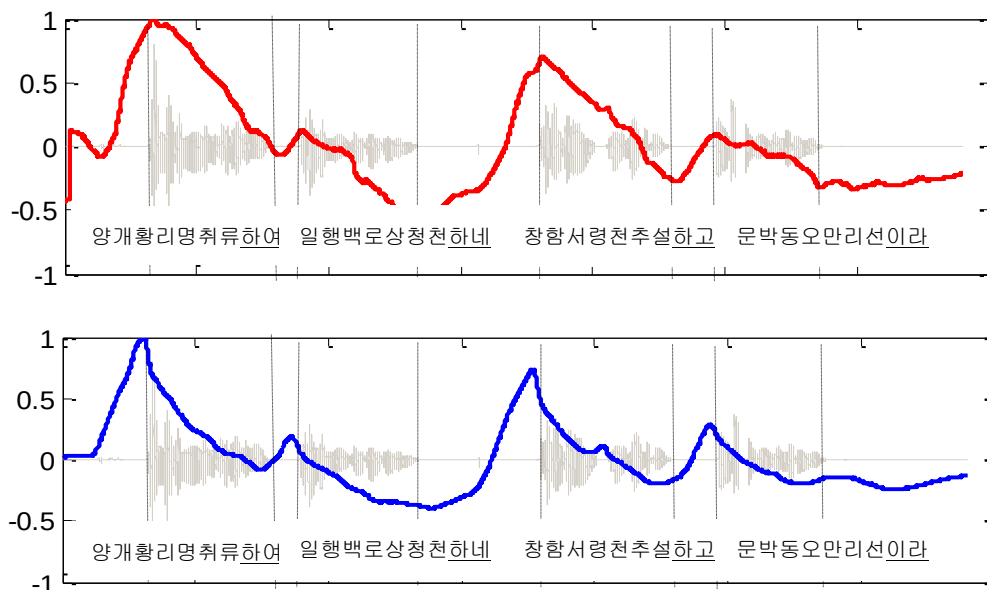


图 3.16 韩语汉字音版七言绝句朗读中胸、腹呼吸与语音波形
(上图：胸呼吸与语音波形，下图：腹呼吸与语音波形)

图 3.16 分开给出胸、腹呼吸的呼吸曲线，上图为胸呼吸曲线和语音波形、下图为腹呼吸曲线与语音波形。

韩语汉字音版七言近体诗朗读中胸呼吸重置分为两级，与五言近体诗的胸呼吸方式一致，大小呼吸交替出现，而在律诗中会常出现一个大呼吸加几个不同幅度的小呼吸的类型。胸呼吸开始点和峰值都比语音开始点早，在大多数情况下，胸呼吸呼气时间点与诗句语音开始和结束点吻合，无论是汉语还是韩语，胸呼吸都有相同的特征。在一般情况下，在韵律单位边界处经常会具有抖动，或者胸呼吸曲线会显示“阶梯式”下降。如图 3.16 所示，胸呼吸将“문박(门泊)”、“동오(东吴)”与“만리선(万里船)”分为不同阶梯进行呼气，在前两个音步边界处出现的间隙对这种呼气方式没有显著影响。在第三句中，“창함서령(窗含西岭)”与“천추설(千秋雪)”之间有较短的停顿，而胸呼吸一直在下降，即停顿和胸呼吸重置不会一一对应，谭晶晶(2008)已在研究中证明在停顿和间歇处并不一定出现呼吸信号。

如图 3.16 的下图所示，韩语汉字音版七言近体诗朗读中腹呼吸重置明显分为两级，其呼吸方式与胸呼吸比较一致，在大多情况下，大小呼吸会交替出现，而根据不同被试在律诗中也会出现一个大呼吸加几个小呼吸的方式。从图中可看出在诗句中间停顿处腹呼吸曲线具有一个不容易测量的极小的呼吸抖动，有时还会出现“阶梯式”呼吸。如“창함서령(窗含西岭)”与“천추설(千秋雪)”之间有短暂的停顿，二者之间腹呼吸曲线有一个明显抖动。

与胸呼吸不同的是，腹呼吸重置时间点和峰值所在位置均早于语音开始端，重置结束点也比语音结束时间早，结束后紧接着开始下一个呼吸重置或者保持平线，这几点属于近体诗朗读中腹呼吸具有的普遍特征。

与汉语七言近体诗朗读中的腹呼吸相比，从方式上朗读韩语汉字音版中腹呼吸最显著的不同点是，不管是五言还是七言近体诗，腹呼吸都分成两级，大体没有显著区别。由此可见，从以韵律短语为基本节奏单位的韩语而言，加一个音步的七言近体诗对大体呼吸安排没有决定性影响，一直根据汉语近体诗的

基本格律来安排呼吸节奏。

从数据来说，朗读韩语汉字音版七言近体诗中胸呼吸的一级大呼吸重置幅度平均为 1.0，斜率为-0.4，二级小呼吸重置幅度平均为 0.3，斜率为-0.2，大呼吸开始处气息强，小呼吸则气息则弱，也较为舒缓，而时长为 2.3 和 2.1，二者的区别不大，如下图 3.9 所示，韩语汉字音版七言近体诗朗读中重置幅度与时长之间没有相关，与汉语版七言近体诗一致。腹呼吸的一级大呼吸重置幅度平均为 1.0，斜率为-0.4，二级小呼吸重置幅度平均为 0.4，斜率为-0.3，如图 3.15 所示，腹呼吸的一二级呼吸特征与胸呼吸的情况大体一致，只是在出现二级呼吸的时候，胸呼吸会比腹呼吸更缓慢一点。

Correlations			
		°«Öï°×Ö°æ ÆßÑÖ½üläÊ« ÐØ°öü·ü¶Ê	°«Öï°×Ö°æ ÆßÑÖ½üläÊ« ÐØ°öüÊ±³»
°«Öï°×Ö°æ ÆßÑÖ½üläÊ« ÐØ°öü·ü¶Ê	Pearson Correlation	1	.075
	Sig. (2-tailed)		.574
	N	118	118
°«Öï°×Ö°æ ÆßÑÖ½üläÊ« ÐØ°öüÊ±³»	Pearson Correlation	.075	1
	Sig. (2-tailed)	.574	
	N	118	118

表 3.9 韩语汉字音版七言近体诗呼吸幅度与时长相关

从功能上说，腹呼吸比胸呼吸早，在一般情况下比胸呼吸结束得更快，在语音信号结束前已经开始慢慢吸气，或者保持平线并准备下一次言语活动。相比，胸呼吸呼气开始、结束点都与语音信号比较吻合，由此我们可以知道虽然胸腹呼吸几乎同时开始，但胸呼吸和言语活动更有密切的关系，结合上文所分析的汉语近体诗朗读中的胸呼吸特征，该特征是胸呼吸具有的普遍特征之一。腹呼吸的表现先给整个言语活动提供能量之后，朗读结束前结束并准备下一次呼吸安排。但是，如上图 3.15 所示，在朗读中的胸、腹呼吸曲线和轨迹比较接近，不能说哪种呼吸主要担任哪种功能，与韩语汉字音版五言近体诗的情况比较一致。

另外，在韵律单位之间出现小呼吸的时候，在大多数情况下，胸、腹呼吸一起出现，因此不能断定哪种呼吸对韵律单位更为敏感，就是说二者的功能没有明显的区别。

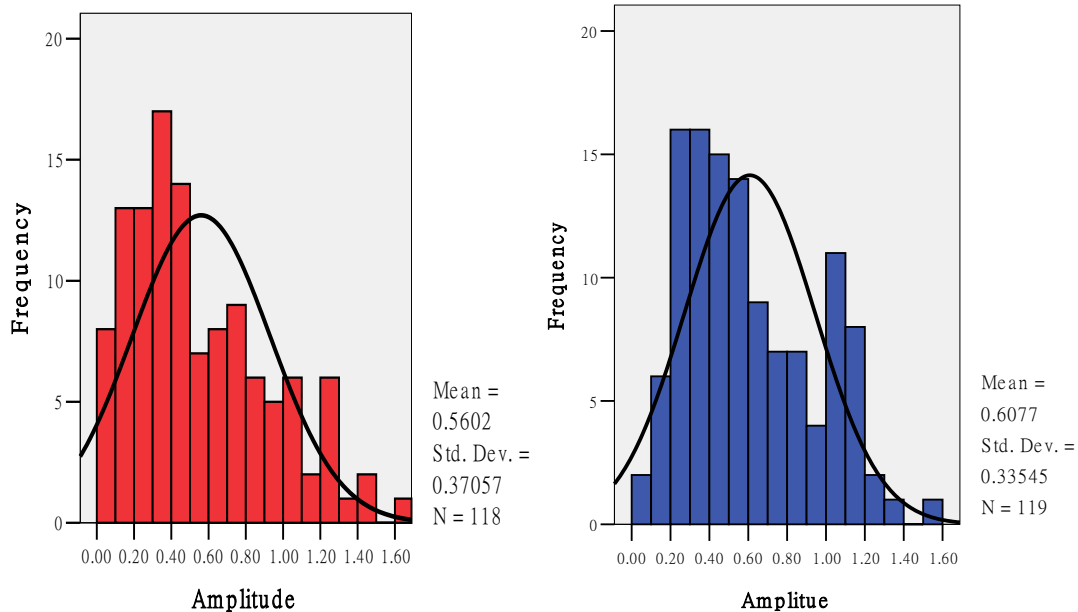


图 3.17 韩语汉字音版七言近体诗呼吸重置分布图
(左图：胸呼吸，右图：腹呼吸)

上图 3.17 给出的是朗读韩语汉字音版七言近体诗中的胸、腹呼吸重置分布，左图为胸呼吸重置分布图，右图则是腹呼吸重置分布图。汉语七言近体诗的胸、腹呼吸重置频度分布曲线高峰呈偏左，因为胸呼吸幅度的平均值比腹呼吸的低，分别为 0.56 和 0.61。基本上，胸呼吸重置幅度主要集中在 0.3-0.4 周围，大峰值分布比较分散，范围比较广，相比腹呼吸重置幅度分布在 0.2-0.4 和 1.0-1.1 处出现比较明显的峰值。如图 3.17 所示，朗读韩语汉字音版七言近体诗中胸、腹呼吸重置次数分别为 118 次和 119 次，每首律诗平均呼吸次数呼吸为 7.7 次、绝句平均为 4.1，就是说七言近体诗的每诗行基本出现一次呼吸，胸、腹呼吸出现的次数基本一致。

与五言近体诗朗读中的呼吸特征相比，朗读韩语汉字音版七言近体诗中胸、腹呼吸重置幅度会变高，重置次数也会变多，腹呼吸平均幅度比胸呼吸的高。如上图 3.17 所示，胸呼吸重置幅度的平均值从 0.52 调到 0.56，腹呼吸平均值从 0.57 调到 0.61，而胸腹呼吸重置次数分别为胸呼吸 105 次变成 118 次，腹呼吸 104 次变成 119 次，无论是重置幅度还是次数，胸、腹呼吸的变化方式与程度比较一致，非常有规律，这是汉语和韩语汉字音版朗读中的呼吸方式具有的共同特征。

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2	1	2
°«Öï°°×ÖÖô/ÆßÑÖ½üläÊ« ÐØ°ôïüÖØÖÄ-ù¶Ë	1.01	.34	Valid	118.000
			Missing	.000

Final Cluster Centers		Number of Cases in each Cluster	
	Cluster		
	1	2	
°«Öï°×ÖÖδÆßÑÖ½üîâÊ« ,°δüÖÖÖÄ·ü¶Ê	.96	.38	
Cluster	1	2	
			46.000
			73.000
Valid			119.000
Missing			.000

表 3.10 韩语汉字音版七言近体诗胸、腹呼吸重置聚类分析结果

上表 3.10 胸呼吸重置聚类分析结果显示胸呼吸重置可分成两级，如呼吸信号与语图所示，朗读韩语汉字音版中五言与七言近体诗的呼吸情况一致，一级和二级呼吸的中心值分别为 1.01 和 0.34，重置次数分别为 39 次和 79 次，如上文分析结果已显示，比五言近体诗的重置幅度与次数都增加了一些。值得注意的是，虽然朗读的文体不同，重置次数也增加了一些，而五言和七言近体诗朗读中的一级和二级胸呼吸重置次数的比例仍然保持为 1:2。这说明，不管是汉语版还是韩语汉字音版，按照汉语近体诗的不同类型，不同阶级的胸呼吸重置幅度值都会有变化，而发音人安排胸呼吸重置的方式是不会变的，颇有规律，可以说是胸呼吸重置的普遍性特征之一。

结合上文给出的直方图的分析来看，与五言近体诗的表现不同的是七言的胸呼吸重置分布没有五言的集中，尤其在一级呼吸重置分布中没有像五言分布图中的一级胸呼吸那样频度极高的部分。由此可见，由于五言近体诗的结构比七言诗的简单，使五言近体诗的胸呼吸表现更有规则，更为稳定。

与此相比，二级胸呼吸的次数比一级胸呼吸的两倍以上，分布范围比较广，而且除了在 0.2-0.4 处最突出的峰值之外，其他不同幅度段的频度也比较高，均高达 15。这说明在朗读汉语七言近体诗中，在诗歌开始处出现一级呼吸之后，二级呼吸出现时会比较自由地调整呼吸重置幅度。因为该特征与朗读五言近体诗中的胸呼吸特征比较一致，所以可以说是朗读汉语近体诗中的胸呼吸特征之一。

腹呼吸呼吸重置分布和胸呼吸的情况比较相似，腹呼吸重置幅度还是分成二级，与五言近体诗朗读中的腹呼吸分类结果一致。如上图 3.17 与表 3.10 所示，一级与二级中心值分别为 0.96 和 0.38，与胸呼吸相比，两个峰值比较明显。不同级别的呼吸重置次数分别为 46 次和 73 次，一级呼吸重置次数还是比二级呼吸重置少，大致 1:1.5 的比例。与五言诗朗读中的腹呼吸分析结果相比，七言诗中的腹呼吸一二级出现频率更均匀一点。由此可见，虽然韩语汉字音版近体诗朗读中胸、腹呼吸趋势比较相似，但是从具体重置安排来看，腹呼吸的变化比胸呼吸。此结果与朗读五言近体诗中的腹呼吸重置特征比较一致，由此我们可以说，汉语近体诗腹呼吸重置的主要特征是：无论由两个级别呼吸组成的五言诗还是由三个级别呼吸组成的七言诗，不同级别的腹呼吸重置出现得比较均匀，即大呼吸小呼吸出现得更更有规律。与五言近体诗的腹呼吸重置不同的是，由于七言近体诗的音节数比五言诗的多，因此整个呼吸重置分成三个不同级别，从而有效体现汉语七言近体诗的节奏感。

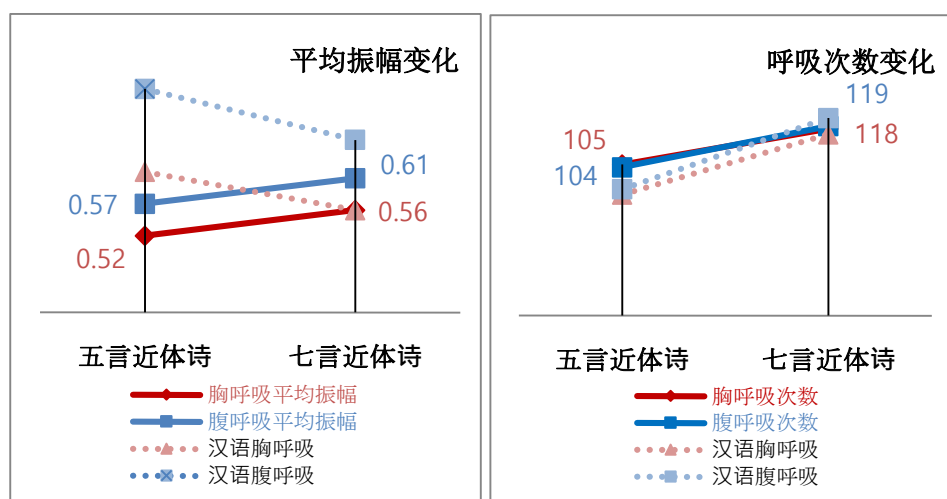


图 3.18 汉语与韩语汉字音版近体诗朗读中胸、腹呼吸幅度与次数变化

结合上图 3.18 所示的对比情况以及上文已分析过的汉语近体诗朗读中的呼吸振幅和次数变化特征来看，汉语版与韩语汉字音版之间的振幅变化趋势截然不同。从图 3.18 给出的对比来看，汉语版朗读中，随着韵律结构变化而次数增加，平均呼吸振幅会下降，相反，韩语汉字音版朗读中，随着音步数的增加而呼吸次数会增加，呼吸幅度也会升高，考虑韩语韵律结构来看，这是因为朗读七言近体诗中所增加的两个音节对韩语韵律结构影响不大，与五言诗呼吸特征相比，韩语汉字音版七言近体诗呼吸特征没有显著变化，仍然分成两级呼吸，因此没有小呼吸次数的增加导致了平均幅度的提高。其具体原因，我们可以结合韩语翻译版中的呼吸情况而进一步探讨。

与汉语七言近体诗朗读中的呼吸特征不同的是，汉语七言近体诗朗读中腹呼吸重置分布比较均匀，并且分成三级呼吸，而韩语汉字音版七言近体诗朗读中的腹呼吸仍然保持二级呼吸，与胸呼吸的表现比较一致。在次数上二者基本没有差别，只是在言语活动中的功能不同，即在朗读汉字音近体诗时胸腹呼吸的配合也是有一定规则的，符合近体诗格律。

从上文所示的波形图和聚类分析的结果来看，汉语七言近体诗的每诗行基本出现一次呼吸，和五言近体诗的呼吸重置不同的是，在七言诗中在韵律短语之间会出现明显的呼吸重置。总呼吸次数上，腹呼吸次数比胸呼吸的稍微多一点，但可以说胸腹呼吸出现的次数基本一致。这说明朗读汉语七言近体诗时，和五言近体诗的胸腹呼吸功能一样，发音人用腹呼吸时的重置幅度大，以给整个言语活动提供主力量，紧接着胸呼吸要开始，坚持一段时间以保持言语活动，正好胸呼吸的终点和语音信号的终点比较一致。在次数上二者基本没有差别，只是在言语活动中的功能不同，即在朗读五言近体诗时胸腹呼吸的配合是有一定规则的，这一点与之前的研究结果（杨锋，2011 等）比较一致。

由此可以总结，朗读韩语汉字音版七言近体诗中的呼吸特点是：

与韩语汉字音版五言近体诗呼吸特征比较相似，不管是五言还是七言，胸呼吸还是腹呼吸，平均呼吸幅度都比汉语版的低，说明韩语朗读者将韩语汉字音版近体诗视为比韵律句等同的大单位，从而体现出不同语言结构的不同节奏特征。但是韩语朗读者安排汉字版近体诗朗读时，严格遵守汉语格律而完成朗读，因此整体呼吸格局与汉语近体诗比较一致。

从呼吸方式看，胸腹呼吸都存在二级呼吸重置，二者的呼吸方式和特征比较相似，与五言诗呼吸特征不同的是，朗读七言诗中大小呼吸出现更有规律，很少出现长呼吸。因此，每联开始处会出现一级呼吸，每联第二句开始处出现小呼吸。由于二级呼吸出现的地方比较自由，往往第二联还会出现小呼吸，因此二级呼吸重置次数比一级呼吸的多。与汉语七言近体诗呼吸特征最明显的不同点是，在韩语汉字音版朗读中，韵律短语之间基本没有出现明显的呼吸。

综合韩语汉字音版近体诗朗读中的呼吸特征，可以说胸腹呼吸的功能相对比较模糊，二者的呼吸曲线有时间差而基本重合。虽然其功能不太显著，但是腹呼吸仍然具有给言语活动提供能量的功能，早开始早结束早准备。相比，胸呼吸主要与韵律边界处有关，胸呼吸的峰值与终点，和言语信号时间点比较一致。

3.2.2 韩语翻译版诗歌的呼吸节奏特征

本次实验总共有 20 首汉语近体诗的韩语翻译版诗歌，虽然韩语翻译版诗歌也是每首由两联四句或四联八句组成，而韩语翻译版五言和七言近体诗每句的平均音节数分别为 11 个和 14 个（汉语五言和七言近体诗音节数分别为 5 个和 7 个）。可见与汉语近体诗相比，韩语翻译版近体诗的音节数远远超过汉语近体诗的音节数，近似达到汉语的两倍以上。从韵律结构来看，韩语翻译版近体诗五言诗每句由 2-6 个音步，七言诗由 3-6 个音步来组成一小句（汉语分别为 2 个和 3 个），显然汉、韩两种版本的差异比较明显。那不同韵律结构如何影响到生理呼吸的安排，如何显示不同语言的特征？

从上文已分析过韩语汉字音版近体诗的呼吸节奏特征以及与汉语近体诗朗读中的呼吸特征的异同。由此我们可以继续探讨汉语与韩语朗读中胸、腹呼吸普遍具有的特征是什么；二者的呼吸特征到底如何不同；近体诗格律能否支配具有不同结构的语言节奏等问题。

（1）韩语翻译版五言近体诗的呼吸节奏特征

我们首先考查韩语翻译版五言近体诗的呼吸特征。下图 3.19 显示的是汉语近体诗《静夜思》的韩语翻译版朗读中的胸、腹呼吸曲线和语音波形。图中红色曲线显示胸呼吸重置变化，蓝色曲线则是腹呼吸变化曲线，黑虚线是每小句语音信号的始点和终点。大体来看，胸、腹呼吸曲线变化趋势比较相似，大小呼吸交替出现。

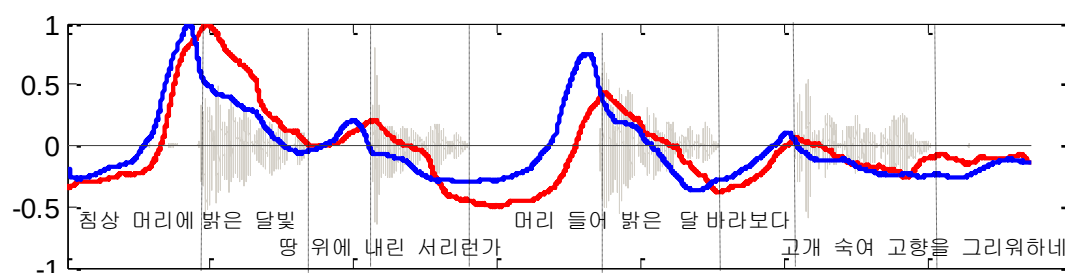


图 3.19 韩语翻译版五言近体诗朗读中胸、腹呼吸与语音波形

静夜思

床前明月光、疑是地上霜。

침상 머리에 밝은 달빛, 땅 위에 내린 서리런가.

举头望明月，低头思故乡。

머리들어 밝은 달 바라보다, 고개숙여 고향을 그리워하네.

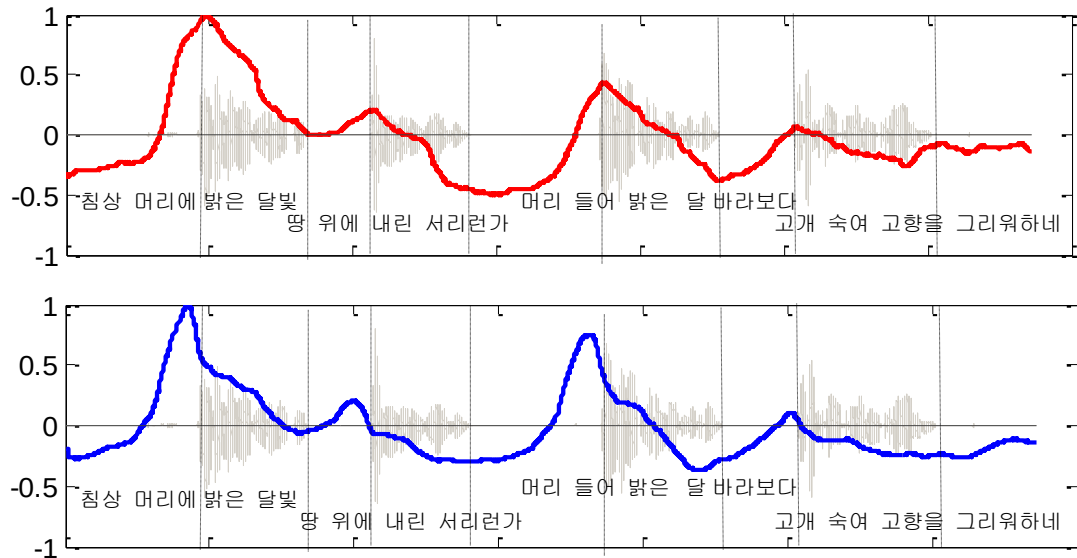


图 3.20 韩语翻译版五言绝句朗读中胸、腹呼吸与语音波形
(上图：胸呼吸与语音波形，下图：腹呼吸与语音波形)

上图 3.20 则分别给出胸、腹呼吸曲线和语音波形图。

从方式上看，朗读中在大多数情况下，韩语翻译版五言近体诗朗读中胸、腹呼吸重置明显分为两级，每联开始处具有一个大呼吸重置，每小句的开始处具有一个小呼吸重置，大小呼吸交替出现的方式颇有规律，更接近于汉语五言近体诗朗读中的呼吸特征。不同的是，胸、腹呼吸二者的曲线变化趋势大体相似，只是腹呼吸早于胸呼吸。由此可见，这点属于韩语近体诗朗读中的胸、腹呼吸配合特征。

如上图 3.20 所示，在大多数情况下，胸呼吸的呼气开始点与诗句语音开始时间吻合，腹呼吸重置时间点与峰值所在位置均早于语音开始时间与胸呼吸，其呼气信号在语音信号结束前结束，同时开始下一次呼吸或者保持平线。综合上文所分析的汉语与韩语版呼吸特征，这是不管是汉语还是韩语，朗读近体诗中的胸、腹呼吸各自具有的普遍性特征。韩语翻译版五言近体诗朗读中，基本上在诗句中间音步边界处不会出现明显的抖动或者小呼吸，可见这是汉语与韩语五言近体诗朗读中的呼吸共同具有的特征。

值得注意的是，虽然韩语翻译版五言近体诗的音节数以及音步数比汉语版更多，每句平均有 11 个音节，包含 2-6 个音步，并且根据韩语诗歌的格律，每两个音步出现较长的停顿，但是本实验显示韩语翻译版五言近体诗朗读中的呼吸方式仍然符合汉语近体诗的格律，就是说不同韵律结构对呼吸方式没有显著

影响。第一句和第三句之前各具有一个大呼吸重置，如图 3. 所示，“침상머리에/밝은 달빛（床前/明月光）”和“머리들어/밝은 달/바라보다（举头/望明月）”分别包括两个和三个音步，而二者的呼吸幅度之间没有显著差异，第二句和第四句之前各具有一个小呼吸重置，“땅 위에 내린/서리런가（疑是/地上霜）”和“고개숙여/고향을/그리워하네（低头/思故乡）”也分别包括两个和三个音步，但是二者的呼吸重置幅度还是比较一致。就是说，不管是几个音步，每联的第一句会出现大呼吸，第二句会出现小呼吸，大小呼吸交替出现的呼吸方式是一种固定模式。如上图 3. 所示，朗读者会调整呼吸幅度和斜率以及每个音节的时长而体现符合近体诗格律的呼吸模式。

但有两种例外，在小句里偶尔会出现两次小呼吸，具体如下：

一种是一句里边音节数与音步数比较多，一小句一般包含 5-6 个音步，15-18 个音节，因此从生理上、语义上都需要呼吸；另一种是按照韩语诗歌的节奏，每两个音步出现一次呼吸，根据朗读者对诗歌的解释，表现出韩语诗歌的节奏感与个人情感。如下图 3. 21 所示，五言近体诗《江雪》第二句“만 갈래 산길은 / 사람발길 끊어졌네.（万径/人踪灭）”之间出现胸、腹呼吸曲线有一倍明显的小呼吸。

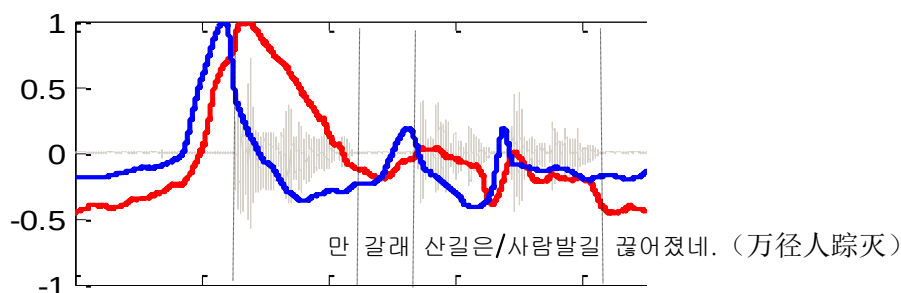


图 3. 21 韩语翻译版五言近体诗《江雪》中出现的小呼吸实例

因为诗歌的文体形式比较特殊，生理呼吸信号尤其是小呼吸信号不会完全符合韩语韵律与停顿规律。目前的韩语韵律研究结果指出，音步之间或者韵律短语之间一般不会出现停顿，而如上图 3. 21 所示，在朗读诗歌中，因诗歌带来的情感或者个人风格不同，还会看到明显的小呼吸信号。

由此可以说，在韩语翻译版五言近体诗朗读中，格律特征优先于不同语言的韵律特征，就是说与其他文体相比，因近体诗格律特征比较严格，而在生理条件允许的情况下，能够支配不同语言的韵律节奏特征。

从具体数据上看，三种不同版本中韩语翻译版呼吸重置幅度平均值最高，朗读中胸呼吸的一级大呼吸重置幅度平均为 1.1，斜率为-0.4，二级小呼吸重置幅度平均为 0.5，斜率为-0.2，每联开始处气息强，而其他小句开始处气息则弱，也较为舒缓。基本上一级呼吸的幅度和斜率值都比二级呼吸的两倍，相比二者的呼气时长区别不大，分别为 2.7 秒和 2.5 秒。腹呼吸的一级大呼吸重置幅度平均为 1.1，斜率为-0.4，二级小呼吸重置幅度平均为 0.5，斜率为-0.2，呼吸时长分别为 2.7 和 2.7，如呼吸曲线与语图分析所示，胸、腹呼吸中调整大小呼吸的趋势比较一致，只是腹呼吸峰值前后流速极快，曲线又高又尖，从语音开始点其斜率会变低一些。

韩语节奏具有“等时性”，即随着韵律短语内部音节数的增加，音节时长要变短，使得保持韵律短语的时长（安秉燮，2010）。该特征构造韩语特有的

节奏。如图 3.11 所示，本次实验结果也充分反映了韩语节奏特征。进一步，韩语朗读中以调节呼吸重置幅度来实现韩语的节奏。

Correlations		°«Öï.-Öë°æ îãÑÖ½ûlâÊ« ĐØ°ôïû.ù¶Ê	°«Öï.-Öë°æ îãÑÖ½ûlâÊ« ĐØ°ôïû.Ê±³¤
°«Öï.-Öë°æ îãÑÖ½ûlâÊ« °ôïû.ù¶Ê	Pearson Correlation Sig. (2-tailed) N	1 122	.172 .059 122
°«Öï.-Öë°æ îãÑÖ½ûlâÊ« ĐØ°ôïû.Ê±³¤	Pearson Correlation Sig. (2-tailed) N	.172 .059 122	1 122

表 3.11 韩语翻译版五言近体诗呼吸幅度与时长相关

从功能上看，腹呼吸比胸呼吸来得早振幅也大，但是比胸呼吸结束得快，在语音信号结束前已经开始准备下一个呼吸。可以说，腹呼吸先给言语活动提供能量，因此平均值也比胸呼吸高一些。相比，胸呼吸比腹呼吸来得晚，而语音信号结束点和胸呼吸呼气结束点比较一致，腹呼吸已结束的时候，胸呼吸依然参与言语活动直到语音结束。可见，腹呼吸主要给言语活动能量，并扩大呼吸量而辅助胸呼吸，根据言语活动的不同长度和结构，胸、腹呼吸协助调整呼气方式而进行言语活动，腹呼吸提前结束并准备下次呼吸，胸呼吸继续完成所安排的言语活动。朗读韩语翻译版五言近体诗中，一般在韵律单位边界处不会出现明显的呼吸，或者胸、腹呼吸一起出现，不能说哪种呼吸对韵律单位更敏感，这一点是韩语朗读者的呼吸所具有的共同特征。与汉语诗歌呼吸特征不同的是，朗读韩语中，除了韵律句外，韵律短语边界也会对应小呼吸。

那朗读韩语五言近体诗中大小呼吸幅度和次数具体怎么安排？下图 3.22 和表 3.12 显示的是韩语翻译版的五言近体诗胸腹呼吸重置幅度分布以及聚类分析结果。

从下图中可以看到，五言近体诗重置频度分布基本符合正态分布，不管是哪种呼吸，重置幅度的分布则在 0.4-0.6 和 1.0-1.2 处分别出现了两个比较明显的高峰，这说明韩语翻译版五言近体诗的呼吸重置大致可以分成两级。上文已提到，这种频度分布可以从近体诗的呼吸节奏特点中得到解释。

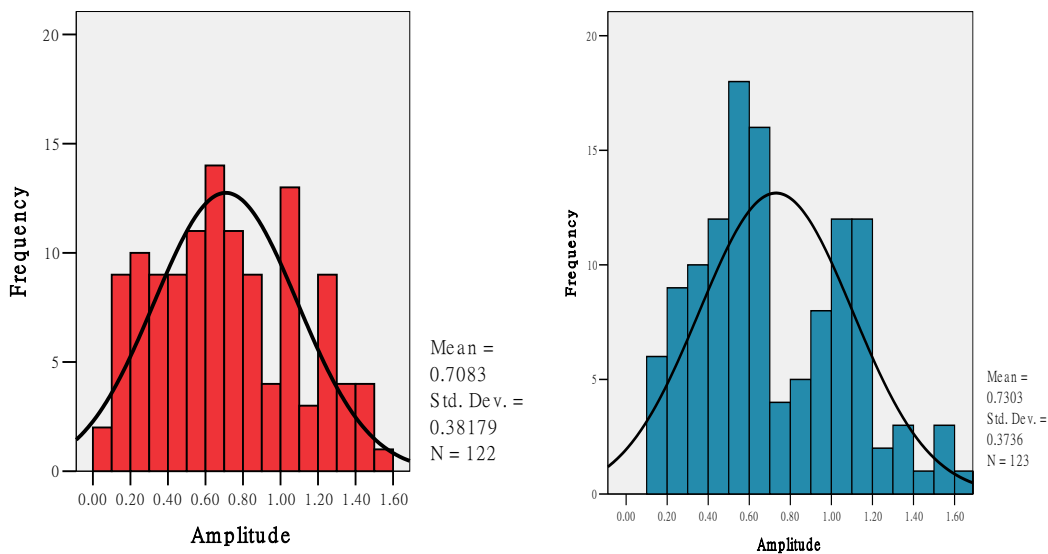


图 3.22 韩语翻译版五言近体诗呼吸重置分布图
(左图：胸呼吸，右图：腹呼吸)

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2		
°«Öi-Öë°æläÑÖ½üläÊ« ÐØ°öü·ù¶Ê	1.13	.47	Cluster 1	44.000
			Cluster 2	78.000
			Valid	122.000
			Missing	.000

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2		
°«Öi-Öë°æläÑÖ½üläÊ« ,¹°öü·ù¶Ê	1.15	.47	Cluster 1	55.000
			Cluster 2	68.000
			Valid	123.000
			Missing	.000

表 3.12 韩语翻译版五言近体诗呼吸重置聚类分析结果

上图 3.22 给出的是朗读韩语翻译版七言近体诗中的胸、腹呼吸重置分布，左图为胸呼吸重置分布图，右图则是腹呼吸重置分布图。韩语翻译版五言近体诗的胸、腹呼吸重置频度分布基本符合正态分布，腹呼吸幅度的平均值比胸呼吸的高一些，分别为 0.70 和 0.73。胸、腹呼吸重置分别出现了 122 处（每首律诗平均呼吸次数为 8 次、绝句平均为 4.2 次）和 123 处（每首律诗平均为 8.1 次、绝句平均为 4.2 次），等于韩语翻译版五言近体诗朗读中每诗行基本出现一次呼吸，如上文给出的呼吸曲线所示，胸、腹呼吸出现的次数基本一致。

从具体数据分析显示，三种不同版本中，韩语翻译版平均幅度最高，音节数与呼吸次数也最多，而严格遵守汉语近体诗格律，呼吸安排颇有规律。可以说，诗歌的特殊文体形式优先于语言结构特征，韩语朗读者根据韩语本身的节奏特征，调整呼吸方式而体现汉语格律诗的节奏。

从上表 3.12 聚类分析所示，胸、腹呼吸重置幅度都可以分成两级，其中心值也比较接近，胸呼吸一、二级分别为 1.13 和 0.47，腹呼吸为 1.15 和 0.47，这些特征印证了上文所探讨的呼吸曲线分析结果，即韩语翻译版五言近体诗朗读

中胸、腹呼吸表现比较一致。不同的是，从分布图与聚类分析所示，虽然胸、腹呼吸的呼吸次数相同，而二者的分布格局有所不同：即胸呼吸的二级呼吸幅度分布较散，小呼吸出现次数比大呼吸的多，分别为 44 次和 78 次，相比腹呼吸的重置幅度分布更集中，大小呼吸的两个峰值更明显，二者的呼吸次数之间相差不大，分别为 55 次和 68 次。由此可见，无论是汉语还是韩语，五言近体诗胸呼吸幅度变化多，尤其朗读韩语翻译版近体诗时，朗读者根据音节数比较自由地调整胸呼吸幅度，相比腹呼吸的大小呼吸分布比例更均匀一些。虽然朗读的文体不同，音节数与重置次数也增加了一些，但这些特征都没有变，可以说这是汉语近体诗格律特征所形成的五言近体诗朗读中呼吸方式的普遍性特征。

总之，朗读韩语翻译版五言近体诗中的呼吸特点是：

胸、腹呼吸都存在二级呼吸重置，表现方式也大体一致，呼吸曲线基本重合。一般不会出现长呼吸，因此每首诗的上联和下联开始位置都会出现一级呼吸，二级呼吸处于每小句的开始位置或者韵律短语边界处，大小呼吸有规律地交替出现。

与汉语版五言近体诗呼吸特征相比，朗读韩语近体诗呼吸重置最显著的不同点是胸、腹呼吸配合方式。虽然二者的呼吸都分成二级，腹呼吸平均重置幅度高于胸呼吸，重置开始时间早于胸呼吸，而在开始朗读韩语近体诗时，胸、腹呼吸有时间差而基本重合，直到语音结束前才会显示二者的不同功能：即胸呼吸持续至发音结束，腹呼吸提前结束并为下次言语活动准备足够气息量。

综合三种版本近体诗呼吸特征，发现了汉语近体诗格律的严格性制约不同结构语言的节奏特征。虽然呼吸配合方式不同，韩语朗读者调整呼吸重置幅度与次数以及朗读速度而体现汉语近体诗节奏特征。另外，值得注意的是，呼吸信号分析结果表明，韵律短语边界处也会出现小呼吸，与声学分析中的停顿时长无关。

（2）韩语翻译版七言近体诗的呼吸节奏特征

上文已提到，由于语言结构不同，在诗歌翻译时，两个汉字基本对应一两个音步。韩语翻译版五言近体诗呼吸特征分析显示，不管每一个小句包含几个音步，韩语朗读者会尽量遵守汉语近体诗格律，表现出颇有规律的呼吸模式。那由此再加一两个音步，等于加了一个韩语诗歌中的一个韵律句的话，其呼吸方式会有什么变化？下面我们继续探讨不同文体，不同语言结构对生理呼吸信号的影响。

下图 3. 显示的是汉语七言近体诗《戏为六绝句（一）》的韩语翻译版朗读中的胸、腹呼吸曲线和语音波形。图中红色曲线显示胸呼吸重置变化，蓝色曲线则是腹呼吸变化曲线，黑虚线是每小句语音信号的始点和终点。

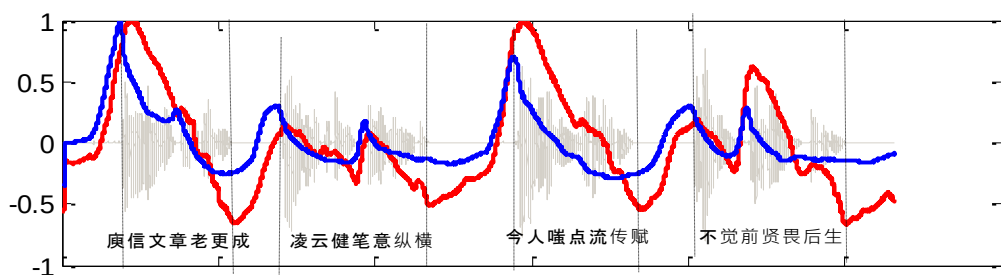


图 3.23 韩语翻译版七言近体诗朗读中胸、腹呼吸与语音波形

戏为六绝句（一）

庚信文章老更成，凌云健笔意纵横。

유신의 문장은 늙어 더욱 격조가 높아져, 구름을 넘는 듯 굳건하고 의미도 종횡부진하였다

今人嗤点流传赋，不觉前贤畏后生。

요즈음 사람들 전하는 부를 꼬집어 비웃지만, 먼저 이룬 사람이 후생을 두려워함은 깨닫지 못하네.

从方式上看，朗读中在大多数情况下，每联开始处具有一个大呼吸重置，随后每小句具有两个小呼吸重置，与汉语七言近体诗的呼吸格局比较一致。不同的是朗读汉语时，只有腹呼吸出现在韵律短语边界处，而韩语朗读时，胸、腹呼吸基本一起出现，只是小呼吸的胸、腹呼吸重置幅度大小不一，只看呼吸曲线变化，二级和三级呼吸的界限比较模糊。

值得注意的是，综合韩语翻译版七言近体诗呼吸方式，发现了小呼吸并不是生理需要而产生的，就是说不是朗读者一口气读不完才产生的。如，上图 3. 和列出的诗句所示，虽然每联的第一句和第二句各自包含的音节数和音步数基本一致，而在每联的第一句开始处一定会出现大呼吸，不会出现小呼吸，只在第二句会出现两个小呼吸，体现出比较固定的呼吸模式。

由此可以说，不管是哪种语言，五言还是七言诗，朗读者按照严格的近体诗格律来安排呼吸方式，体现近体诗的情感以及近体诗格律带来的节奏感，该特征具有一定的普遍性。

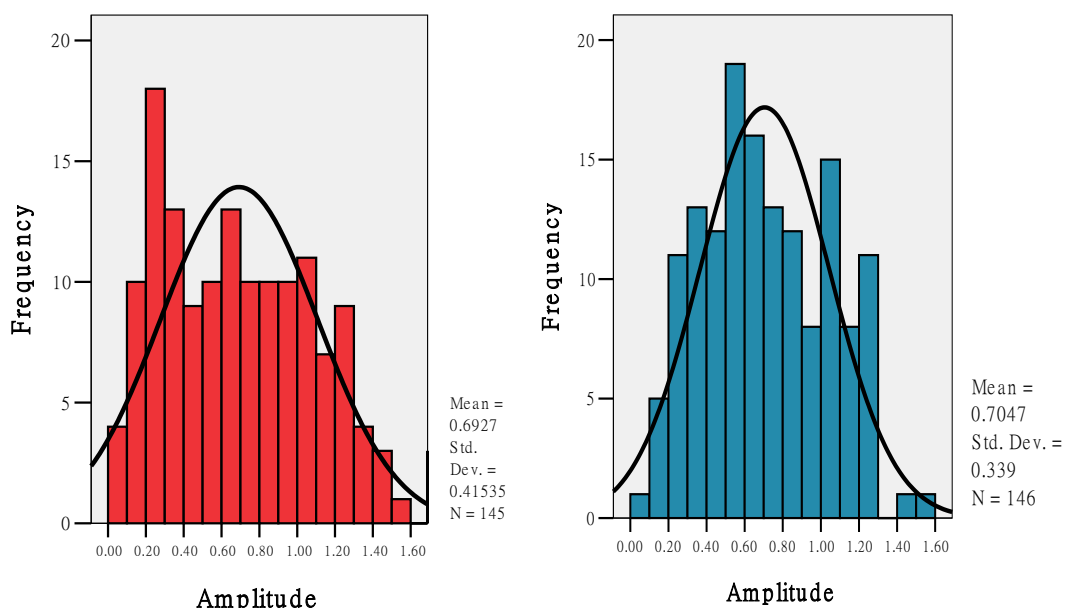


图 3.24 韩语翻译版七言近体诗呼吸重置分布图

（左图：胸呼吸，右图：腹呼吸）

Final Cluster Centers				Number of Cases in each Cluster	
	Cluster			Cluster	
	1	2	3		
°«Öï-Öë°æÆßÑÖ½üîäÊ« ÐØ°öîüÖØÖÃ-ùŋË	1.24	.74	.27	1	38.000
				2	52.000
				3	54.000
				Valid	144.000
				Missing	.000

Final Cluster Centers			Number of Cases in each Cluster	
	Cluster		Cluster	
	1	2		
°«Öï-Öë°æÆßÑÖ½üîäÊ« °¹öîüÖØÖÃ-ùŋË	1.10	.49	1	56.000
			2	90.000
			Valid	146.000
			Missing	.000

表 3. 13 韩语翻译版七言近体诗胸呼吸重置聚类分析结果

下面为了探讨韩语翻译版七言近体诗大小不同的呼吸重置幅度的频度分布，提取了 20 首汉语近体诗中所有的呼吸重置幅度，左图为胸呼吸重置分布图，右图则是腹呼吸重置分布图。从图 3.24 与表 3. 13 给出的聚类分析中可以看出，朗读韩语翻译版七言近体诗中胸呼吸重置大体分成三级，腹呼吸重置分成二级。从呼吸曲线已看出，七言近体诗呼吸比五言诗的呼吸方式复杂一些，尤其二级呼吸重置大小不同，如图 3.24 所示，胸呼吸的分布比较散，除了三级呼吸有明显的峰值，一级与二级呼吸分布次数比较均匀。与它相比，腹呼吸重置幅度有两个较明显的峰值，分别出现在 0.5-0.7 和 1.0-1.1 处。如图 3.24 所示，就像韩语翻译版五言近体诗呼吸重置分析结果一样，腹呼吸平均幅度还是高于胸呼吸，从五言到七言，胸、腹呼吸次数都增加了许多，随之平均幅度也会变低了一些，胸呼吸从 0.70 降到 0.69，基本一致，而腹呼吸从 0,73 降到 0,70，比胸呼吸的变化大一些。

而胸腹呼吸重置次数分别为胸呼吸 122 次变成 145 次（每首律诗平均呼吸次数为 9.5 次、绝句平均为 5.1 次）、腹呼吸 123 次变成 146 次（每首律诗平均为 9.4 次、绝句平均为 5.2 次），基本上比每首五言近体诗中出现的呼吸重置次数要多了一次。上文已分析过，汉语版朗读中，无论是重置幅度还是次数，从五言到七言，胸腹呼吸重置幅度与次数的变化趋势比较一致，朗读七言诗中胸、腹呼吸的幅度都降低了 10%，相反呼吸次数两个都增加了 25%，很有规律。与它相比，韩语翻译版七言近体诗朗读时，胸呼吸幅度基本没有变化，腹呼吸幅度降低了 5%左右，胸、腹呼吸的次数增加了 20%。可以说，汉语和韩语版呼吸次数变化趋势是相似的，只是韩语近体诗朗读中，文体或者不同结构的变化对呼吸幅度和次数变化程度要比汉语近体诗低一些。尤其对韩语翻译版胸呼吸重置平均幅度几乎没有影响。

由此可见，朗读韩语近体诗中，增加了音节而韵律结构出现变化时，根据不同结构，胸呼吸会安排不同大小的呼吸幅度体现近体诗的节奏，结果胸呼吸就分成三级呼吸。与此相比，随着小呼吸的增加，腹呼吸平均幅度也会变低，而大小呼吸的固定比例不会改变。

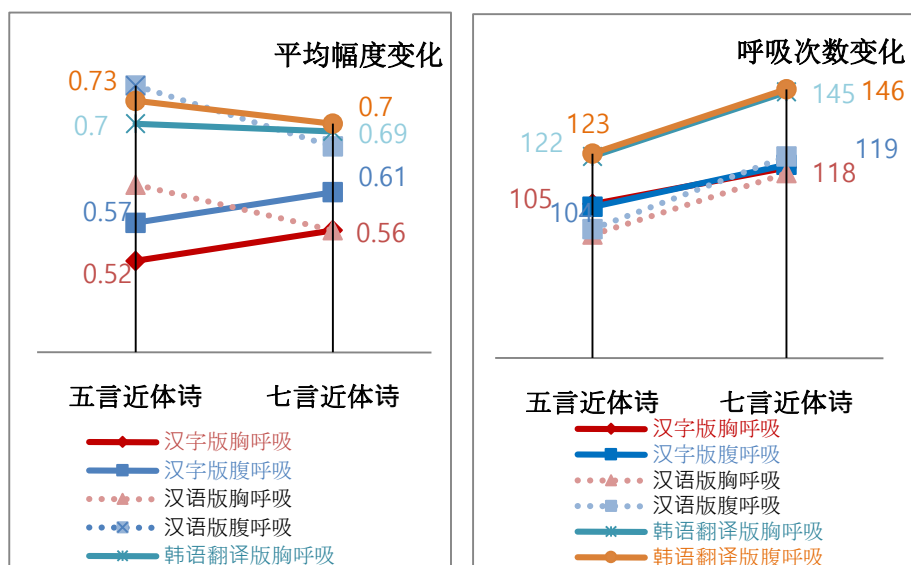


图 3.25 汉语版与韩语翻译版近体诗朗读中胸、腹呼吸幅度与次数变化

上图 3.25 所示的对比情况以及上文已分析过的汉语近体诗朗读中的呼吸特征来看，朗读不同文体的时候中韩朗读者之间最显著的差异是调整呼吸的方式。

可见，语言结构变化与呼吸幅度变化方式有密切的关系，与谭晶晶（2008）的结果一样，呼吸重置幅度对不同文体（结构）的影响比较大。

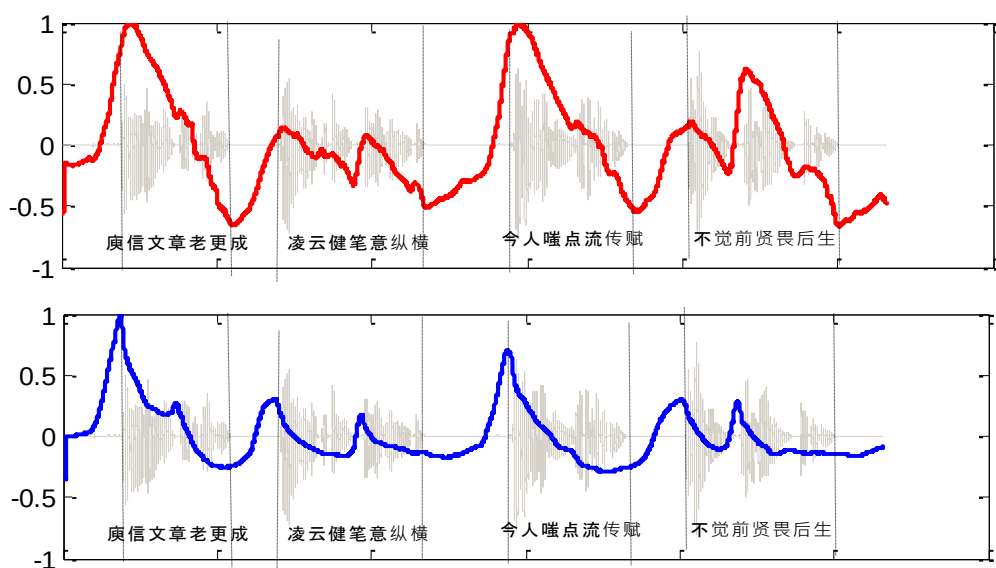


图 3.26 韩语翻译版五言绝句朗读中胸、腹呼吸与语音波形
（上图：胸呼吸与语音波形，下图：腹呼吸与语音波形）

从数据来看，朗读韩语翻译版七言近体诗中胸呼吸的一级大呼吸重置幅度平均为 1.2，斜率为-0.4，二级中呼吸重置幅度平均为 0.7，斜率为-0.2，三级小呼吸重置幅度平均为 0.3，斜率为-0.2，大呼吸开始处气息强，斜率也高，而小呼吸则气息弱，也较为舒缓，时长为二三级斜率比较一致，三级呼吸的时长分别为 3.0、2.8 和 2.6 秒，三个级别之间相差不大，如下图 3.14 分析显示，韩语翻译版七言近体诗重置幅度与时长之间的相关系数为 0.155，没有显著性，从

而再次印证了韩语节奏的等时性特征，充分反映韩语诗歌本身的规律。

腹呼吸的一级大呼吸重置幅度平均为 1.1，斜率为-0.3，二级小呼吸重置幅度平均为 0.5，斜率为-0.3，因为不同幅度以同样的斜率呼气，从而二者的时长之间的差距比较大，分别为 2.6 和 2.1 秒。

Correlations			
		°«Öÿ.-Öë°æ ÆßÑÖ½üläÊ« ÖØÖÃ·ù¶Ê	°«Öÿ.-Öë°æ ÆßÑÖ½üläÊ« ÖØÖÃÊ±³□
°«Öÿ.-Öë°æ ÆßÑÖ½üläÊ« ÖØÖÃ·ù¶Ê	Pearson Correlation	1	.155
	Sig. (2-tailed)		.070
	N	138	138
°«Öÿ.-Öë°æ ÆßÑÖ½üläÊ« ÖØÖÃÊ±³□	Pearson Correlation	.155	1
	Sig. (2-tailed)	.070	
	N	138	138

表3.14 韩语翻译版七言近体诗呼吸幅度与时长相关

从功能上说，腹呼吸重置时间点早于诗句语音起始点时间，在一般情况下比胸呼吸结束得更快，在语音信号结束前已经开始慢慢吸气，或者保持平线并准备下一次言语活动。如图 3.26 下图所示，朗读第二联第一句“요즈음 사람들 전하는 부를 꼬집어 비웃지만（今人嗤点流传赋）”中，到了最后一个音步“비웃지만”开始处腹呼吸已呼到最低点并开始慢慢吸气，朗读第二联第二句“먼저 이룬사람이 후생을 두려워함은 깨닫지 못하네（不觉前贤畏后生）中第二个腹呼吸到了小句的中间“두려워함은 깨닫지 못하네”处呼气已结束而曲线保持平线。

相比，胸呼吸呼气起点和终点都与语音信号比较吻合，呼气结束点与诗句语音结束点一致。由此我们可以知道，胸呼吸和言语活动更有直接的关系，结合上文所分析的汉语近体诗朗读中的胸呼吸特征，这点是胸呼吸具有的普遍特征之一。但是，如上图 3.25 所示，在朗读中的胸、腹呼吸曲线和轨迹比较接近，而且有时候胸、腹呼吸都会单独出现在韵律短语之间，不像汉语版一样，二者的分工比较明确，因此不能说哪种呼吸主要担任哪种功能。

- 由此可以总结，朗读韩语翻译版近体诗中的呼吸重置特点是：
1. 腹呼吸起始点与峰值均早于胸呼吸，重置幅度高于胸呼吸，在开始朗读时给胸呼吸以及言语活动提供能量，比胸呼吸早点结束并开始准备下次呼吸；一开始朗读，胸呼吸晚于腹呼吸，其结束时间点和言语信号终点比较一致，可以说虽然胸、腹呼吸的曲线比较重合，而胸呼吸对言语活动有更密切的关系。
 2. 不管朗读五言近体诗还是七言近体诗，腹呼吸存在二级呼吸重置，一级呼吸处于每首诗的每联的始点，二级呼吸处于每联第二句以及韵律短语边界处。
 3. 胸呼吸根据不同类型的近体诗，存在三级呼吸重置，一级呼吸处于每首诗的每联开始位置，二级呼吸处于每长句的开始位置，三级呼吸处于长句末尾小句。每长句对应一个胸呼吸重置，韵律边界处出现呼吸中段抖动。
 3. 从整体呼吸安排来说，汉语近体诗格律优先于不同语言节奏特征，而呼吸重

置幅度变化以及相关分析结果还显示出韩语诗歌本身的节奏特征，韩语音节结构所具有的“等时性”特征比较明显。

4. 朗读韩语近体诗时，呼吸安排决定于格律和朗读者对诗歌内容的不同解释，不在于生理上的需求（生理呼吸）。

5. 从呼吸信号曲线的分析发现，声学分析上的停顿与否和生理呼吸信号往往没有关系，值得进一步研究。

6. 在朗读汉语近体诗中，汉语近体诗的不同类型对胸腹呼吸的重置幅度和次数以及腹呼吸的级别有影响。

3.3 结果

本章基于呼吸信号主要从呼吸节奏与韵律表现的角度考查具有不同语言结构的汉语和韩语近体诗朗读中的呼吸节奏特征的不同，并试图探讨呼吸节奏和韵律单位的对应以及不同语言结构对呼吸节奏的影响。通过本章的分析，我们发现从宏观上看，因为汉语近体诗的格律严格，形式固定，从而不管哪种语言哪种形式，朗读近体诗中整体呼吸格局有一定的一致性：朗读近体诗中胸、腹呼吸基本分成二级，大小呼吸交替出现，每联的第一句一定会出现大呼吸，在大多数情况下，第二句和第四句开始处具有小呼吸，第二联开始处有时具有大呼吸，有时会出现小呼吸。与之前的呼吸研究结果最显著的不同点是，不管是哪种语言的呼吸曲线，韵律短语边界处不仅会出现抖动，还会出现明显的小呼吸。有趣的是，这些小呼吸的出现的位置往往与声学分析中的停顿的位置和时长无关，有待进一步考查。

从胸、腹呼吸配合的角度看，汉语版和韩语版（汉字版与翻译版）朗读中腹呼吸起始点都早于胸呼吸，重置幅度高于胸呼吸，在开始朗读时给整个言语活动提供能量，有辅助胸呼吸的作用，而提前结束并准备下一次呼吸和言语活动；一开始朗读，胸呼吸晚于腹呼吸，胸呼气开始点与终点与言语信号时间点比较一致。可以说，这些特征具有普遍性。

不同语言版本之间主要差别在于：朗读汉语近体诗时，胸、腹呼吸分工更加明确，就是说在开始朗读时给整个言语活动提供能量，比胸呼吸早点结束而朗读中对韵律单位的界限更敏感，韵律短语边界处常出现腹呼吸，从而呼吸次数比胸呼吸多一些；呼气之后保持一定时间使言语活动持续下去，因此胸呼吸的终点和言语信号终点基本吻合，胸呼吸与整个言语活动的完成有关。相比，朗读韩语版近体诗中，韩语版近体诗朗读中胸、腹呼吸下降趋势比较一致，二者的分工比汉语版呼吸特征模糊。

从五言近体诗到七言近体诗，汉语版和韩语版（汉字版和翻译版）呼吸变化趋势也截然不同：汉语版五言近体诗朗读中胸、腹呼吸都存在二级呼吸，而七言近体诗朗读中，因小呼吸次数增加而平均幅度偏低，腹呼吸具有三级呼吸。汉字版五言近体诗朗读中的呼吸特征与汉语版的比较一致，而七言近体诗朗读中，增加的音节数对韩语朗读者对呼吸安排的影响不够大，结果呼吸次数增加，幅度也增加，而并没有出现显著变化。韩语翻译版近体诗的音步数已远远出韩语诗歌本身结构中的音步数和音节数，但朗读五言近体诗中，胸、腹呼吸的格局颇有规律，均分成二级呼吸。韩语版七言诗朗读中，因为增加的两个音步影响到韩语朗读者的呼吸安排，结果像汉语版的变化一样，随着小呼吸次数的增

加，平均呼吸幅度也降低了一些。与汉语版呼吸变化不同的是，韩语版胸呼吸重置大小不一而变成三级呼吸，腹呼吸比较稳定。由此可见，汉语近体诗格律的严格性制约不同结构语言的节奏特征。虽然呼吸配合方式不同，但韩语朗读者调整呼吸重置幅度、次数和朗读速度而体现汉语近体诗格律特征。

从呼吸幅度与时长的相关关系上，汉语和韩语之间存在显著性差异：汉语版近体诗和汉字版近体诗朗读中，呼吸幅度与时长之间的相关分析结果比较一致，即五言近体诗朗读中，幅度与时长之间成正比，存在显著性相关。与此相比，韩语版近体诗朗读中，不管是五言还是七言近体诗，不同级别的呼吸时长之间差别不明显，朗读者按照固定的呼吸模式，调整整个呼吸安排体现汉语近体诗格律，结果相关分析显示，幅度与时长之间不存在显著性相关。

可见，此结果既可以显示汉语近体诗格律的特征，也可以表现韩语音节结构的特征“等时性”。从整体呼吸安排来说，汉语近体诗格律优先于不同语言节奏特征，不管是哪种语言，不管是五言诗还是七言诗，朗读者按照严格的近体诗格律来安排呼吸方式，体现近体诗的情感以及近体诗格律带来的节奏感，该特征具有一定的普遍性。这还说明，呼吸安排决定于格律和朗读者对诗歌内容的不同解释而产生的情感，基本上不在于生理上的需求。

第四章 汉语和韩语新闻的呼吸与韵律特征对比

本文选取了 20 篇的新闻语料,进行了语音和呼吸信号的录制,新闻的文章结构有其固定的格式,是典型的陈述语体,通常没有特别的强调重音和特殊的语气,发音人在朗读新闻时也容易形成相对固定的呼吸节奏模式。因此,新闻语料特别适用于考察节奏的基本特性和呼吸信号(李永宏等,2010)。

对话篇的韵律特征的研究也已经提上了日程。王蓓等(2005)的研究探讨了语篇中大尺度韵律单元的声学线索,郑秋豫(2005)也明确提出口语语流单位为多短语句群,而非单一语句或短语,她还提出,多短语句群受到自上而下的管辖制约,人们在说话时除了受到发声器官和生理机制的制约外,还要受到认知能力的制约,也就是说,人们在说话之前要对所说的内容做一个规划,划分韵律单元的时候应该也会考虑到内容的完整性和语法的制约。

基于上述的认识,本次实验分析了较典型的新闻语料的呼吸节奏,每篇都是对某一事件的完整报道,每篇语料大约 250 字,一般分为 2-4 个自然段。利用这样的语料,我们可以探讨语篇的呼吸规划和韵律结构,也可以考察人们在表达某一个完整而相对复杂的内容时,如何规划呼吸节奏。

掌握“停顿”对流畅自然的汉语表达具有重要意义。参考许毅(1986)、华武(1998)、叶军(2001)、曹剑芬(2001)、吴洁敏(2001)、刘现强(2007)等论述,汉语停顿可归为三类:

- (1) 说话时因生理需要而停顿的“生理停顿”;
- (2) 由言语结构决定的、体现出言语结构层次的“语法停顿”;
- (3) 为了强调某一事物或突出表达某种感情,而改变语法停延位置或时值的“强调停顿(心理停顿)”。

李爱军(1999)、曹剑芬(2001)、钱摇等(2001)、熊子瑜(2001)等在各自的研究中都涉及到了对停延声学特征的考查,归纳起来,普通话停顿的声学表现有以下几点:(1) 无声段的出现;(2) 停顿前言语片段的相对延长,这里主要指停顿前最后一个音节的时长变化,主要是韵母的延长;(3) 停顿前后音节的频率值的调整,即停顿前一音节和停延后一音节的基频 F0 重置(张延成,2013)。可见,决定停顿出现位置的因素有很多,或是根据语法或是根据语义、或是由说话者的气息决定,此外,说话者对意群的划分直接影响停顿的出现。同时,为了清楚地表达语言的结构层次,说话者会使用语法停延,这是语音链上出现频率最高的一种停延(张延成,2013)。

根据汉韩语停顿和呼吸节奏研究结果,汉语每 9-10 个音节出现一次停顿,大概每 15 个音节出现一次呼吸重置(谭晶晶,2008)。韩语则每 13 个音节或每两个韵律短语出现一次停顿(李浩英,1996;金成硅,1999),而韩语朗读时以什么样的规律安排呼吸节奏目前尚未有相关研究。

因此,本文基于生理呼吸信号,将探讨汉韩语中新闻朗读的生理呼吸信号如何对应声学信号,以及表现是否一致等问题。下面我们将一一进行探析。

4.1 汉语新闻的呼吸与韵律特征

新闻的文章结构有其固定的格式，发音人在朗读新闻时也相对容易形成固定的呼吸节奏模式。在本次试验中，我们分析了汉语发音人朗读的 20 篇新闻的呼吸节奏，朗读中语音信号以及胸腹呼吸曲线表现情况如下图所示：

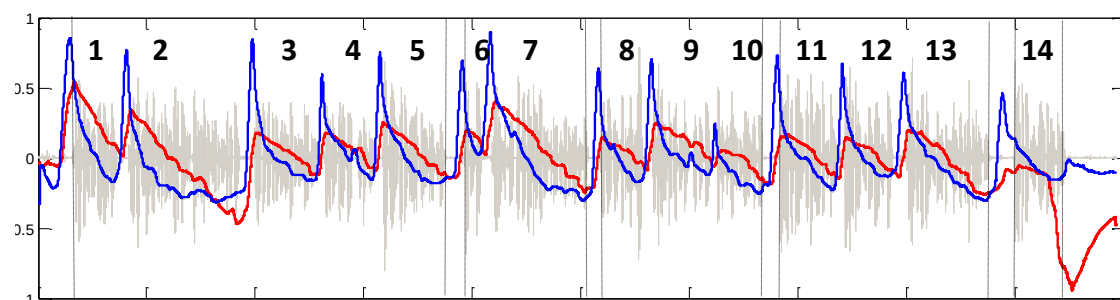


图 4.1 汉语新闻朗读中胸、腹呼吸与语音波形

文本：(1)新华社北京 3 月 6 日电：(2)十一届全国人大四次会议 6 日下午在人民大会堂举行记者会，(3)国家发展和改革委员会主任张平、(4)副主任朱之鑫、(5)徐宪平就“十二五”规划纲要回答记者提问。(6)张平说，(7)“十一五”时期我国经济社会发展取得了举世瞩目的成绩。(8)同时，我们清醒地认识到，(9)我国发展中不平衡、不协调、(10)不可持续的问题依然突出。(11)在编制“十二五”规划纲要过程中，(12)针对这些问题做了深入研究，(13)期望在今后五年的发展进程中解决好这些问题。(14)这也就构成了“十二五”规划的特点。

上图 4.1 中的红色曲线显示胸呼吸重置变化，蓝色曲线则是腹呼吸变化曲线，黑色虚线是每小句语音信号的始点和终点。从方式上看，汉语新闻朗读中腹呼吸主要以大呼吸为主，而胸呼吸重置幅度比腹呼吸显得低一点，即胸腹呼吸重置幅度之间的差异比较明显。在大多数情况下，胸呼吸的呼气开始点与新闻语音开始时间吻合，腹呼吸重置时间点与峰值所在位置早于语音开始时间与胸呼吸，其呼气信号在语音信号结束前结束，紧接着开始下一次呼吸或者保持平线。综合上文所分析的近体诗呼吸特征，这是不管是汉语还是韩语，朗读中胸、腹呼吸各自具有的普遍性特征。

与上文已分析的近体诗呼吸重置特征最明显的差异是，朗读新闻时大小呼吸没有交替出现，而根据朗读内容和发音人的安排，比较自由地调整呼吸重置幅度和时长。

呼吸重置与下列的文本结合看，我们容易发现汉语新闻朗读中呼吸重置位置与表示停顿的标点符号出现的位置基本一致，只有两个例外：一个是在第三个分句开始处“同时”后的“，”，另一个是“我国发展中不平衡”后的“、”。由此可以看出，朗读汉语新闻时，标点符号等于呼吸重置，但是如果标点符号之间的音节数不够多并且前后呼吸重置的距离非常短的话，其标点处不出现呼吸重置。此外，在其他新闻朗读文本中出现的例外只有一类，即其呼吸重置出现在主语和谓语之间。如，“在中国与东盟贸易中位居前三位的东盟国家是马来西亚、泰国、新加坡。”无其他例外，也没有个人差异。可见，朗读汉语新闻时，呼吸重置安排受标点符号的制约。

在本次实验结果中出现的与呼吸重置有关的主要标点符号有句号（。）、逗号（，）、顿号（、）和冒号（：）。它们的基本用法如下：

1. “句号”用于句子末尾，表示陈述语气。使用句号主要根据语段前后有较大停顿、带有陈述语气和语调，并不取决于句子的长短。

2. “逗号”用于复句内各分句之间，以及各种语法位置。例如，较长的主语之后、句首的状语之后、较长的宾语之前、较长的主语中间、谓语中间或宾语中间、某些序次语（“第”字头、“其”字头及“首先”类序次语）之后，以及用顿号表示较长、较多或较复杂的并列成分之间的停顿时。

3. “顿号”主要用于并列词语之间。

4. “冒号”用于总说性或提示性词语之后，表示提示下文。

《标点符号用法》（1995）这本书给的标点符号的意义：句子，前后都有停顿，并带有一定的句调，表示相对完整的意义。句子前后或中间的停顿，在口头语言中，表现出来就是时间间隔，在书面语言中，就用标点符号来表示。

与本次实验结果结合来讲，标点符号是书面语言的有机组成部分，是书面语言不可缺少的辅助工具，从而可以帮助人们确切地表达思想感情和理解书面语言。因此，朗读汉语新闻中呼吸重置出现在标点处是合理的。人们一般是倾向于把一个完整的语义放在一个韵律单元里实现的，在说话时人们可以通过调节呼吸节奏保证语义表达的完整性。

由此，值得我们注意的是，一直以来大多数关于韵律以及停顿的研究提出，在一般情况下，句中停顿有两个目的：换气和表意。换气停顿有的叫做气息停顿，句子太长，为了读得从容，自然要运用换气的技巧，长句子如果不在中间换气，一口气读下来的话，势必打乱朗读者的节奏，影响后面的正常发挥。而正确的运用了换气技巧，一方面朗读者读得从容自然，听来也会觉得轻松顺畅。然而，从本次实验结果显示，在大多数情况下，句中基本不存在所谓朗读者一口气读不完而发生的换气停顿，也叫“生理停顿”，都以标点符号来替代该功能，因此主要是由标点符号导出的“语义停顿”。可以说，汉语标点符号支配朗读汉语新闻中的呼吸重置安排，从而保证新闻信息有效地传达，这也是一种汉语独有的呼吸节奏特征。

对本实验分析的呼吸曲线进行对照，我们可以发现三级呼吸单元分别和不同大小的语义单元相对应。大呼吸单元对应整段话，也就是对应着最大的意义单元，表达了一个最完整的意思，下图 4.2 显示，在朗读汉语新闻中出现的大呼吸，无论是胸呼吸还是腹呼吸，在每篇新闻的开始处，呼吸信号会较快地上升到这个呼吸单元的最高处，第一出现的大呼吸往往会持续到一个句子的结尾。

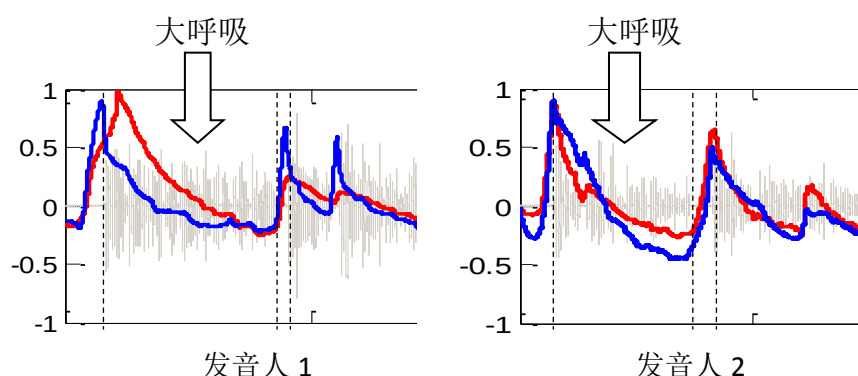


图 4.2 在朗读汉语新闻中出现的大呼吸实例

而下图 4.3 所示的胸呼吸曲线不符合该特征，那是与第一个胸呼吸重置出现的新闻的特殊结构有关。不管是那种语言，新闻的写作有其固定的模式，一般说来新闻的第一段都是用一句话交代出某一新闻事件的概况，后面再展开叙述。如，第一个胸呼吸对应的文本是“（1）新华社北京 3 月 6 日电”，由 10 个音节组成的提示句，它对应一个中呼吸。一般而言，每个中呼吸单元对应一个分句，即一个中等的语义单元，小呼吸主要对应于分句中的某个句子成分，即并列词语或者句首出现的状语等。比如，“（4）副主任朱之鑫”和“（6）张平说”等。就是说，和上文已分析的近体诗呼吸重置特征一样，胸呼吸重置的峰值和语音起点吻合，根据标点符号的位置和朗读内容的安排，发音人会调整呼吸重置方式。

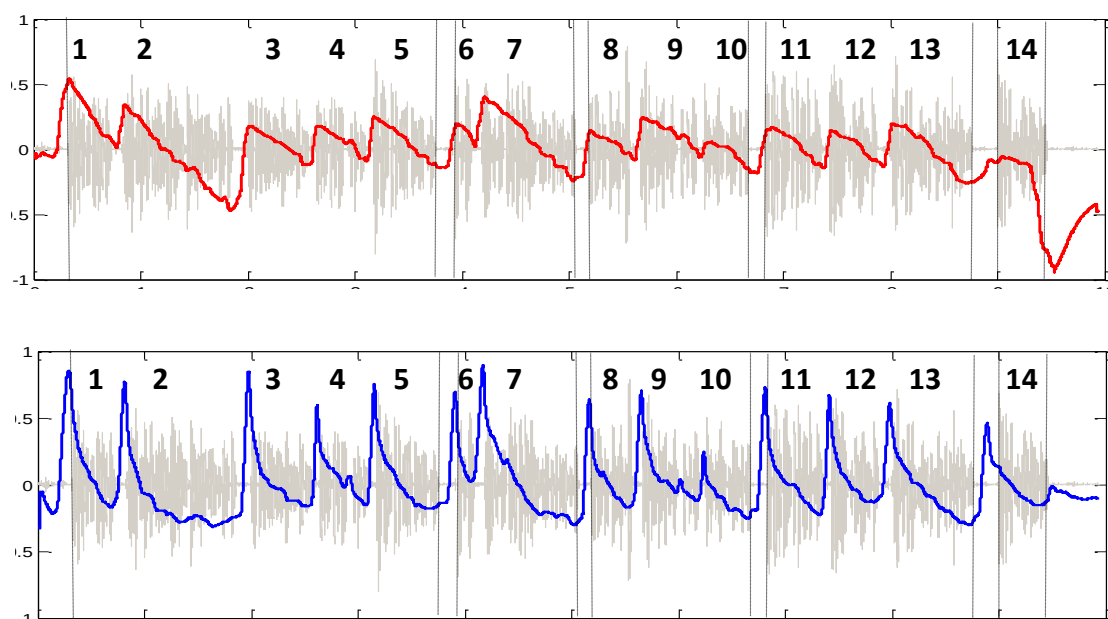


图 4.3 汉语新闻朗读中胸、腹呼吸与语音波形
（上图：胸呼吸与语音波形，下图：腹呼吸与语音波形）

文本：(1)新华社北京 3 月 6 日电：(2)十一届全国人大四次会议 6 日下午在人民大会堂举行记者会，(3)国家发展和改革委员会主任张平、(4)副主任朱之鑫、(5)徐宪平就“十二五”规划纲要回答记者提问。(6)张平说，(7)“十一五”时期我国经济社会发展取得了举世瞩目的成绩。(8)同时，我们清醒地认识到，(9)我国发展中不平衡、不协调、(10)不可持续的问题依然突出。(11)在编制“十二五”规划纲要过程中，(12)针对这些问题做了深入研究，(13)期望在今后五年的发展进程中解决好这些问题。(14)这也就构成了“十二五”规划的特点。

朗读汉语新闻中，腹呼吸与胸呼吸的重置特征截然不同。如图 4.3 下图所示，虽然胸腹呼吸出现的位置基本一致，而朗读汉语新闻时腹呼吸最显著特征是其重置幅度又高又尖，显得大呼吸和中呼吸的频率极高，而且大呼吸除了句子之外，还会出现在句首状语，如“（6）张平说”。就是说，朗读新闻中，不管是整个句子还是分句，腹呼吸的大呼吸和中呼吸都比较自由地出现，小呼吸主要对应于并列词语、状语或者分句，但是不能确定其规律性。该特征很可能

与腹呼吸功能和文体的特殊性有关。

从功能上看，腹呼吸比胸呼吸来得早幅度也大，但是比胸呼吸结束得稍微快，与朗读近体诗中的呼吸特征比较一致，只是没有近体诗朗读中的分工明确。由于新闻主要有几个大段落组成，信息传递功能比较强，因此发音人需要相对较强而长的呼吸重置来完成一段话。朗读汉语新闻时，胸呼吸和腹呼吸配合进行，而腹呼吸给言语活动提供较大的能量，呼气之后保持一定时间使言语活动持续下去，胸呼吸与整个言语活动的完成有关，这点与朗读近体诗中的呼吸重置特征比较相似。

那胸腹呼吸的不同功能和文体是否影响呼吸重置具体安排方式？

下面我们提取了每个发音人朗读的 20 篇汉语新闻中胸呼吸共 454 处、腹呼吸共 481 处的呼吸重置幅度，并将其画成了频度分布直方图：

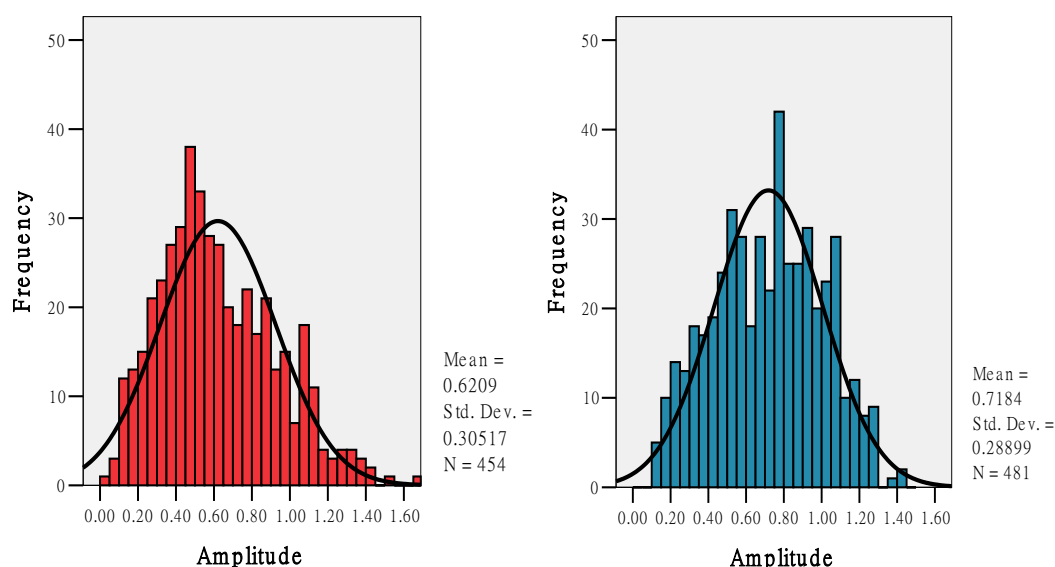


图 4.4 汉语新闻呼吸重置分布图
(左图为胸呼吸；右图为腹呼吸)

从图 4.4 中我们可以看到，汉语新闻的胸腹呼吸重置频度分布基本符合正态分布，而由于胸呼吸重置幅度的平均值比腹呼吸的小，胸呼吸重置幅度的平均值为 0.62，腹呼吸平均值为 0.71，因而胸呼吸呼吸重置分布稍微偏左一点。基本上，胸呼吸重置幅度分布主要集中在小呼吸，大峰值在 0.3-0.4 周围，相比腹呼吸重置幅度分布在 0.8 处出现明显的大峰值，0.6-1.2 周围也有几个峰值，可见腹呼吸重置幅度分布集中在中呼吸和大呼吸上，图 4.2 也直观显示胸腹呼吸重置幅度之间的差异。

因此，与汉语近体诗呼吸重置特征结合来看，不管是韵文还是非韵文，朗读汉语中的腹呼吸重置幅度总是大于胸呼吸，胸呼吸重置中小呼吸较多，腹呼吸重置中不同大小呼吸的分布相对比较均匀，可以说这些特征都是朗读汉语中出现的普遍性特征，没有个人差异。

如图 4.4 所示，腹呼吸重置出现的次数比较多，但是从每段新闻的具体情况来看，每段新闻中胸呼吸平均出现 12 次，腹呼吸平均出现 13 次，胸腹呼吸出现的次数基本一致，腹呼吸重置次数会稍微多一点。由于腹呼吸幅度比较大，有时虽然胸腹呼吸一起出现，而胸呼吸重置曲线只表现出小呼吸或者一个抖动，结果胸呼吸重置总次数少于腹呼吸重置。例如，从下图 4.5 所示，朗读完整的

一句话时，腹呼吸为了保证完成言语活动，分句边界处再次出现，给整个言语活动提供能量，而胸呼吸的大呼吸来完成一个完整的句子，在分句边界处只出现一个小抖动。也可以说，胸呼吸与完成言语活动，即语音信号关系更密切，腹呼吸对韵律结构更为敏感。总之，在朗读新闻时胸腹呼吸各有不同功能，二者的配合是有一定规律的。

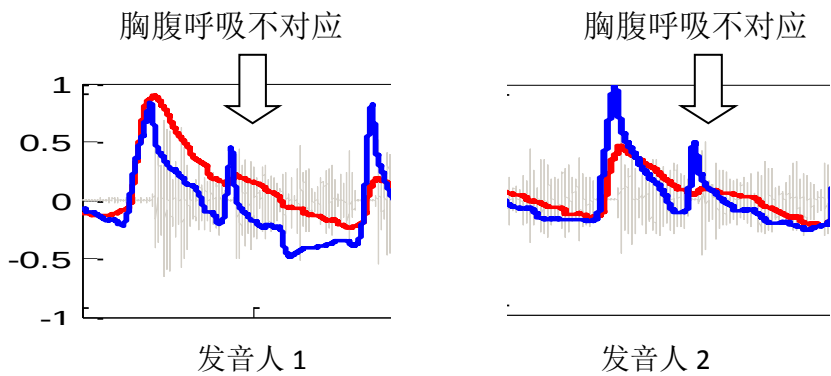


图 4.5 朗读汉语新闻中胸腹呼吸不对应实例

如下表 4.1 所示，呼吸重置聚类分析结果显示胸呼吸和腹呼吸重置都分成三级，不同级别的胸呼吸呼吸重置幅度中心值分别为 1.1、0.7 和 0.36，腹呼吸的分别为 1.07、0.74 和 0.39，可见胸腹呼吸重置幅度的中心值之间区别不大，胸腹呼吸重置最显著的差别在于不同大小呼吸重置安排的不同。如上图 4.4 和下表 4.1 聚类分析结果已显示，胸呼吸重置中三级呼吸重置次数最多，胸呼吸重置共 454 次，三级呼吸次数高达 209 次，其次是二级呼吸，160 次，次之一级呼吸，85 次。与此相比，表 4.1 显示腹呼吸重置中二级呼吸次数最多，共 481 次中出现 182 次，其峰值非常明显。一级和三级呼吸分别为 159 次和 140 次，三种不同级别的呼吸重置频度相对比较分散。可见，聚类分析结果再次证实朗读汉语新闻中的呼吸重置特征。

Final Cluster Centers				Number of Cases in each Cluster	
	Cluster			Cluster	
	1	2	3		
∞ÓiÐÂîÂÐØ°ð				1	85.000
îüÖØÖÃ·ù¶Ë	1.10	.71	.36	2	160.000
				3	209.000
				Valid	454.000
				Missing	.000

Final Cluster Centers				Number of Cases in each Cluster	
	Cluster			Cluster	
	1	2	3		
∞ÓiÐÂîÂ¹°ð				1	159.000
îüÖØÖÃ·ù¶Ë	1.07	.74	.39	2	182.000
				3	140.000
				Valid	481.000
				Missing	.000

表 4.1 汉语新闻呼吸重置幅度聚类分析结果
(上表：胸呼吸，下表：腹呼吸)

从数据来说，朗读汉语新闻中胸呼吸的一级大呼吸重置幅度平均为 1.1，二级中呼吸重置幅度平均为 0.7，三级小呼吸重置幅度平均为 0.4。开始处大呼吸气息较强，小呼吸则气息较弱，也较为舒缓，时长为 3.5、3.2 和 3.0，随着幅度平均值变小，时长也变短。如表 4.2 所示，汉语新闻朗读中胸呼吸重置幅度与时长呈显著正相关，与之前的研究结果一致（谭晶晶，2008；杨锋 等，2013）。

而腹呼吸的情况有所不同，腹呼吸重置幅度与时长无显著相关。从上文已分析的腹呼吸曲线可以发现，发音人安排腹呼吸时，不好归纳呼吸重置和句子成分的对应，例如不管是三个音节的状语成分还是较长的分句，都安排大呼吸来完成。可以说，该特征也是朗读汉语中的胸腹呼吸之间的不同之处。

Correlations			
		胸呼吸重置幅度 重置幅度	胸呼吸重置幅度 重置幅度
胸呼吸重置幅度 重置幅度	Pearson Correlation	1	.189**
	Sig. (2-tailed)		.007
	N	437	437
腹呼吸重置幅度 重置幅度	Pearson Correlation	.189**	1
	Sig. (2-tailed)	.007	
	N	437	437

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations			
		腹呼吸重置幅度 重置幅度	腹呼吸重置幅度 重置幅度
腹呼吸重置幅度 重置幅度	Pearson Correlation	1	.123
	Sig. (2-tailed)		.061
	N	462	462
胸呼吸重置幅度 重置幅度	Pearson Correlation	.123	1
	Sig. (2-tailed)	.061	
	N	462	462

表 4.2 汉语新闻朗读中的呼吸重置幅度和时长平均和相关
（上表为胸呼吸；下表为腹呼吸）

因此，我们为了了解朗读汉语新闻中的声学信号和生理信号的对应关系，根据语音信号提取了两个发音人的新闻中所有的停顿，又根据已提取的呼吸重置幅度值，分别统计出他们的平均值，如下表4.2所示。

	生理信号		声学信号
	胸呼吸重置次数	腹呼吸重置次数	停顿数
发音人1	12	13	27
发音人2	13	13	26
平均值	12.5	13	26.5

表 4.3 每篇汉语新闻出现的呼吸重置与停顿数比较

从表中我们可以看到，每篇新闻中一平均有 13 个呼吸呼吸重置，而停顿

平均则达到了 27 次，停顿数与呼吸重置次数相差极大，停顿数多一倍，也没有个人差异。

一般而言，朗读时人们按照语言内容和结构调节呼吸规划。比如，

“同时，/我们清醒地认识到，/我国发展中/不平衡、/不协调、/ 不可持续的问题依然突出。”

在文本中，阴影部分对应呼吸重置，/表示停顿，可以显示停顿和呼吸重置如何对应。从文本中可以看出，在一句话里，出现 6 次停顿，2 次呼吸重置，与上表 4.3 的结果比较一致。汉语朗读中出现的停顿与呼吸重置都和标点符号重合，但二者的单位不同。

分析结果说明，对汉语发音人而言，停顿与呼吸重置是不同的节奏概念，呼吸重置对应于大的语义单位，而停顿对应于小的语义单位。首先根据已规划的标点符号来安排大单位，然后按照语法标记、朗读内容以及个人对内容的理解来安排停顿。

由此可以总结，朗读汉语新闻中的呼吸重置特点是：

1. 标点符号是停顿的书面提示，通过对汉语新闻朗读中的呼吸节奏特征，我们看到汉语发音人朗读时，呼吸安排比较依赖于标点符号，主要标点处都有呼吸重置。可见，朗读汉语新闻中的呼吸重置属于“语义（语法）停顿”，几乎没有出现生理停顿。
2. 从方法上，朗读汉语新闻时，胸腹呼吸都存在三级呼吸重置，不管是胸呼吸还是腹呼吸，每条新闻第一自然段的呼吸类型比较固定，都是先有一个大呼吸重置。大呼吸对应于整个句子或者分句等，中呼吸主要对应于一个分句，小呼吸主要对应于分句中的某个句子成分，即并列词语或者句首出现的状语等。在腹呼吸重置中，不管是整个句子还是分句，大呼吸和中呼吸都比较自由地出现，小呼吸主要对应于并列词语、状语或者分句。
3. 从数据上，腹呼吸幅度与时长值均大于胸呼吸，腹呼吸重置次数也多于胸呼吸，该特征属于朗读中的呼吸重置具有的普遍性特征。在胸呼吸重置中，小呼吸占的比例最大，而在腹呼吸重置中，中呼吸次数占的比例最大，不同级别的呼吸重置分布比较均匀。
4. 从功能上，腹呼吸起始点与峰值均早于胸呼吸，重置幅度高于胸呼吸，在开始朗读时给胸呼吸以及言语活动提供能量，比胸呼吸结束早并开始准备下次呼吸。朗读时，胸呼吸晚于腹呼吸，其结束时间点和言语信号终点比较一致，而胸呼吸对言语活动有更密切的关系。
5. 汉语新闻朗读中，胸呼吸重置幅度与时长呈显著正相关，而腹呼吸重置幅度与时长无显著相关，发音人朗读汉语新闻时，不管是整个句子还是词语，主要以大呼吸和中呼吸来完成。
6. 从呼吸信号曲线的分析发现，声学分析上的停顿数比呼吸重置次数多一倍。对汉语发音人而言，停顿对应于小语义单位，相比呼吸重置对应于大语义单位，主要以大呼吸或者中呼吸来完成言语活动，朗读汉语新闻中的呼吸节奏主要受标点符号的制约。

4.2 韩语新闻的呼吸与韵律特征

不管是那种语言，新闻的写作有其弧顶的模式，一般来说新闻的第一段都是用一句话交待出某一新闻事件的概况，后面再展开叙述（谭晶晶，2008）。

根据汉韩停顿和呼吸节奏研究结果，汉语每 10 个音节出现一次停顿，每 15-16 个音节出现一次呼吸重置。而韩语则每 13 个音节或每两个韵律短语出现一次停顿（李浩英，1996；金成硅，1999），那么这些生理呼吸信号和声学信号如何对应？与汉语新闻朗读中的呼吸特征有何不同？

下面给出的是韩语新闻朗读中胸腹呼吸与语音信号的对应情况：

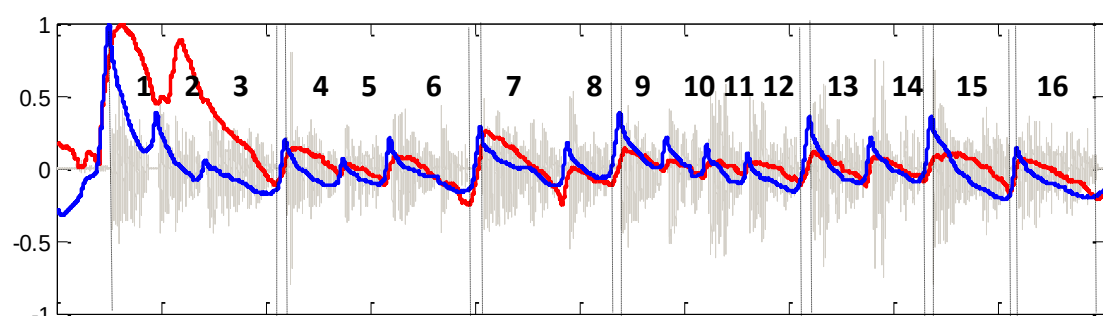


图 4.6 韩语新闻朗读中胸、腹呼吸曲线与语音波形

文本：(1) 신화사 베이징 3 월 6 일 보도: (2) 11 기 전국인대 4 차 회의는 (3) 6 일 오후 인민 대회당에서 기자 간담회를 진행했다. (4) 국가 발전 및 개혁위원회 장평(張平)주임, (5) 주즈신(朱之鑫), 쉬셴핑(徐憲平) 부주임이 (6) ‘12 차 5 개년’ 계획 강요와 관련해 기자들의 질문에 답했다. (7) 장평은 “11 차 5 개년 계획 시기 중국은 경제사회 발전에서 (8) 세계가 주목하는 성과를 이룩했다. (9) 동시에 중국이 발전하는 과정에 (10) 불균형, 부조화, (11) 발전이 지속적으로 이루어지지 않는 등 (12) 몇 가지 문제점이 여전히 존재하고 있다. (13) ‘12 차 5 개년’ 계획을 작성하는 과정에서 (14) 이런 문제와 관련하여 깊이 연구했다. (15) 앞으로 5 년간의 발전에서 이와 같은 문제를 잘 해결해야 한다. (16) 이는 ‘12 차 5 개년’ 계획을 구성하는 특징이기도 하다” 고 밝혔다.

(中文文本：(1) 新华社北京 3 月 6 日电：(2) 十一届全国人大四次会议 6 日下午在人民大会堂举行记者会，(3) 国家发展和改革委员会主任张平、(4) 副主任朱之鑫、(5) 徐宪平就“十二五”规划纲要回答记者提问。(6) 张平说，(7) “十一五”时期我国经济社会发展取得了举世瞩目的成绩。(8) 同时，我们清醒地认识到，(9) 我国发展中不平衡、不协调、(10) 不可持续的问题依然突出。(11) 在编制“十二五”规划纲要过程中，(12) 针对这些问题做了深入研究，(13) 期望在今后五年的发展进程中解决好这些问题。(14) 这也就构成了“十二五”规划的特点。)

上图 4.6 中的红色曲线显示胸呼吸重置变化，蓝色曲线则是腹呼吸变化曲线，黑色虚线是每小句语音信号的始点和终点。

从呼吸方式上看，朗读汉语和韩语新闻中的呼吸重置特征之间最显著的差异是韩语新闻朗读中的胸腹呼吸主要以小呼吸为主，而且胸呼吸重置幅度相等

或者比腹呼吸还低一点，显得胸腹呼吸配合比较均称。

而二者的呼气方式是不一致的，二者呼气时胸呼吸曲线相对比较缓慢，持续到语音结束，相比腹呼吸曲线下降地比较快，在语音结束前已开始下一次呼吸重置。可见，该特征是不管汉语还是韩语，近体诗还是新闻，朗读中呼吸重置具有的普遍性特征。

韩语除了标点符号之外，以语法词来表示句法成分的不同，如，文本“(2)11기 전국인대 4차 회의는(十一届/全国人大/四次会议)”中“는”表示主语或者话题的语法词，在大多数情况下，在主语或者话题后边都会出现呼吸重置，而朗读汉语新闻中没有出现任何呼吸。再如，文本“(9)동시에 중국이 발전하는 과정에(同时/我国/发展/过程中)”中“에”表示状语的语法词，后边出现呼吸重置。朗读一篇新闻中，共16次呼吸，其中9次呼吸是标点符号，句号和逗号等，其余的7次重置是发音人根据新闻内容和语法标记来安排的。

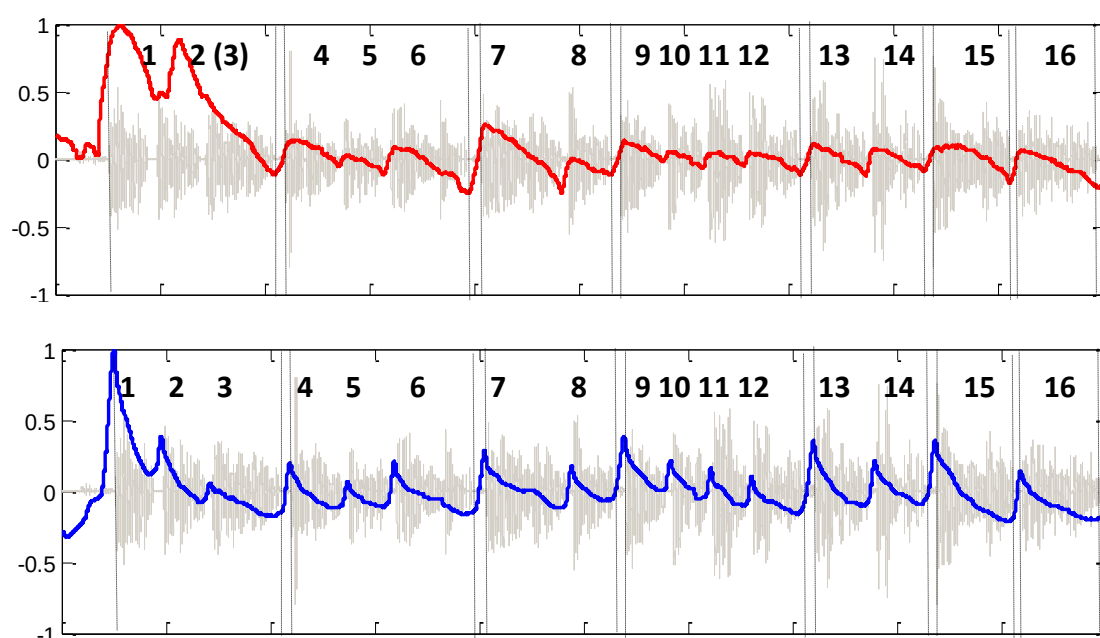


图 4.7 韩语新闻朗读中胸、腹呼吸与语音波形
(上图：胸呼吸曲线与语音波形，下图：腹呼吸曲线与语音波形)

文本：(1)신화사 베이징 3월 6일 보도：(2)11기 전국인대 4차 회의는 (3)6일 오후 인민 대회당에서 기자 간담회를 진행했다. (4)국가 발전 및 개혁위원회 장평(張平)주임, (5)주즈신(朱之鑫), 쉬쎌핑(徐憲平) 부주임이 (6) ‘12차 5개년’ 계획 강요와 관련해 기자들의 질문에 답했다. (7)장평은 “11차 5개년 계획 시기 중국은 경제사회 발전에서 (8)세계가 주목하는 성과를 이룩했다. (9)동시에 중국이 발전하는 과정에 (10)불균형, 부조화, (11)발전이 지속적으로 이루어지지 않는 등 (12)몇 가지 문제점이 여전히 존재하고 있다. (13) ‘12차 5개년’ 계획을 작성하는 과정에서 (14)이런 문제와 관련하여 깊이 연구했다. (15)앞으로 5년간의 발전에서 이와 같은 문제를 잘 해결해야 한다. (16)이는 ‘12차 5개년’ 계획을 구성하는 특징이기도 하다”고 밝혔다.

(中文文本：(1)新华社北京 3 月 6 日电：(2)十一届全国人大四次会议 6 日下午在人民大会堂举行记者会，(3)国家发展和改革委员会主任张平、(4)副主任朱之鑫、(5)徐宪平就“十二五”规划纲要回答记者提问。(6)张平说，(7)“十一五”时期我国经济社会发展取得了举世瞩目的成绩。(8)同时，我们清醒地认识到，(9)我国发展中不平衡、不协调、(10)不可持续的问题依然突出。(11)在编制“十二五”规划纲要过程中，(12)针对这些问题做了深入研究，(13)期望在今后五年的发展进程中解决好这些问题。(14)这也就构成了“十二五”规划的特点。)

从图 4.7 所示，韩语新闻朗读中胸、腹呼吸重置明显分为三个级别，三个级别的呼吸单元分别和不同大小的语义单元相对应。大呼吸单元对应整段话，也就是对应着最大的语义单元，表达一个最完整的意思。在韩语朗读中，无论是胸呼吸还是腹呼吸，大呼吸出现在每篇新闻的开始处，其呼吸曲线经常呈“阶梯型”，如下图 4.8 所示。第一出现的大呼吸往往会持续到一个句子的结尾，这些特征与汉语新闻朗读中的呼吸特征比较一致，也就是说，大呼吸的分布和表现最为稳定。

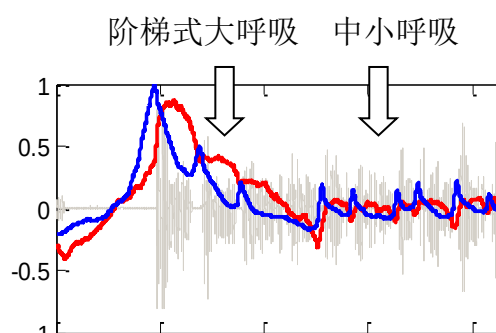


图 4.8 朗读韩语中的呼吸重置特征实例

图 4.7 所示，与汉语新闻呼吸特征不同的是中小呼吸的安排，朗读韩语新闻时不同大小的中小呼吸比较多，尤其胸呼吸重置中，小呼吸重置占大比例。

胸呼吸重置曲线显示，每个中呼吸单元对应一个分句或者较长的状语成分，即一个中等的语义单元。比如，图 4.7 中的“(7)장평은 “11 차 5 개년 계획 시기 중국은 경제사회 발전에서” (张平说，“十一五”时期我国经济社会发展)。

小呼吸主要对应于分句中的某个句子成分，即主语、并列词语或者句首出现的状语等。比如，“(5)주즈신(朱之鑫),쉬첸핑(徐憲平) 부주임이 (副主任朱之鑫、徐宪平)”和“(10)불균형,부조화 (不平衡、不协调)”等。但是除了每段落开始处之外，胸呼吸主要以小呼吸来配合言语活动，从小的句子成分到整个句子，都可以对应小呼吸重置。比如，在胸呼吸重置曲线图中，“(16)이는 ‘12 차 5 개년’ 계획을 구성하는 특징이기도 하다”고 밝혔다. (这也就构成了“十二五”规划的特点。)”也对应小呼吸重置。

与汉语呼吸特征结合来看，可以说，不管是汉语还是韩语，小呼吸是朗读新闻中的胸呼吸重置具有的共同特征。

腹呼吸重置特征与胸呼吸的有所不同。如图 4.6 和 4.7 所示，虽然胸腹呼吸出现的位置基本一致，但腹呼吸重置的中小呼吸和不同语义单位对应的规律比

胸呼吸明确。腹呼吸重置中，中呼吸像大呼吸一样对应于整个句子，也可以对应分句或者较大的状语成分。像胸呼吸一样，小呼吸主要对应于较小的语义单位；与胸呼吸不同的是，腹呼吸的小呼吸不与完整的句子相对应，不同级别的呼吸之间分工比较明确。

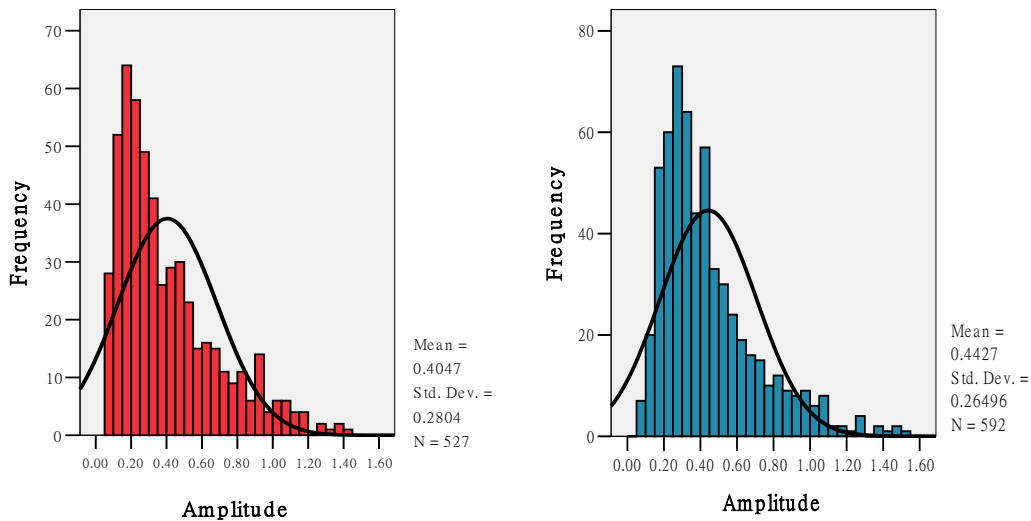


图 4.9 韩语新闻呼吸重置分布
(左图为胸呼吸；右图为腹呼吸)

Final Cluster Centers				Number of Cases in each Cluster	
	Cluster			Cluster	
	1	2	3		
°«ÓiĐÂîÄ				1	73.000
ˆˆôüüÖØ				2	156.000
ÖÄ·ü¶Ë				3	298.000
	.96	.52	.21	Valid	527.000
				Missing	.000

Final Cluster Centers				Number of Cases in each Cluster	
	Cluster			Cluster	
	1	2	3		
°«ÓiĐÂîÄ				1	72.000
ˆˆôüüÖØ				2	201.000
ÖÄ·ü¶Ë				3	319.000
	1.00	.54	.26	Valid	592.000
				Missing	.000

表 4.5 韩语新闻呼吸重置聚类分析结果
(上表为胸呼吸；下表为腹呼吸)

从具体数据来说，上图 4.9 所示，胸腹呼吸的总呼吸重置次数分别为 527 次和 592 次，也就是说每篇新闻朗读中出现的呼吸重置次数分别为 17.5 和 19 次。

由此可见，腹呼吸重置幅度和频度都高于胸呼吸的，这些特征都与汉语新闻中的呼吸特征一致。

不同的是，与汉语相比，韩语朗读新闻中的呼吸重置幅度值非常低，胸腹

呼吸幅度分别为 0.40 和 0.44（汉语胸呼吸幅度为 0.62，腹呼吸为 0.71），因此二者的分布明显呈左偏分布，小呼吸重置的频度占最大比例，从图 4.9 中我们看到呼吸重置幅度在 0.2-0.4 处有一个很明显的高峰，不管是胸呼吸还是腹呼吸，其频度均接近于 70。

韩语新闻朗读中胸呼吸重置聚类分析结果显示三种不同级别呼吸的中心值分别为 0.96、0.52 和 0.21，重置次数分别为 73 次、156 次和 298 次。可见，该结果符合上图 4.7 显示的呼吸曲线情况，即胸呼吸三级呼吸次数占最大比例，高达 57%，比一级呼吸的次数多三倍，中呼吸大致占 30%，大呼吸占的比例最小。由此我们可以发现，该特征与汉语胸呼吸特征比较相似，不管是汉语还是韩语，朗读中胸呼吸重置安排集中在小呼吸。

腹呼吸重置分布和胸呼吸的情况比较相似，胸呼吸也分成三级，其中心值分别为 1.0、0.54 和 0.26，比胸呼吸稍微高一点，而差别不大。表 4.5 呼吸重置聚类分析结果显示，三级呼吸出现 319 次，占 54%，其次是二级呼吸，出现 201 次占 34%，与胸呼吸的分布情况比较一致。朗读韩语新闻时，胸腹呼吸之间的差别在于其不同功能上。

从功能上看，韩语朗读中的胸腹呼吸之间的不同功能没有汉语的明确，但是腹呼吸还是主要提供大能量并帮助胸呼吸持续言语活动，胸呼吸与整个言语活动的开始和完成有关，这点是言语呼吸具有的普遍性特征，只是根据语言或者不同发音人有所区别。

Correlations			
		·ùŋĚ	Ê±³α
°«ÓĬĎÂĬĂ ĐĐ°ôĭũÖØ ÖĂ·ùŋĚ	Pearson Correlation	1	.270**
	Sig. (2-tailed)		.000
	N	493	493
°«ÓĬĎÂĬĂ ĐĐ°ôĭũÖØ ÖĂ·ùŋĚ	Pearson Correlation	.270**	1
	Sig. (2-tailed)	.000	
	N	493	493

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations			
		·ùŋĚ	Ê±³α
°«ÓĬĎÂĬĂ ,¹°ôĭũÖØ ÖĂ·ùŋĚ	Pearson Correlation	1	.244**
	Sig. (2-tailed)		.000
	N	568	568
°«ÓĬĎÂĬĂ ,¹°ôĭũÖØ ÖĂ·ùŋĚ	Pearson Correlation	.244**	1
	Sig. (2-tailed)	.000	
	N	568	568

** . Correlation is significant at the 0.01 level (2-tailed).

表 4.6 不同级别的呼吸重置幅度和时长平均与相关
（上表为胸呼吸；下表为腹呼吸）

从数据来说，朗读韩语新闻中胸呼吸的一级大呼吸重置幅度平均为 1.0，二

级中呼吸重置幅度平均为 0.5，三级小呼吸重置幅度平均为 0.2，对应的时长为 3.6、3.2 和 2.5，呼吸幅度与时长强度呈正相关，与朗读汉语时的呼吸情况比较一致，符合之前的研究结果（谭晶晶，2008；杨锋 等，2013）。而腹呼吸的表现有所区别，朗读汉语中的腹呼吸幅度与时长呈无相关关系，从上文已分析的腹呼吸曲线可以看到，发音人安排腹呼吸时，不好归纳呼吸重置和句子成分的对应，例如不管是三个音节的状语成分还是较长的分句，都安排大呼吸来完成。相反，朗读韩语新闻中，无论是胸呼吸还是腹呼吸，幅度和时长呈显著正相关，表 4.6 所示。

此外，与汉语新闻每个不同级别的呼吸重置分析结果进行比较，我们发现两种不同语言新闻朗读中，二者的一级呼吸的幅度和时长平均值比较一致，即大呼吸的表现比较稳定，主要区别在于中小呼吸的幅度和时长的表现。

因此，我们为了了解朗读韩语新闻中的声学信号和生理信号的对应关系，根据语音信号提取了两个发音人的新闻朗读中的所有停顿，又根据已提取的呼吸重置幅度值，分别统计出他们的平均值，如下表4.4所示。

	生理信号		声学信号
	胸呼吸 重置次数	腹呼吸 重置次数	停顿数
发音人1	15	16	19
发音人2	20	22	25
平均	17.5	19	22

表 4.4 韩语新闻呼吸重置与停顿数比较

从表中我们可以看到，每篇新闻中一平均有 18-19 个呼吸呼吸重置，而停顿平均则达到了 22 次，停顿数与呼吸重置次数相差不大，与汉语的对应情况完全不同。韩语的呼吸次数多于汉语的，而停顿数比汉语少。

较之上文已分析的汉语新闻朗读中的呼吸重置特征结合来看，韩语的呼吸次数比汉语的多一些，那是因为韩语书写的最大特点是每个词语之间都会出现一个“空格”，根据说话者的意图，比较自由地选择停顿的位置。但是与汉语的情况不同的是，其停顿数没有汉语一样大量增加，如表 4.4 所示，也没有个人差异。

李浩英（1996）和金成硅（1999）研究表明，韩语的停顿单位平均包括 13-14 个音节，或者有四个重音短语的话，在第三个开始处会出现一个停顿，而汉语的停顿平均包括 10 个音节（谭晶晶，2008），可见在声学研究中，韩语的停顿单位大于汉语的，与本此实验结果一致。

有趣的是，通过生理呼吸信号分析的结果发现，汉语的呼吸单位反而大于韩语，与声学分析结果相反。下例的语句可以显示韩语新闻朗读中，发音人如何将呼吸信号与停顿相对应。

“동시에 중국이 발전하는 과정에/불균형,부조화,/ 발전이
지속적으로 이루어지지 않는 등/몇 가지 문제점이 / 여전히 존재
하고 있다.”

（“同时，/我们清醒地认识到，/我国发展中/不平衡、/不协调、/ 不可持续的问题依然突出。”）

在文本中，阴影部分对应呼吸重置，“/”表示停顿，可以显示停顿和呼吸重置如何对应。从文本中可以看出，在一句话里，出现 4 次停顿，3 次呼吸重置，二者的位置基本重合，主要在标点处和韵律句边界处。安秉燮（2007）的研究表明，朗读韩语时，在一般的情况下，在韵律句边界处会出现停顿，而韵律短语边界处无停顿。如表 4.4 的结果所示，朗读韩语新闻中呼胸腹呼吸重置出现的次数和停顿数之间没有显著差异，就是说在朗读韩语时，停顿和呼吸重置出现的地方比较一致，主要在韵律句边界处，一般出现在句子、小句、主语以及状语的后边。

汉语新闻朗读中，对汉语发音人而言，停顿和呼吸节奏是两个不同的单位。除了主要标点符号出现的位置之外，停顿会出现在其他小语义单位边界处，而呼吸重置等于大语义单位，因此每次呼吸重置都比较大，基本上只受标点符号的制约，很少参照其他标准。

相比之下，对韩语发音人而言，停顿和呼吸节奏是相近的单位。韩语的韵律短语包含的音节数比较多，3-6 个，因此由 3,4 个韵律短语组成的韵律句和呼吸重置节奏比较一致。并且发音人已知，由于韩语结构的特征，除了语法词之外，像汉语的标点符号，没有较多的明显的标记，因此朗读时，主要用小呼吸来比较自由地安排停顿的位置，发音人朗读韩语新闻中不需要刻意安排大呼吸来完成长句。

总之，由于二者的语言结构不同，发音人呼吸安排时采取的方式也完全不一样，反映了二者的不同认知方式。

由此我们可以总结，朗读韩语新闻中的呼吸重置特点是：

1. 通过对韩语新闻朗读中的呼吸节奏特征，我们看到韩语发音人朗读时，呼吸重置除了标点符号之外，主要出现在韵律句边界处。但是相对而言，在韩语朗读中能比较自由地安排呼吸重置位置。因此，在一般情况下，朗读汉语新闻中的呼吸重置属于“语义（语法）停顿”，即生理停顿很少见。
2. 从方法上，朗读汉语新闻时，胸腹呼吸都存在三级呼吸重置，不管是胸呼吸还是腹呼吸，每条新闻第一自然段的呼吸类型比较固定，都先有一个大呼吸重置。朗读韩语新闻时不同大小的中小呼吸比较多，除了每段落开始处之外，胸呼吸主要以小呼吸来配合言语活动，从小的句子成分到整个句子，都可以对应小呼吸重置。
3. 从数据上，腹呼吸幅度与时长值均大于胸呼吸，腹呼吸重置次数也多于胸呼吸，该特征属于朗读中的呼吸重置具有的普遍性特征。在胸呼吸重置中，小呼吸占的比例最大，而在腹呼吸重置中，中呼吸次数占的比例最大，不同级别的呼吸重置分布比较均匀。
4. 从功能上，腹呼吸起始点与峰值均早于胸呼吸，重置幅度高于胸呼吸，在开始朗读时给胸呼吸以及言语活动提供能量，比胸呼吸结束早并开始准备下次呼吸。朗读时，胸呼吸晚于腹呼吸，其结束时间点和言语信号终点比较一致，而胸呼吸与对言语活动关系更密切。
5. 在汉语和韩语的呼吸重置幅度和时长相关特征之间最显著的差异在于腹呼吸的表现上。与语音段重合的胸呼吸重置幅度与时长呈正相关，无论是汉语

还是韩语。与此相比，汉语的腹呼吸重置幅度与时长不相关，不管是朗读的时长是多少，汉语腹呼吸主要以大而尖的呼吸为主。

6. 从呼吸信号曲线的分析可以发现，声学分析上的停顿数与呼吸重置次数相差不大，停顿位置与呼吸重置位置基本重合。由此可见，对韩语发音人而言，停顿与呼吸重置是功能相等的节奏单位，主要以“小呼吸”来完成每个语义单位，出现在韵律句边界处，如出现在句子、小句、主语以及状语的后边。

4.3 结果

本章基于呼吸信号主要从呼吸节奏与韵律表现的角度中考查具有不同语言结构的汉语和韩语新闻朗读中的呼吸节奏特征的差异，并试图探讨呼吸节奏和韵律单位的对应以及不同语言结构对呼吸节奏的影响。

通过本章的分析，我们发现：

不管哪种语言，朗读新闻中整体呼吸格局有一定的一致性。朗读中胸、腹呼吸基本分成三级，腹呼吸重置的幅度、频度都高于胸呼吸。在大多数情况下，每段话的第一句会出现大呼吸，表现形式主要有阶梯式呼吸或者一个大呼吸。

朗读汉语新闻时，胸、腹呼吸分工更加明确，即在开始朗读时给整个言语活动提供能量，比胸呼吸结束早而朗读中对韵律单位的界限更敏感，韵律短语边界处常出现的腹呼吸，从而呼吸次数比胸呼吸多一些；呼气之后保持一定时间使言语活动持续下去，因此胸呼吸的终点和言语信号的终点基本吻合，胸呼吸与整个言语活动的完成有关。虽然韩语胸腹呼吸之间的分工比汉语版呼吸特征模糊，但还是可以说，这些不同功能属于朗读言语中出现的普遍性特征。

从呼吸安排与语言结构而言，汉语与韩语之间最显著的共同点是，不管朗读哪种语言，文本中出现的标点符号引发生理呼吸重置，如，句号（。）、逗号（，）和顿号（、）处都会出现呼吸重置。

值得注意的是，以往的声学研究表明，韩语的停顿单位大于汉语，但是通过生理呼吸信号分析结果发现，汉语的呼吸单位反而大于韩语，与声学分析结果相反。

汉语新闻朗读中，停顿数远远多于呼吸重置数。对汉语发音人而言，停顿对应于小语义单位，相比呼吸重置对应于大语义单位，主要以“大呼吸”或者“中呼吸”来完成言语活动，朗读汉语新闻中的呼吸节奏主要受标点符号的制约。与此相比，韩语新闻朗读中，声学分析上的停顿数与呼吸重置次数相差不大，停顿位置与呼吸重置位置基本重合。由此可见，对韩语发音人而言，停顿与呼吸重置是功能相等的节奏单位，主要以“小呼吸”来完成每个语义单位。不同认知方式直接影响汉语与韩语两种不同语言的呼吸重置平均幅度，汉语胸腹呼吸重置幅度平均分别为 0.62 和 0.71，韩语的分别为 0.40 和 0.44，二者之间相差极大。

从呼吸幅度与时长的相关关系上，汉语和韩语之间存在显著性差异：新闻朗读中，不管是哪种语言，胸呼吸幅度与时长之间成正比，存在显著性相关。与之相比，腹呼吸重置与时长相关分析结果显示，汉语新闻朗读中，大呼吸与中呼吸出现的频度极高，不管是小语义单位还是整个句子，几乎无规律地出现。因此，汉语腹呼吸重置幅度与时长不存在显著正相关，只有韩语新闻朗读中的腹呼吸重置与时长呈正相关。

第五章 结论

韵律不仅具有辨义功能，而且是影响语音表现力的主要因素，声音的轻重缓急、抑扬顿挫、不同语气和口气的传达、个人语音特征的辨别，都要依靠韵律的作用。因此，韵律研究在当前的语音研究中占有重要位置，深入理解韵律不仅是语音学理论发展的需要，也是语音合成和语音识别发展的需要。

本文从语言学和生理学的交叉研究的角度出发，采录了韩国人和中国人共 4 个发音人朗读的 20 首汉语与韩语近体诗与 20 篇新闻。我们使用言语呼吸韵律分析系统，对语音和胸、腹呼吸信号添加韵律层级和呼吸重置标记，提取胸腹呼吸的重置幅度、时长、斜率等参数，分析朗读在韵律和呼吸两方面的差异，探索语言结构与呼吸节奏之间的关系和不同语言的呼吸方法。

5.1 研究结果

本文研究结果显示，由于近体诗和新闻的形式比较固定，因此不管哪种语言哪种形式，朗读中整体呼吸格局有一定的一致性。

下表 5.1 给出的是，朗读汉语和韩语近体诗和新闻中的不同呼吸重置特征。结果显示，不管是哪种语言，哪种文体，朗读中腹呼吸重置幅度和频度都高于胸呼吸，这些特征都与汉语新闻中的呼吸特征一致，无例外。

		幅度		次数		级别	
		胸呼吸	腹呼吸	胸呼吸	腹呼吸	胸呼吸	腹呼吸
汉语	五言	0.65	0.72	94	96	2 级	2 级
	七言	0.52	0.67	116	122	2 级	3 级
	新闻	0.62	0.71	454	481	3 级	3 级
汉字音	五言	0.52	0.57	105	104	2 级	2 级
	七言	0.56	0.61	118	119	2 级	2 级
韩语	五言	0.70	0.73	122	123	2 级	2 级
	七言	0.69	0.70	145	146	3 级	2 级
	新闻	0.40	0.44	527	592	3 级	3 级

表 5.1 汉语和韩语朗读中的呼吸节奏特征比较

由此我们可以发现，从胸、腹呼吸配合的角度看，不管是哪种语言还是哪种文体，言语以胸腹联合式呼吸为主，朗读中腹呼吸起始点都早于胸呼吸，在开始朗读时给整个言语活动提供能量，有辅助胸呼吸的作用，而提前结束并准备下一次呼吸和言语活动；一开始朗读，胸呼吸晚于腹呼吸，胸呼气开始点与终点与言语信号时间点比较一致。也可以说，胸呼吸与生理信号对应，腹呼吸与韵律单位对应。这些特征具有普遍性，胸腹呼吸各有不同功能，二者的配合是有一定规律的。只是，朗读韩语中，胸、腹呼吸表现比较一致，二者的分工比汉语呼吸特征模糊。

朗读近体诗中胸、腹呼吸基本分成二级，大小呼吸交替出现，每联的第一

句一定会出现大呼吸，在大多数情况下，第二句和第四句开始处具有小呼吸。

与之前的呼吸研究结果最显著的不同点是，不管是哪种语言的呼吸曲线，韵律短语边界处不仅会出现抖动，还会出现明显的小呼吸。有趣的是，这些小呼吸的出现的位置往往与声学分析中的停顿的位置和时长无关，有待进一步研究。

对本实验分析的呼吸曲线进行对照，我们可以发现三个级别的呼吸单元分别和不同大小的语义单元相对应。大呼吸单元对应整段话，也就是对应着最大的意义单元，表达了一个最完整的意思，大呼吸的特征符合之前的研究结果（谭晶晶，2008；杨锋 等，2013）。不管是哪种语言，哪种文体，大呼吸的表现最为稳定，主要出现在近体诗的第一联或者新闻的第一段。

从五言近体诗到七言近体诗，汉语版和韩语版呼吸变化趋势截然不同：如表 5.1 所示，汉语版五言近体诗朗读中胸、“腹呼吸”都存在二级呼吸，而七言近体诗朗读中，只有腹呼吸变成三级呼吸。与此相比，韩语版七言诗朗读中，因为增加的两个音步影响到韩语朗读者的呼吸安排，结果像汉语版的变化一样，随着小呼吸次数的增加，平均呼吸幅度也降低了一些。与汉语版呼吸变化不同的是，韩语版“胸呼吸”重置大小不一而变成三级呼吸，腹呼吸比较稳定。由此可见，汉语近体诗格律的严格性制约不同结构语言的节奏特征。虽然呼吸配合方式不同，但韩语朗读者调整呼吸重置幅度、次数和朗读速度而体现汉语近体诗格律特征。

在朗读新闻时，汉语发音人主要依赖于“标点符号”。呼吸重置出现的位置与标点处基本重合，可以说标点符号能制约汉语新闻朗读中的呼吸节奏。与此相比，韩语发音人主要根据标点符号和语法词的位置规划整个呼吸节奏方式，比汉语的安排方式自由一点。一般而言，大呼吸对应于整个句子，每个中呼吸单元对应一个分句，即一个中等的语义单元，小呼吸主要对应于分句中的某个句子成分，即并列词语或者句首出现的状语等。而朗读时，根据不同语言结构而呼吸安排方式也会改变。

汉语新闻朗读中，对汉语发音人而言，停顿和呼吸节奏是两个不同的单位。除了主要标点符号出现的位置之外，停顿会出现在其他小语义单位边界处，而呼吸重置等于大语义单位，因此每次呼吸重置都比较大，主要以“大呼吸”或者“中呼吸”来完成言语活动，基本上只受标点符号的制约，很少参照其他标准。

相比之下，对韩语发音人而言，停顿和呼吸节奏是相近的单位。朗读时，主要用小呼吸来比较自由地安排停顿的位置，发音人朗读韩语新闻中不需要刻意安排大呼吸完成长句。因此，韩语朗读新闻中的呼吸重置幅度值非常低，重置次数比较多，这说明在朗读韩语新闻时，发音人主要用“小呼吸”来安排呼吸协奏。韩语新闻朗读中，声学分析上的停顿数与呼吸重置次数相差不大，停顿位置与呼吸重置位置基本一致。由此可见，不同认知方式直接影响到汉语与韩语两种语言的呼吸重置方式。

与本次实验结果结合来讲，标点符号是书面语言的有机组成部分，是书面语言不可缺少的辅助工具，可以帮助人们确切地表达思想感情和理解书面语言。因此，朗读汉语新闻中呼吸重置出现在标点处是合理的。人们一般是倾向于把一个完整的意义放在一个韵律单元里实现的，在说话时人们可以通过调节呼吸节奏保证语义表达的完整性。

在朗读汉语新闻朗读中，基本不出现所谓朗读者一口气读不完而发生的换气停顿，也叫“生理停顿”，都以标点符号来替代该功能，因此主要是由标点符号导出的“语义停顿”。可以说，汉语标点符号支配朗读汉语新闻中的呼吸重置安排，从而保证新闻信息有效地传达，这也是一种汉语独有的呼吸节奏特征。

从呼吸幅度与时长的相关关系上，汉语和韩语之间存在显著性差异：朗读新闻时，汉语和韩语的相关特征比较一致，幅度与时长呈显著正相关。但是近体诗中的表现有所不同。

汉语版近体诗和汉字版近体诗朗读中，呼吸幅度与时长之间的相关分析结果比较一致，即五言近体诗朗读中，幅度与时长之间成正比，存在显著性相关。与此相比，韩语版近体诗朗读中，不管是五言还是七言近体诗，不同级别的呼吸时长之间差别不明显，朗读者按照固定的呼吸模式，调整整个呼吸安排体现汉语近体诗格律，结果相关分析显示，幅度与时长之间不存在显著性相关。

此结果既可以显示汉语近体诗格律的特征，也可以表现韩语音节结构的特征“等时性”。从整体呼吸安排来说，汉语近体诗格律优先于不同语言节奏特征，不管是哪种语言，五言还是七言，朗读者按照严格的近体诗格律来安排呼吸方式，体现近体诗的情感以及近体诗格律带来的节奏感，该特征具有一定的普遍性。这还说明，呼吸安排决定于格律和朗读者对诗歌内容的不同解释而产生的情感，基本上不在于生理上的需求，出现的呼吸重置都属于“语义停顿”。

本研究证实了郑秋豫（2005）人们在说话时还要除了受到发声器官和生理机制的制约外，还要受到认知能力的制约，也就是说，人们在说话之前要对所说的内容做一个规划，划分韵律单元的时候应该也会考虑到内容的完整性和语法的制约。

总而言之，本文的研究显示朗读汉语和韩语时出现的不同呼吸节奏特征，还发现不同语言结构会影响到整个呼吸节奏的安排和呼吸方式。另外，从生理信号与声学信号的对应中发现，声学分析结果还是不能完全解释言语过程中的认知方式与过程，因此基于生理学的韵律研究还需要更深层次的实验研究。

5.2 创新之处

本文的创新点在于选题和研究方法的新颖性：

1. 目前大部分的韵律研究主要以声学分析为主，基于生理信号研究韵律的很少。本文从呼吸节奏这一新的角度着手，研究汉韩两种语言的不同文体及语流中的呼吸节奏的类型和特点，并试图寻找呼吸信号和声学信号怎么对应、呼吸节奏和韵律结构的关系以及不同句法结构对呼吸节奏的影响。以此丰富结合生理语音学方法的语音多模态研究。

2. 本实验充分利用实验仪器以及相关软件，将胸呼吸、腹呼吸、语音信号同步录制，同步分析，便于考查语音产生的机制和相互间的配合关系。

3. 此外，本文对汉语和韩语的呼吸节奏特征进行对比研究，因为在以往的语音对比研究中都是针对声学特征进行研究的，没有深入分析不同语言之间的生理机制以及认知上的区别。本文试图探讨语言结构与生理信号之间的关系，归纳汉语与韩语呼吸韵律特征的异同。

5.3 问题与展望

本文采录的呼吸信号非常敏感，很容易受到生理和心理以及个人情况的影响。因此，在数据中难免出现较多的例外，也会和其他实验结果有出入。此外，本实验中还同步采录了 EGG 嗓音信号，时间所限未能分析，下一步将分析不同语言的韵律特征和嗓音信号，探讨呼吸-发声-调音之间的关系。

呼吸信号的研究作为韵律研究的新方法和手段，进一步还可以与气流气压计相结合，更加精确地计算气流量、声门下压力、肺的容积变化等问题。同时，结合螺旋 C T 拍摄的肺部图像建立肺部的动态 3 维几何模型，对深入研究言语产生的呼吸机制具有重要作用（杨锋 等，2013），可以说多种仪器相结合的多莫泰研究将成为语音学研究的发展方向。

呼吸信号的研究，不仅能揭示人的发音动力系统的机理，作为韵律研究的新方法和手段，能为国内语音学的韵律研究开辟新的课题，为工程语音学提供更多的理论支持（李永宏 等，2010）。

参考文献

中文文献

- 曹剑芬（2001）汉语韵律切分的语音学和语言学线索，《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
- 陈娟文（2004）上海普通话和普通话韵律特征对比研究，浙江大学硕士学位论文
- 陈默（2007）韩国留学生汉语句子停延习得的实验研究，《暨南大学华文学院学报》第二期
- 初敏、王韞佳、包明真（2004）普通话节律组织中的局部语法约束和长度约束，《语言学论丛》第三十辑，北京
- 端木三（1999）重音理论和汉语的词长选择，《中国语文》，第4期
- 冯葆富、齐忠政、刘运墀（1981）《歌唱医学基础》，上海科技出版社，上海
- 冯胜利（1997）《汉语的韵律、词法与句法》，北京大学出版社，北京
- 冯胜利（2000）《汉语韵律句法学》，上海教育出版社，上海
- 国家语言文字工作委员会（1995）《标点符号用法》，中国标准出版社，北京
- 郭锦桴（1993）《综合语音学》，福建人民出版社，福建
- 黄京洙（2009）《韩国语语言学概论》，辽宁民族出版社，沈阳
- 齐声乔（1956）汉语的字调停顿与语调的交互关系，《中国语文》
- 孔江平（2007）中国语言学的历史与现状，《语音乐律研究报告》，北京大学中文系语言学实验室
- 孔江平（2008）语音多模态研究和多元化语音学研究，《中国语音学报》，商务印书馆，北京
- 李翊燮、李相亿、蔡琬（韩）（2008）《韩国语概论》，世界图书出版公司，北京
- 李永宏、于洪志、孔江平（2008）现代语音学仪器及生理语音学研究，《生命科学仪器》，008.6，54-58
- 李永宏、张金爽、王姝文、于洪志（2010）藏语新闻朗读的呼吸信号声学分析，《西北民族大学学报 自然科学版》，第31卷 第78期
- 林茂灿（1988）普通话声调的声学特性与其知觉征兆，《中国语文》，第3期
- 林茂灿（2000）普通话语句中间断和语句韵律短语，《当代语言学》，第4期
- 林茂灿（2001）普通话孤立句的韵律结构和F0下倾，《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
- 林茂灿、颜景助、孙国华（1984）北京话两字组正常重音的初步实验，《方言》，第1期
- 林焘、王理嘉（1992）《语音学教程》，北京大学出版社，北京
- 林焘（1996）《语音研究和对外汉语教学》，世界汉语教学，1996(3)：18-21
- 卢纹岱（2006）SPSS for Windows 统计分析（第3版），电子工业出版社，北京
- 罗常培、王均（1981）《普通语言学纲要》第147页，商务印书馆
- 刘亚斌（2003）汉语自然口语的韵律分析和自动标注研究，中国社会科学院硕士学位论文
- 朴正俸（2009）《韩国汉字音系与厦门方言音系比较研究》，巴蜀书社，成都

沈炯（1985）北京话的声调的音域和语调，《北京语音实验录》，林焘，王理嘉编

沈晓楠（1992）汉语普通话声调的协同发音，《国外语言学》，第2期

宋吟宣（1999）《汉韩韵律成分的对比》，北京大学博士学位论文

谭晶晶（2008）汉语普通话朗读时的呼吸节奏研究，北京大学硕士学位论文

谭晶晶、孔江平（2006）新闻朗读的呼吸节奏初探，《语音乐律研究报告》，北京大学中文系语言学实验室

谭晶晶、李永宏（2007）汉语普通话不同文体朗读时的呼吸重置研究，《语音乐律研究报告》，北京大学中文系语言学实验室

谭晶晶、李永宏、孔江平（2008）汉语普通话不同文体朗读时的呼吸重置特征，《清华大学学报》，2008.4

王蓓，杨玉芳（2002）汉语重读音节的时长变化模式，《声学学报》

王蓓、杨玉芳、吕士楠（2004）汉语韵律层级边界结构的声学相关物，《声学学报》第1期，北京

王蓓、杨玉芳、吕士楠（2005）语篇中大尺度信息单元边界的声学线索，《声学学报》第3期，北京

王洪君（1999）《汉语非线性音系学》，北京大学出版社，北京

王洪君（2000）汉语的韵律词和韵律短语，《中国语文》第6期，北京

王洪君（2002）普通话中节律边界与节律模式、语法、语用的关联，《语言学论丛》第二十六辑，北京

王力（2002）《汉语诗律学》，上海教育出版社，上海

王毓钧（2012）生理呼吸与语言信号，《南开语言学刊》第2期

王温佳、初敏、贺琳、冯勇强（2001）语句中双音节韵律词重音感知的初步研究，《新世纪的现代语音学》一第五届全国现代语音学学术会议论文集，清华大学出版社。

王温佳，初敏，贺琳（2003）汉语语义重音分布的初步研究，第六届全国现代语音学学术会议论文集

吴洁敏（1990）停延初探，《语文建设》，第3期

吴洁敏（2001）《汉语节律学》，语文出版社，北京

吴洁敏、吕士楠（1997）《汉语停延的“内三层”和“外三层”结构模型》，《语文现代化论丛》第三辑，语文出版社

吴宗济（1982）普通话语句中的声调研究，《中国语文》，第6期

吴宗济（1985）普通话三字组变调规律，《中国语言学报》，第2期

吴宗济（1997）从声调与乐律的关系提出普通话语调处理的新方法，《中国语文》

吴宗济（2002）中国音韵学和语音学在汉语言语合成中的处用，《语言教学与研究》，第1期

吴宗济、林茂灿（1989）《实验语音学概要》，高等教育出版社

熊子瑜（2003）韵律单元边界特征的声学语音学研究，《语言文字应用》第2期

许毅（1986）普通话音联的声学语音学特征，《中国语文》，第五期

杨锋（2012）《中国传统吟诵研究-从节奏、嗓音和呼吸角度》，北京大学博士学位论文

杨锋、侯兴泉、孔江平（2013）言语产生的胸腹呼吸机制，《清华大学学报 自

然科学版》，第 53 卷，第 6 期

杨玉芳（1997）句法边界的韵律学表现，《声学学报》第 5 期

叶军（2008）《现代汉语节奏研究》，上海世纪出版集团，上海

张锦玉（2010）普通话语篇停顿与呼吸特征初探，《第九届中国语音学学术会议》

张锦玉，石锋，白学军（2012）讲述与朗读状态下呼吸差异的初步分析，《南开语言学刊》2012（1）：pp56-63

赵元任（1933）汉语的字调跟语调，《中央研究院历史语言研究所集刊》第 4 期

赵元任（1980）《语言问题》，商务印书馆，北京

郑秋豫（2001）语流中韵律结构的主要征信，第六届全国人机语音通讯学术会议

周明强（2002）句子的停延和句法结构的关系，《语言教学和研究》，第 3 期

仲晓波，杨玉芳（2003）汉语普通话句子重音在时长方面的声学表现，《自理学报》，第 35 卷 第 2 期

韩文文献

고도흥（高道兴）（2007）《언어기관의 해부와 생리，语言器官的解剖与生理机制》，도서출판 소화

고승연（高承妍）（2010）《중국인 학습자의 한국어 끊어읽기 양상，中国学习者的韩语阅读停顿研究》，釜山外国语大学 硕士学位论文

김성규（金成珪）（1999）잠재적 휴지의 실현，潜在停顿的实现，《선청语文》27-1，首尔大学 国语教育系，pp155-183

김진만（金真满）（2012）《한국어 낭독체의 발화속도와 휴지의 상관관계

연구：한국인과 중국인 학습자를 중심으로，韩语朗读体的速度与停顿的相关关系研究：以韩国母语者与中国学习者为主》，汉阳大学 教育学院 硕士学位论文

성철재（成哲在）（1991）《표준 한국어 악센트의 실험음성학적 연구：청취

테스트 및 음향 분석，标准韩语重音声学实验研究：听辨实验和声学分析》

首尔大学 语言学系 硕士学位论文

성철재（成哲在）（1994）《한국어 리듬의 실험음성학적 연구-시간 구조와

관련하여，韩语节奏声学实验研究-从时间结构的角度》首尔大学 语言学系 博士学位论文

심소희 (沈小喜) (1998) 中国语의 리듬单位와 语法构造 (汉语节奏单位和语法构造), 《中国语文学誌》5-1, 中国语文学会, pp. 307-346

안병섭 (安秉燮) (2007) 휴지의 역할에 대한 반성적 검토 (再考停顿的功能), 《우리어문연구》, 우리어문학회 28, pp67-87

안병섭 (安秉燮) (2008) 《국어 운율부의 구조와 기능에 대한 연구, 韩语韵律结构与功能研究》, 高丽大学 博士学位论文

안병섭 (安秉燮) (2010) 《한국어 운율과 음운론, 韩语韵律与音韵论》, 한국문화사

이숙향 (李淑香) (1985) 韩国语文尾抑扬에 대한 연구 (韩国语文尾抑扬研究) 《말소리》9-10, 大韩语音学会

이승녕 (李崇宁) (1960) 现代서울말의 Accent 의考察 (现代首尔语的重音考察), 《国语学论考》, 东洋出版社

이지향 (李智香) (2004) 《한국어 낭독체 서술문장의 운율특성분석: 음성합성용 PBS (Phonetically Balanced Sentences) 문장을 대상으로, 韩语朗读体陈述句文章的韵律特征分析: 语音合成用 PBS》, 忠南大学 硕士学位论文

이현복 (李炫馥) (1973) 现代 韩国의 Accent (现代韩国的重音), 《首尔大 文理大学报》19, 首尔大学 文理学院

이현복 (李炫馥) (1974) 서울말의 리듬과 억양 (首尔语的节奏与语调), 10-2, 首尔大学 语学研究所

이현복 (李炫馥) (1976) 한국어 단음절어의 억양연구 (韩语单音节词的语调研究), 《语言学》1, 韩国语言学会

이현복 (李炫馥) (1982) 한국어 리듬의 음성학적 연구 (韩语节奏声学研究), 《말소리》4, 大韩语音学会

이현복 (李炫馥) (1984) 《한국어의 표준발음, 韩语标准发音》, 大韩语音学会

이호영 (李浩英) (1996), 《국어음성학, 国语音声学》, 太学社

- 이호영 (李浩英) (1997) 《국어운율론, 国语韵律论》, 韩国研究院
- 인지영 (仁智英) (2007) 《한국어 정규뉴스 방송문장의 운율특성연구, 韩语新闻的韵律特征研究》, 忠南大学 硕士学位论文
- 정명숙 (郑明叔) (2007) 한국어 학습자를 위한 전략적 발음 교육 (针对韩语学习者的发音教育策略), 《韩国语学》第 38 期, 韩国语学会, pp346-369
- 정인섭 (郑仁燮) (1965) 우리말 Accent 는 高低 Accent 다 (国语重音属于高低重音), 《论文集》10, 中央大学
- 조동일 (赵东日) (1982), 《한국시가의 전통과 율격, 韩国诗歌的传统与格律》, 한길사
- 조동일 (赵东日) (1996), 《한국민요의 전통과 시가율격, 韩国民谣的传统与诗歌的格律》, 지식산업사
- 조두현 (曹斗铉) (1976) 《한시의 이해, 汉诗的理解》, 一志社, 首尔
- 최현배 (崔铉培) (1971) 《우리말본, 国语本》, 정음사
- 최현식 (崔贤识) (2013) 한국 근대시와 리듬의 문제 (韩国近代史与节奏问题) 《韩国学研究》30
- 허웅 (许雄) (1985) 《국어음운학, 国语音韵学》, 샘문화사

英文文献

- Baken R J, Cavallo S A, Weissman K L. Chest wall movements prior to phonation [J] Journal of speech, Language and Hearing Research, 1979, 22(4):862-872.
- C. Tseng et al. (2005) Fluent speech prosody: framework and modeling. Speech Communication 46 284-309
- C. Tseng (2006) Prosody analysis. *Advances in Chinese spoken language Processing*. World Scientific Publishing Co. Pte. Ltd. Singapore. pp: 57-76
- Denes, Peter B., Pinson, Elliot N. (1993) Speech Chain., W H Freeman
- 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[J]. Journal of Speech Language and Hearing Research, 16(1):78-115

- Hixon T J. (1987) Respiratory Function in Speech and Song. College-Hill Press. San Diego.
- Jung Myung-Sook(2002) “The Teaching Method of Korean Intonation by Basic Pattern ” , Korean Education, Education Institute of International Korean
- Ladefoged P., Whitteridge D. (1959) Respiratory muscles in speech [J]. Journal of Speech, Language and Hearing Research, 16-35
- Lee Hyun Bok(1965) A Study of Korean(Seoul) Intonation, M.A. Thesis, University of London
- Liberman P. (1967) Intonation, Perception and Language. Cambridge. Massachusetts, The MIT Press
- Nooteboom S (1997) The prosody of speech:melody and rhythm. In:;in Hardcastle L, editor. The handbook of phonetic sciences. Oxford: Blackwell. Pp. 640-673
- Ohala J J. (1990) Respiratory activity in speech. Speech Production and Speech Modelling. Dordrecht, The Netherlands:Springer, 1990:23-53.
- Philip Liberman (1967) *Intonation, perception, and language*. The M. I. T. Press, Cambridge Massachusettes
- Ron J. Baken et al. (1979) Chest wall movements prior to phonation. *Journal of Speech and Hearing Research*, 22, 862-872
- Sun-Ah, Jun. (2000) “K-ToBI (Korean ToBI) Labelling Conventions” , *UCLA Working Papers in Phonetics* 99
- Sun-hee, Kim. (2000) “ Intonation Patterns of Korean Spontaneous Speech” , The Voice and Phonetics Science, vol.1 No.4

附录 1 录音材料（汉语版）

诗 20 首

诗 1：望岳

岱宗夫如何，齐鲁青未了。
造化钟神秀，阴阳割昏晓。
荡胸生层云，决眦入归鸟，
会当凌绝顶，一览众山小。

诗 2：旅夜书怀

细草微风岸，危樯独夜舟。
星垂平野阔，月涌大江流。
名岂文章著？官应老病休。
飘飘何所似，天地一沙鸥。

诗 3：春望

国破山河在，城春草木深。
感时花溅泪，恨别鸟惊心。
烽火连三月，家书抵万金。
白头搔更短，浑欲不胜簪。

诗 4：天末怀李白

凉风起天末，君子意如何？
鸿雁几时到，江湖秋水多。
文章憎命达，魑魅喜人过。
应共冤魂语，投诗赠汨罗。

诗 5：月夜

今夜鄜州月，闺中只独看。
遥怜小儿女，未解忆长安。
香雾云鬟湿，清辉玉臂寒。
何时倚虚幌，双照泪痕干？

诗 6：客至

舍南舍北皆春水，但见群鸥日日来。
花径不曾缘客扫，蓬门今始为君开。
盘飧市远无兼味，樽酒家贫只旧醅。
肯与邻翁相对饮，隔篱呼取尽馀杯！

诗 7：登高

风急天高猿啸哀，渚清沙白鸟飞回。
无边落木萧萧下，不尽长江滚滚来。
万里悲秋常作客，百年多病独登台。
艰难苦恨繁霜鬓，潦倒新停浊酒杯。

诗 8：登楼

花近高楼伤客心，万方多难此登临。
锦江春色来天地，玉垒浮云变古今。
北极朝庭终不改，西山寇盗莫相侵！
可怜后主还祠庙，日暮聊为梁父吟。

诗 9：蜀相

丞相祠堂何处寻？锦官城外柏森森。
映阶碧草自春色，隔叶黄鹂空好音。
三顾频频天下计，两朝开济老臣心。
出师未捷身先死，长使英雄泪满襟！

诗 10：闻官军收河南河北

剑外忽传收蓟北，初闻涕泪满衣裳。
却看妻子愁何在，漫卷诗书喜欲狂。
白日放歌须纵酒，青春作伴好还乡！
即从巴峡穿巫峡，便下襄阳向洛阳。

诗 11：江雪

千山鸟飞绝，万径人踪灭。
孤舟蓑笠翁，独钓寒江雪。

诗 12：登鹳雀楼

白日依山尽，黄河入海流。
欲穷千里目，更上一层楼。

诗 13：悯农

锄禾日当午，汗滴禾下土。
谁知盘中餐，粒粒皆辛苦。

诗 14：春晓

春眠不觉晓，处处闻啼鸟。
夜来风雨声，花落知多少。

诗 15：静夜思

床前明月光、疑是地上霜。
举头望明月，低头思故乡。

诗 16：绝句

两个黄鹂鸣翠柳，一行白鹭上青天。
窗含西岭千秋雪，门泊东吴万里船。

诗 17：戏为六绝句（一）

庾信文章老更成，凌云健笔意纵横。
今人嗤点流传赋，不觉前贤畏后生。

诗 18：戏为六绝句（二）

王杨卢骆当时体，轻薄为文哂未休。
尔曹身与名俱灭，不废江河万古流。

诗 19：戏为六绝句（五）

不薄今人爱古人，清词丽句必为邻。

窃攀屈宋宜方驾，恐与齐梁作后尘。

诗 20：戏为六绝句（六）

未及前贤更勿疑，递相祖述复先谁？

别裁伪体亲风雅，转益多师是汝师！

新闻 20 段

汉语为原文

新闻 1：《人民日报海外版》（2011 年 03 月 07 日 第 04 版）

新华社北京 3 月 6 日电：十一届全国人大四次会议 6 日下午在人民大会堂举行记者会，国家发展和改革委员会主任张平、副主任朱之鑫、徐宪平就“十二五”规划纲要回答记者提问。张平说，“十一五”时期我国经济社会发展取得了举世瞩目的成绩。同时，我们清醒地认识到，我国发展中不平衡、不协调、不可持续的问题依然突出。在编制“十二五”规划纲要过程中，针对这些问题做了深入研究，期望在今后五年的发展进程中解决好这些问题。这也就构成了“十二五”规划的特点。

新闻 2：《人民日报海外版》（2011 年 03 月 21 日 第 04 版）

本报北京 3 月 20 日电：商务部部长陈德铭今天出席在此间举行的“2011 年中国发展高层论坛”上表示，“十二五”期间，中国将坚持进口和出口并重、吸收外资和对外投资并重，把提高对外开放水平作为转变经济发展方式的新动力。陈德铭表示，十年来，中国从一个世贸组织的新成员，通过不断学习规则、熟悉规则、遵守规则，逐渐成为规则制订的参与者和推动者。在这个过程中，中国政府付出了艰辛的努力，清理了 3000 多部法律法规和部门规章，平均关税从 15%降到 10%，开放了 100 多个服务贸易部门。

新闻 3：《人民日报海外版》（2011 年 04 月 14 日 第 02 版）

国家能源局局长刘铁男指出“中巴在油气、电力和新能源等领域的合作前景非常广阔。”首先，中国有稳定而且不断增长的市场；第二，中国在勘探开发和加工方面有资金优势、技术优势、管理优势。”刘铁男说，目前来看，巴西已经是中国石油进口的第十大国家。“如果双方加大来往的频率和贸易量的平衡，特别是在重化工领域和矿石石油方面，双方的互补性会很好地表现出来。”刘铁男表示，“十二五”时期，中国经济发展和能源发展蕴含着巨大的商机，欢迎巴方企业与中国企业开展各种方式合作。

新闻 4：《人民日报海外版》（2011 年 04 月 13 日 第 04 版）

4 月 12 日上午，国家旅游局在京举行新闻发布会，宣布根据《国务院关于同意设立“中国旅游日”的批复》，自 2011 年起，每年 5 月 19 日为“中国旅游日”，这也是《徐霞客游记》的开篇日。国家旅游局副局长祝善忠表示，将《徐霞客游记》的开篇日确定为“中国旅游日”，在文化内涵上与旅游联系密切。徐霞客是我国明代伟大的旅行家、地理学家、史学家、文学家，《徐霞客游记》既是系统考察祖国地貌地质的地理名著，又是描绘华夏风景资源的旅游巨篇，在国内外具有深远影响。

新闻 5：《人民日报海外版》（2011 年 02 月 18 日 第 01 版）

据新华社北京 2 月 17 日电：国家外汇管理局 17 日发布报告称，我国大规模跨境资金净流入与实体经济活动基本相符。虽然存在蚂蚁搬家式的“热钱”违法违规流入，但尚未发现正规金融机构大规模集中流入的情况。国家外汇管理局国际收支分析小组发布的《2010 年中国跨境资金流动监测报告》估算，2010 年我国“热钱”净流入 355 亿美元，占外汇储备增量的 7.6%。但这一数据并不都是纯粹套利的、违法违规的或者不可解释的跨境资金流动。

新闻 6：《人民日报海外版》（2011 年 11 月 21 日 第 02 版）

本报北京电：记者从中国—东盟商务理事会获悉，今年从 1 月份到 10 月份，中国与东盟贸易额达 2959 亿美元，较上年同期增长 25.7%。其中，中国从东盟进口增长了 27.5%，中国向东盟出口增长了 23.7%。在中国与东盟贸易中位居前三位的东盟国家是马来西亚、泰国、新加坡。在双方贸易产品结构中，机电产品占 50%，其中电子产品是双方第一大类贸易产品。

截至今年 8 月，中国与东盟双向投资已近 900 亿美元，其中，中国企业对东盟的累计投资达 223 亿美元。中国企业在东盟的投资较东盟在华投资起步晚、数额少，但发展很快。

新闻 7：《人民日报海外版》（2011 年 11 月 10 日 第 02 版）

据新华社联合国 11 月 8 日电：联合国秘书长潘基文 8 日在纽约联合国总部呼吁中国企业履行自己的社会使命，共同推动社会的可持续发展，实现绿色增长的目标。

潘基文当天在同“中国企业家俱乐部”代表团举行的午餐会上发表讲话说，旨在推动经济可持续发展和社会效益共同提高的联合国“全球契约”计划已经吸引了 200 多家中国企业。“全球契约”的参加者认识到，当今世界全球化程度越来越高，企业的眼光应当放得长远一些，不仅要注重短期的财务收益，还要在社会、环境和道德方面体现自己的长期价值。

新闻 8：《人民日报海外版》（2011 年 11 月 07 日 第 04 版）

据新华社华盛顿电：中国驻美国大使馆日前举办活动，支持美方提出的“10 万人留学中国计划”。

中国驻美国大使张业遂说，在今年4月中国国务委员刘延东和美国国务卿希拉里·克林顿主持召开的中美人文高层磋商上，中方就表达了对“10万人留学中国计划”的支持，并承诺为留华美国学生提供更多奖学金，这一计划眼下已经取得不少进展。他说，中方将继续采取措施，支持“10万人留学中国计划”。

“10万人留学中国计划”由美国总统奥巴马在2009年最早提出。根据这一计划，美国将在4年内派遣10万名美国学生到中国学习。

新闻 9：《人民日报海外版》（2011年11月18日 第01版）

本报内蒙古四子王旗11月17日电：17日19时32分，圆满完成与天宫一号目标飞行器两次交会对接使命的神舟八号飞船，星夜降落于内蒙古四子王旗主着陆场。

至此，这艘11月1日5时58分从酒泉卫星发射中心升空的飞船，以在轨运行16天又13小时的时间和1100万公里的行程，成为迄今中国在太空飞行时间最久、飞行距离最长的飞船。

神八返回舱着陆后，中共中央、国务院、中央军委电贺天宫一号与神舟八号交会对接任务圆满成功。

新闻 10：《人民日报海外版》（2011年11月10日 第01版）

本报北京11月9日电：经济合作与发展组织（OECD）日前发布的《2011年科学、技术与产业计分榜：知识经济中的创新与增长》报告显示，2000年至2010年，全球主要国家的专利质量比1990年至2000年平均下降了20%，但以中国、印度和韩国为代表的亚洲国家的专利质量正显著提升。在过去10年中，中国成长为信息通信技术创新三强之一。

报告显示，来自美国、德国和日本的发明人所提交的专利，其引用率最高，但一些亚洲国家，如中国、印度和韩国的专利质量正显著提升。

韩语为原文

新闻 11：〈중앙일보〉(2011.03.05)

找到了在2月22日新西兰克莱斯特切奇发生的强震中失踪的韩国人兄妹中的哥哥刘吉焕的尸体。4日，外交部称：“韩国驻新西兰大使鲁光镒从新西兰警方得知，已找到了失踪的哥哥刘吉焕的尸体。”刘吉焕的尸体是在坎特伯雷电视台大楼残骸堆里找到的。一起失踪的刘吉焕的妹妹刘罗温的尸体还没有被找到。刘氏兄妹在2月22日白天12时51分发生6.2级强烈地震时，正在坎特伯雷电视台内的语言辅导班上課，大楼倒塌时被埋在了下面。

新闻 12：〈중앙일보〉(2011.03.25)

据了解，在多国军队的连续打击下，卡扎菲政权正在寻找“退出战略”，试图通过秘密接触美国等国家高层解决利比亚事态。另外，英国日刊《泰晤士报》(The Times) 22日报道称，卡扎菲政权的核心成员们为了挽救陷入危机的利比亚，似乎开始有背叛卡扎菲或亡命的倾向。但是，该报并未具体透露利比亚相关高层领导的身份。当日，美联社AP等外媒称：“为了解决旷日持久

的利比亚事态，卡扎菲方面正在同美国高层官员进行协商谈判的情况不断发生”，“最近，被卡扎菲军队扣留的《纽约时报》的 4 名记者被释放也是这些外交努力的结果”。

新闻 13: <중앙일보> (2011.03.28)

日本政府将于 30 日公布对加强描述独岛主权的中学教科书的审查结果。对此，韩国政府在 28 日着手准备对策。国务总理室和外交通商部、教育科学技术部等相关部门在当天举行了“独岛领土管理对策团”会议，讨论了着手维修独岛直升机机场工程和阻止问题教科书扩散的方案等。政府当局者表示，“政府一直在通过多种途径，敦促日本政府慎重对待此事”，“然而如果日本方面执意描述对独岛的主权，那么政府将不会仅停留在语言抗议上，将用行动说话”

新闻 14: <중앙일보> (2011.03.29)

在首尔等韩半岛也检测出了从日本福岛核电站泄露出的碘（iodine）和铯（cesium）等放射性物质。3 月 28 日深夜时分，韩国原子能安全技术院向外界紧急公布了消息，称“上周对从首尔等国内 12 个地方收集到的空中的尘埃进行了分析，结果显示，从其中几个地方检测出了放射性碘”。但同时韩国原子能安全技术院方也表示，已经检测出的放射性碘含量非常少，不会对人体造成伤害，准确的分析结果预计将在本月 29 日公布。对此，某原子能专家表示“放射性碘一旦被泄露，基本上铯也会同时外泄”，确认了铯已经扩散至韩半岛的事实。

新闻 15: <중앙일보> (2011.03.30)

3 月 29 日，韩国和朝鲜召开了以白头山火山问题为议题的民间专家会议，会议上双方就火山研究及准备对策的必要性交换了意见。在京畿道汶山都罗山韩朝出入事务所召开的会议上，朝鲜方面的团长尹英根火山研究所副所长在开场白中谈及了气象异常等问题，之后表示“气象现象和地震都不容易掌握”，提出了白头山火山协议的必要性。尹副所长称“日本地震之后，我们白头山地下水观测孔中的约 60 厘米的水在晃动，泉水中出现泥的现象很多”。

新闻 16: <중앙일보> (2011.07.08)

韩国和中国政府达成协议，宣布将于今年内开始自由贸易协定（FTA）的协商。随着协商开始消息的宣布，瞄准零关税或低关税的“早期收获（early harvest）事项”将立即适用于不太敏感的工业产品。

北京的某消息灵通人士表示“韩中两国将于 7 月 8 日在北京达成最终协议，宣布年内开始 FTA，并签署谅解备忘录（MOU）”，“为此，韩国外交部通商交涉本部 FTA 局长李泰镐已于本月 7 日抵达了北京”。他还转达说“在将于 8 日与中国商务部举行的局长级会谈中，双方将在 MOU 上签字”。

新闻 17: <중앙일보> (2011.10.05)

将年物价上涨率控制到 4% 的目标事实上已不可能实现。10 月 4 日，韩国统计厅数据显示，9 月的消费者物价同比上涨了 4.3%，虽然涨幅较 8 月下降了 1 个百分点，但是已经连续九个月超过了 4%。金戒指及房屋出租价格牵引了物价上涨。

得益于夏季大幅上涨的果蔬类价格下挫，9 月的新鲜食品价格同比下降了 7.4%。SK 电讯将移动通信费下调了 1000 韩元，这也削弱了物价上涨的势头。企划财政部经济政策局长崔相穆表示“中秋节后，由于需求下降和气候条件改善等原因，对果蔬类的需求逐渐平稳，这也有助于物价回落”。

新闻 18: <중앙일보> (2011.11.05)

首尔市政府称已决定把 2800 名非正式员工阶段性转成正式员工。首尔市计划马上确定转成正式员工的人力规模、所需预算和日程等方面所需的研究人员。将非正式员工转成正式员工，这是首尔市长朴元淳的承诺。

当天，就业政策课课长周勇泰表示“计划于明年初展开研究工作，如果能够在明年 7 月份得出结果，那么将通过部门协商确定最终方案”。首尔市政府的员工共有 1 万 6228 名，其中非正式员工有 1500 余名，首尔地铁、SH 等投资、出捐机构和办事处有非正式员工 1300 名，他们便是转正的对象。

新闻 19: <중앙일보> (2011.10.27)

无党派候选人朴元淳成功当选首尔市长是“政权审判论”和“安哲秀风暴”共同作用的成果。在这次选举中，在野圈通过选出唯一候选人放手一搏，最终时隔 10 年重新从执政党手中夺回了首尔市。

朴候选人和民主党方面分析认为，李明博政府进入第四年任期以来，民众对于政府和执政党的支持度急剧下降，这种情况下又爆出李明博总统的“内谷洞私宅问题”，更是给选举造成了致命的影响。朴候选人选举阵营的发言人禹相虎 26 日表示“内谷洞私宅风波进一步刺激了本来就因目前的经济情况而火大的市民们”，“这是对政权和大国家党的审判”。

新闻 20: <중앙일보> (2011.11.10)

韩国国立环境科学院 9 日发布了从 2004 年起进行的国家长期生态研究项目的中期调查结果，表示“韩半岛的年均气温正逐渐上升，韩国国内自然生态界的变化也日渐明显”

环境科学院预计，特别是目前仅出现在济州岛和南海岸部分地区的亚热带气候，到 2071 年以后会扩大到除白头大干的部分高山地区之外的韩国全境。

研究组指出，类似气候变暖的影响将使分布在智异山海拔 400 米地带的全南求礼郡土旨面的松树林中温带树种的松树的密度降低，取而代之的是亚热带树种不断增多。

附录 2 录音材料（韩语版）

诗 20 首

诗 1: 望岳

岱宗夫如何，齐鲁青未了。

仄平平平平 平平平仄仄
去平平去去 去上平入上

造化钟神秀，阴阳割昏晓。

仄仄平平仄 平平仄平仄
平平平平去 平平入平上

荡胸生层云，决眦入归鸟。

仄平平平平 仄仄仄平仄
去平去平平 入去入平上

会当凌绝顶，一览众山小。

仄平平仄仄 仄仄仄平仄
去去平入上 入上平平上

대종부여하 인가, 제노청미료 하네.

dAzoQbuyEha inga, zenocEQmiryo hane.

조화종신수 에, 음양할훈효 라.

zohwazoQsinsu e, lmyaQharhonhyo ra.

탕흥생충운 하고, 결자입귀조 하네.

taQhyuQsAQcIQun hago, gyErzaibgwizo hane.

회당능절정 하여, 일람중산소 하리.

HwedaQnIQzErzEQ hayE, irramzuQsanso hari.

태산의 큰줄기 어떠한가.

TAsanU kInzurgi ETEhanga.

제나라에서 노나라까지 푸르름 끝 없네.

zenaraesE nonaraKazi purIrIm Kit Ebne.

하늘의 조화 신묘하게 모아 놓으니

hanIrU zohwa sinmyohage moa noIni

산빛과 그림자 어둠과 새벽을 가르는구나.

sanbidgwa gIrimza Edungwa sAbyEgIr garInInguna
 솟아 오르는 층층구름에 가슴 벅차오르고
 sosa orInIn cIQcIQgurIme gasIm bEgcaorIgo
 두 눈을 부릅뜨고 돌아가는 새를 본다.
 du nunIr burIbTIgo doraganIn serIr bonda
 언젠가 반드시 정상에 높이 올라
 Enzenga bandIsi zuQsaQe nopi orra
 주위의 작은 산들을 굽어 보리라.
 zuwiU zagIn sandIrIr gubE borira

诗 2: 旅夜书怀

细草微风岸，危檣独夜舟。

仄仄平平仄 平平仄仄平
 去上平去去 平平入去平

星垂平野阔，月涌大江流。

平平平仄仄 仄仄仄平平
 平平平上入 入上去平平

名岂文章著？官应老病休。

平仄平平仄 平平仄仄平
 平上平平入 平去上去平

飘飘何所似，天地一沙鸥。

平平平仄仄 平仄仄平平
 平平上上上 平去入去平

세초미풍안 에, 위장독야주 라.

secomipuQan e, wizaQdogyazu ra.

성수평야활 하고, 월용대강류 하네.

suQsupyEQyahwar hago, wEryoQdAgaQryu hane.

명개문장저 인가, 관응노병휴 거늘.

myEQgAmunzaQzE inga, gwanIQnobyEQhyu gEnIr.

표표하소사 런가, 천지일사구 라네.

Pyopyohasosa rEnga, cEnziirsagu rane.

가는 풀이 미풍에 흔들리는 물가
 ganIn puri mipuQe hIndIrrinIn murga
 위태로운 돛대 아래 외로운 밤배
 witQroun doTA arA weroun bambA
 별이 내리니 들판이 광활하고
 byEri nArini dIrpani gwaQhwarhago
 달이 뜨니 거대한 강물이 흐른다
 dari TIIni gEdAhan gaQmuri hIrInda
 명성을 어찌 문장에서 얻겠는가
 myEQsuQIr ECi munzaQesE EdgednInga
 벼슬도 높고 병들면 그만 두는 것이 옳거늘
 byEsIrdo nIrKo byEQdIrmyEn gIman dunIn gEsi orkEnIr
 바람에 떠도는 몸 어디에 비길까
 BaramE TEdonIn mom Edie bigirKa
 천지간에 외로운 갈매기 한마리 뿐
 Cunzigane weroun garmAgi hanmari Pun

诗 3: 春望

国破山河在，城春草木深。

仄仄平平仄 平平仄仄平
 入去平平去 平平上入去

感时花溅泪，恨别鸟惊心。

平平平仄仄 仄仄仄平平
 上平平平去 去入上平平

烽火连三月，家书抵万金。

仄仄平平仄 平平仄仄平
 平上平去入 平平上入平

白头搔更短，浑欲不胜簪。

平平平仄仄 仄仄仄平平
 入平平平上 上入入平平

국파산하재 하여, 성춘초목심 하네.

gugpasanhazA hayE, suQcuncomogsim hane.

감시화천루 하고, 한별조경심 이라.

gamsihwacunru hago, hanbyErzokyEQsim ira.

봉화연삼월 하고, 가서저만금 이네.

boQhwayEnsamwEr hago, gasEzEmangIm ine.
 백두소경단 하여, 흑옥불승잠 이네.
 bAgdusogyEQdan hayE, hogyogbursIQzam ine.

나라는 망했지만 산천은 남아서
 naranIn maQhAdziman sancunIn namasE
 성에는 봄이 와 초목이 무성하네
 sEQenIn bomi wa comogi museQhane
 시대가 슬퍼 꽃조차 눈물을 뿌리게 하고
 sidAga sIrpE Kodzoca nunmurIr Purige hago
 슬픈 이별에 나는 새도 놀란다
 sIrpIn ibyEre nanIn sedo norranda
 봉화는 석달을 연이어 오르고
 boQhwanIn sEgdarIr yEniE orIgo
 집에서 온 편지는 너무나도 소중한구나
 zibesE on pyEnzinIn nEmunado sozuQhaguna
 흰 머리는 빗을수록 짧아져
 hin mErinIn blsIrsurog CarbazyE
 비녀조차 꽃을 수 없겠구나.
 blnyEzoca KozIr su EbgeKuna

诗 4: 天末怀李白

凉风起天末，君子意如何？

平平仄仄仄 平仄仄平平
 平去上平入 平上去去上

鸿雁几时到，江湖秋水多。

平仄仄仄仄 平平平仄平
 上去上平去 平平平上平

文章憎命达，魑魅喜人过。

平平仄仄平 平仄仄仄仄
 平平平去去 平去上平平

应共冤魂语，投诗赠汨罗。

平仄平平仄 平平仄仄平
 去平平平去 平平去入平

량풍기천말 한데, 군자의여하 인가.

ryaQpuQgicEnmar hande, gunzaUyEha inga.
 흥안기시도 한지, 강호추수다 라.
 hoQangisido hanzi, gaQhocusuda ra.
 문장증명달 하고, 리매희인과 하네.
 munzaQzIQmyEQdar hago, rimAhUingwa hane.
 응공원훈어 하니, 투시증먹라 라.
 IQgoQwonhonE hani, tusizIQmyEQna ra.

하늘가에 찬바람 벌써 부는데
 hanIrgae canbaram bErSE bunInde
 님의 아픈마음 어떠하신가
 nimU apInmaIm ETEhasinga
 기러기는(편지) 어느때나 찾아올런지
 girEginIn EnITAna cazaorrEnzi
 강호엔 가을물(풍파) 많기도 하더라만
 gaQhoen gaIrmul mankido hadEraman
 글을 잘해 운명이 시기를 했느니
 gIrIr zalhA unmyEQi sigirIr hAnnIni
 소인배들이 남의 곤경을 기뻐한다지
 soInbAdIri namU gongyEQIr giPEhandazi
 지나는길 원훈(굴원)이 말하고자 하리니
 zinanIngir wonhoni marhagozE harini
 시 한수 먹라강에 던져 주시라
 si hansu myEQnagaQe dEnzyE zusira

诗 5: 月夜

今夜鄜州月，闺中只独看。

平仄平平仄 平平仄仄平

平去平平入 平去上入平

遥怜小儿女，未解忆长安。

平平仄平仄 仄仄仄平平

平平上平上 去去入上平

香雾云鬟湿，清辉玉臂寒。

平仄平平仄 平平仄仄平

平去平平入

平平入去平

何时倚虚幌，双照泪痕干？

平平仄平仄

平仄仄平平

上平去平上

平去去平平

금야부주월 은, 규중지독간 하리.

gImyabuzuwEr In, gyuzuQzidoggan hari.

요련소아녀 는, 미해억장안 이라.

yoryEnsoanyE nIn, mihAEgzaQan ira.

향무운환습 하고, 청취옥비한 하리.

haQmuunhwansIb hago, cEQhwiogbihan hari.

하시의허황 에, 쌍조누흔건 가.

hasiUhEhwaQ e, SaQzonuhIngEn ga.

오늘 밤 부주의 달을

onIr bam buzU darIr

규중의 아내만 홀로 보고 있으리.

kyuzuQU anAman horro bogo iSIri

멀리 있는 가엾은 어린 아이들은

mErri innIn gayEbSIn Erin aidIrIn

장안의 나를 그리는 어미 마음 알지도 못하리.

zaQanU narIr gIrinIn Emi maIm alzido motari

향기로운 안개 구름같은 머리 결을 적시고

hyaQgiroun angA gurImgatIn mEri gyErIr zEgsigo

맑은 달빛에 옥같이 고운 팔 차갑겠소.

margIn darbice oKaci gown par cagabKeSo

어느 때 앳은 휘장 창가에 기대어

EnI TA yarbIn hwizaQ caQgae gidAE

달빛에 양 볼 흐르는 눈물자국 지울까...

darbice yaQ bor hIrInIn nunmurzagug ziurKa

诗 6: 客至

舍南舍北皆春水，但见群鸥日日来。

仄平仄仄平平仄 仄仄平平仄仄平
上平上入平平上 上去平平入入平

花径不曾缘客扫，蓬门今始为君开。

平仄平平平仄仄 平平平仄仄平平
平去入平去入上 平平平上平平平

盘飧市远无兼味，樽酒家贫只旧醅。

平平仄仄平平仄 平仄平平仄仄平
平平上去平去去 平上平平上去平

肯与邻翁相对饮，隔篱呼取尽馀杯！

仄仄平平平仄仄 仄平平仄仄平平
上上平平去去去 入平平上上平平

사람사복개춘수 하고, 단견군구일일래 하네.

sanamsabuggAcunsu hago, dangyEngunguirirra hane.

화경부증연객소 이나, 봉문금시위군개 하네.

hwagyEQbuzIQyEngAgso ina, boQmungImsiwigunga hane.

반손시원무겸미 하고, 준주가빈지구배 라.

bansonsiwonmugyEmmi hago, zunzugabinziguba ra.

금여인웅상대음 하라, 격리호취진여배 하리.

gIQyEinoQsaQdAIm harya, gyEQnihocwizinyEbA hari.

집의 앞뒤는 봄 물이 가득하고

zibU abdwinIn bom muri gadIghago

날마다 갈매기 떼 날아와 놀뿐

narmada garmAgi Te narawa norPun

꽃이 길을 덮도록 쓴 적 없는데

Koci girIr dEbTorog SIn jEg EmnInde

그대 오니 오늘에야 사립문 여네

gIdA oni onIreya sarimmun yEne

시장이 멀리 있어 상은 조출하고

sizaQi mErri iSE saQIn zocorago

가난하니 술통엔 묵은 탁주 뿐

gananhani surtoQen mugIn tagzu Pun

옆집 노인과도 같이 마실까
yEbzib noingwado gaci masirKa
울타리 너머 불러 마저 마시세
urtari nEmE burrE mazE masise

诗 7: 登高

风急天高猿啸哀，渚清沙白鸟飞回。

平仄平平仄仄仄 仄平平仄仄平平
去入平平平去平 上平去入上平平

无边落木萧萧下，不尽长江滚滚来。

平平仄仄平平仄 仄仄平平仄仄平
平平入入平平去 入上上平上上平

万里悲秋常作客，百年多病独登台。

仄仄平平平仄仄 仄平平仄仄平平
入上平平平去入 入平平去入平平

艰难苦恨繁霜鬓，潦倒新停浊酒杯。

平平仄仄平平仄 仄仄平平仄仄平
平去去去平平去 去去平平入上平

풍급천고원소애 하고, 저청사백조비회 라.

puQgIbcEngowonA hago, jEcEQsabAgzobihwe ra.

무변낙목소소하 하고, 부진장강곤곤내 라.

mubyEnnagmogsosoha hago, buzinzaQgongonnA ra.

만리비추상작객 이요, 백년다병독등대 라.

manribicusaQzaggAg iyo, bAgyEndabyEQdogdIQdA ra.

간난고한번상빈 하니, 요도신정탁주배 라.

gannangohanbEnsaQbin hani, yodosinzEQtagzubA ra.

바람 부는 쓸쓸한 가을에 원숭이 애닦게 울고,

baram bunIn SIrSIrhan gaIre wonsuQi Adarke urgo,

물 맑은 흰 모래 위에 갈매기 나네.

mur margin hin morA wie garmAgi nane.

사방의 나무에선 잎이 우수수 지는데,

sabaQU namuesEn ipi ususu zinInde

끝없는 강물은 굽이쳐 흐르네.
 KIdEbnIn gaQmurIn gubicyE hIrIne
 고향을 떠나 이 슬픈 가을나그네 신세로,
 gohyaQIr TE na i sIrpIn gaIrnagIne sinsero
 병든 몸을 이끌고 이곳에 올랐네.
 byEQdIn momIr iKIrgo igose orranne
 고생도 한스러운데 머리도 희어지니
 gosAQdo hansIrEunde mErido hiEzini
 늙은 몸이라 탁주잔도 들지 못하네.
 nIrgIn momira tagzuzando dIrzi motane

诗 8: 登楼

花近高楼伤客心，万方多难此登临。

平仄 平平平仄平 仄平平仄仄平平
 平去平平去入平 入平平去上平去

锦江春色来天地，玉垒浮云变古今。

仄平平仄平平仄 仄仄平平仄仄平
 上平平入平平去 入去平平去上平

北极朝廷终不改，西山寇盗莫相侵！

仄仄平平平仄仄 平平仄仄仄平平
 入入平平平入上 平平去去入去平

可怜后主还祠庙，日暮聊为梁父吟。

仄平仄仄平平仄 仄仄平平平仄平
 上平去上平平去 入去平平平上去

화근고루상객심 하고, 만방다난차등림 하네.

hwagIngorusaQgAgsim hago, manbyEQdanancadIQrim hane.

금강춘색래천지 하고, 옥루부운변고금 이라.

gImgaQcunsAgrAcEnzi hago, ogrubuunbyEngogIm ira.

북극조정종불개 하니, 서산구도막상침 이라.

buKIgzozEQzoQburgA hani, sEsangudomagsaQcim ira.

가련후주환사묘 하니, 일모료위양부음 이라.

garyEnhuzuhwansamyo hani, irmoryowiyaQbuIm ira.

누각 가까이 핀 꽃 나그네 맘 슬프게 하고

nugag gaKai pin Kod nagIne mam slrplge hago
 천하가 소란하여 나는 누각에 오르네.
 cEnhaga soranhayE nanIn nugage orIne
 금강의 봄빛은 천지에 가득하고,
 gImgaQU bombicIN cEnzie gadIghago
 옥루봉에 뜬 구름 여전히 변하고있네.
 oQnuboQe TIn gurIm yEzEnhi byEnhagoinne
 북극성같은 우리 조정 바뀔리 없으니,
 buggIgsEQgatIn uri zozEQ baKwirri EbSIni
 서산의 도적들아 침노하지 마라.
 sEsanU dozEgdIra cimnohazi mara
 가련한 후주도 종묘사직을 지켰나니
 garyEnhan huzudo zoQmyosazigIr zikyEdnani
 해 저물어 용사의노래가 읊어지노라.
 hA jEmurE yoQsaU norAga IrpEzinora

诗 9: 蜀相

丞相祠堂何处寻？锦官城外柏森森。

仄仄平平仄仄	仄仄平平仄仄
去去平平上上平	上平平去入平平

映阶碧草自春色，隔叶黄鹂空好音。

仄仄仄仄仄仄	仄仄平平仄仄
上平入上去平入	入入平平去去平

三顾频频天下计，两朝开济老臣心。

仄仄平平仄仄	仄平平仄仄平平
去去平平平去去	去平平上上平平

出师未捷身先死，长使英雄泪满襟！

仄仄仄仄平平	平平平平仄仄
去平去入平去上	上去平平去上平

승상사당하처심 인고, 금관성외백삼삼 이네.

sIQsaQsadaQhacEsim ingo, gImgwansuQwebAgsamsam ine.

영계벽초자춘색 하고, 격엽황리공호음 하네.

yEQgyebyEgcozacunsAg hago, gyEgyEbhwaQrigoQhoIm hane.

삼고빈번천하계 이요, 량조개제로신심 이네.

samgobinbEncEnhagye iyo, ryaQzogAzerosinsim ine.

출사미첩신선사 이니, 장사영웅루만금 이네.

cursamicubsinsEnsa ini, zaQsayEQuQrumangIm ine.

승상의 사당을 어디에서 찾을까

sIQsaQU sadaQIr EdiesE cazIrKa

금관성 밖 잣나무 우거진 곳이라네

gImgwansEQ bag zannamu ugEzin gosirane

섬돌에 비친 푸른 풀 절로 봄빛이요

sEmdore bicin purIn pur zErro bombiciyo

나뭇잎 사이의 꾀꼬리 무심히 즐겨 노래한다

namunnib saiU KweKori musimhi zIrgyE norAhanda

세 번 다시 찾은 번거로운 일도 천하 위한 계책이요

se bEn dasi cazIn bEngEroun irdo cEnha wihan gyecAgiyo

두 임금 섬겨 나랏일 해나감은 노신의 충성심이라네

du imgIm sEmgyE narannir hAnagamIn nosinU cuQsuQsimirane

군사를 내었으나 이기기도 전에 몸 먼저 죽으니

gunsarIr nAESIna igigido zEne mom mEnzE zugIni

길이 후대의 영웅들 옷깃에 눈물 가득 채우게 하네

giri hudAU yEQuQdIr oegise nunmur gadIg cAuge hane

诗 10: 闻官军收河南河北

剑外忽传收蓟北，初闻涕泪满衣裳。

仄仄平平平仄仄 平平仄仄仄平平

去去入去去去入 平去上去上去平

却看妻子愁何在，漫卷诗书喜欲狂。

平平仄仄平平仄 仄仄平平仄仄平

入平去上平上去 去去平平上入去

白日放歌须纵酒，青春作伴好还乡！

仄仄平平平仄仄 平平仄仄仄平平

入入上平平去上 平平去去去平平

即从巴峡穿巫峡，便下襄阳向洛阳。

平平仄仄平平仄 仄仄平平仄仄平

入平平入去平入 平去平平去入平

검외홀전수계배 하니, 초문체루만의상 이라.

gEmwehorzEnsugyebA hani, comuncerumanUsaQ ira.

각간처자수하재 인고, 만권시서희욕광 이라.

gaggancEzasuhazA ingo, mangwonsisEhiyoggwaQ ira.

백수방가수종주 하고, 청춘작반호환향 하리.

bAgsubaQgasuzoQzu hago, cEQcunzagbanghohwanhyaQ hari.

즉종파협천무협 하여, 편하양양향낙양 하리.

zIgzoQpahyEbcEnmuhyEb hayE, pyEnhayaQyaQhyaQnagyaQ hari.

홀연히 검각산 밖에서 하남하북 수복 소식이 전해지니
horyEnhi gEmgagsan baKesE hanamhabug subog sosigi zEnhAzini
처음 듣고는 눈물이 옷에 가득하여라
cEIm dIdgonIn nunmuri ose gadIghayEra
돌아가 처자를 만나면 무슨 걱정일까
doraga cEzarIr mannamyEn musIn gEgzEQirKa
아무렇게나 책 덮고 기뻐서 미칠 것 같아라
amurEkena cAg dEbgo giPEsE micir gEd gatara
흰머리로 노래하며 미친 듯 술을 마시며
hinmEriro norAhamyE micin dId surIr masimyE
한창의 봄을 벗삼아 기분좋게 고향에 돌아가리라
hancaQU bomIr bEdsama gibunzoke gohyaQe doragarira
곧장 파협을 따라 무협을 뚫고 지나
godzaQ pahyEbIr Tara muhyEbIr Turko zina
바로 양양으로 내려가 낙양을 향하리로다
baro yaQyaQIro nAryEga nagyaQIr hyaQhariroda

诗 11: 江雪

千山鸟飞绝，万径人踪灭。

平平仄平平 仄仄平平仄
平平上平入 入去平平入

孤舟蓑笠翁，独钓寒江雪。

平平平仄平 平仄平平仄
平平平入平 入去平平入

천산조비절 이요, 만경인종멸 이라.
 cEnsanzobizEr iyo, mangyEQinzoQmyEr ira.
 고주사립웅 이, 독조한강설 이라.
 gozuseriboQ i, dogzohangaQsEr ira.

천산에는 새 한 마리 날지 않고,
 cEnsanenIn sA han mari narzi anko
 만 갈래 산길은 사람 발길 끊어졌네.
 man garrA sangirIn saram bargir KInEzyEnne
 외로운 배에 삿갓 쓴 노인
 weroun bAe sadgad Sin noin
 눈 쌓인 강가에 홀로 낚시하네.
 nun Sain gaQgae horro nagsihane

诗 12: 登鹳雀楼

白日依山尽，黄河入海流。

仄仄平平仄 平平仄仄平
 入入平平上 平平入上平

欲穷千里目，更上一层楼。

仄平平仄仄 仄仄仄平平
 入平平上入 平上入平平

백일의산진 이요, 황하입해류 라.
 bAgirUsanzin iyo, hwaQhaipAryu ra.
 육궁천리목 하여, 갱상일층루 라.
 yogguQcEnrimog hayE, gAQsaQircIQnu ra.

해는 산 너머 지고,
 hAnIn san nEmE zigo
 황하는 바다로 흘러 들어가는데.
 hwaQhanIn badaro hIrrE dIrEganInde
 저 멀리 고향 길 굽어보려,
 zE mErri gohyaQ gir gubEboryE
 다시 한 층 더 누각을 오르네.
 dasi han cIQ dE nugagIr orIne

诗 13: 悯

锄禾日当午，汗滴禾下土。

平平仄平平	仄仄平平仄
平平入去上	平入平去上

谁知盘中餐，粒粒皆辛苦。

平平平平平	仄仄平平仄
平平平去平	入入平平去

서화일당년 하고, 한적화하토 하네.

sEhwairdaQnyEn hago, hanzEkwahato hane.

수지반중찬 이, 입입개신고 라네.

suzibanzuQcan i, ibibgAsingo rane.

작열하는 태양 아래 김을 매고

zaQnyErhanIn tAyaQarA gimIr mAgO

땀방울과 함께 버를 심는다

TambaQurgwa hamKe byErIr simnInda

누가 알리오 상 위에 오른 밥이

nuga arrio saQ wie orIn babi

한 알 한 알 모두 고된 노동의 산물임을

han ar han ar modu godwen nodoQU sanmurimIr

诗 14: 春晓

春眠不觉晓，处处闻啼鸟。

平平仄仄仄	仄仄平平仄
平平入去上	上上去平上

夜来风雨声，花落知多少。

仄平平仄平	平仄平平仄
去平去去平	平入平平去

춘면불각효 하니, 처처문제조 라.

cunmyEnburgakyo hani, cEcEmunzezo ra.

야래풍우성 에, 화락지다소 라.

yarApuQusEQ e,hwaragzidaso ra.

봄잠에 취해 날이 밝는 줄 몰랐네
bomzame cwihA nari baQnIn zur morranne
여기저기서 들려오는 새 울음소리
yEgizEgisE dIrryEonIn sA urImSORI
밤새 비바람이 몰아쳐
bamsA bibarami moracyE
꽃은 또 얼마나 떨어졌을까.
KocIn To Ermana TErEzyESIrKa

诗 15: 静夜思

床前明月光，疑是地上霜。

平平平仄平 平仄仄仄平
平平平入去 平上去上平

举头望明月，低头思故乡。

仄平仄平仄 平平平仄平
上平平平入 平平去去平

상전명월광 은, 의시지상상 이라.

saQzEnmyEQwErgwaQ In,UsizisaQsaQ ira.

거두망명월 이요, 저두사고향 이라.

gEdumaQmyEQwEr iyo,zEdusagohyaQ ira.

침상 머리에 밝은 달빛
cimsaQ mErie bargIn darbid
땅 위에 내린 서리런가.
TaQ wie nArin sErirEnga
머리 들어 밝은 달 바라보다,
mEri dIrE bargIn dar baraboda
고개 숙여 고향을 그리워하네.
gogA sugyE gohyaQIr gIriwEhane

诗 16: 绝句

两个黄鹂鸣翠柳，一行白鹭上青天。

仄仄平平平仄仄 仄平仄仄仄平平
去去平平平去上 入去入去上平平

窗含西岭千秋雪，门泊东吴万里船。

平平平仄平平仄 平仄平平仄仄平
平平平平平平入 平入平平入上平

양개황리명취류 하여, 일행백로상청천 하네.

yaQgAhwaQrimyEQcwiryu hayE, irhAQbAQnosaQcEncEn hane

창함서령천추설 하고, 문박동오만리선 이라.

caQhamsEryEQcEncusEr hago, munbagdoQomanrisEn ira.

한쌍 꾀고리가 푸른 버들에서 지저귀며,

hanSaQ KweKoriga purIn bEdIresE zizEgwimyE

한 줄 나는 백로는 푸른 하늘로 오르네.

Han zur nanIn bAQnonIn purIn hanIrro orIne

창은 서쪽 고래의 천년의 눈을 머금고,

caQIn sECog gorAU cEnnyEnU nunIr mEgImgo

문앞엔 동오의 만리길 달려온 배가 있네.

munapen doQoU manrigir darryEon bAga inne

诗 17: 戏为六绝句 (一)

庾信文章老更成，凌云健笔意纵横。

仄仄平平仄仄平 平平仄仄仄平平
上去平平上平平 平平去入去去平

今人嗤点流传赋，不觉前贤畏后生。

平平仄仄平平仄 仄仄平平仄仄平
平平平上平去去 入去平平去去去

유신문장로갱성 하여, 릉운건필의종횡 하네.

yusinmunzaQrogAQsEQ hayE hayE, rIQungunpirUzoQhweQ hane

금인치점류전부 나, 불각전현외후생 이라.

gImincizEmryuzEnbu na, burgagzEnhyEnwehusAQ ira

유신의 문장은 늙어 더욱 격조가 높아져

yusinU munzaQIn nIrgE dEug gyEgjoga nopazyE

구름을 넘는 듯 굳건하고 의미도 종횡부진하였다

gurImIr nEmnIn dId gudgEnhago Umido zoQhweQbuzinhayEdda

요즈음 사람들 전하는 부를 꼬집어 비웃지만

yoziIm saramdIr zEnhanEn burIr KozibE biudziman
 먼저 이룬 사람이 후생을 두려워한다는 깨닫지 못하네.
 mEnjE irun sarami husAQir duryEwEhandanIn KAdadzi motane

诗 18: 戏为六绝句 (二)

王杨卢骆当时体，轻薄为文晒未休。

平平仄仄平平仄 平仄仄平仄仄平
 去平平入去平上 去入平平上去平

尔曹身与名俱灭，不废江河万古流。

仄平平仄平仄仄 仄仄平平仄仄平
 上平平上平平入 入去平平入上平

양왕노락당시체 에, 경박위문신미휴 하네.

yaQwaQnoragdaQsice e, gyEQbagwimunsinmihyu hane

이조신여명구멸 하나, 불폐강하만고류 하리.

izosinyEmyEQgumyEr hana, burpyegaQhamangoryu hari

양왕과 노락의 당시의 문체를

yaQwaQgwa noragU daQsiU muncerIr

경박하게 글을 지어 아름답지 않다고 비웃네

gyEQbakage gIrIr ziE arImdabzi antago biunne

너희들은 몸과 이름 다 없어지나

nEhidIrIn momgwa irIm da EbSEzina

강물은 만고에 흐름을 그치지 않으리.

gaQmurIn mangie hIrImIr gIcizi anIri

诗 19: 戏为六绝句 (五)

不簿今人爱古人，清词丽句必为邻。

仄仄平平仄仄平 平平仄仄仄平平
 入入平平去上平 平平去去入平平

窃攀屈宋宜方驾，恐与齐梁作后尘。

仄平仄仄平平仄 仄仄平平仄仄平
 入平入去平平去 去上去平去去平

불박금인애고인 하여, 청사려구필위린 하네.

burbaggIminAgoIn hayE, cyEQsaryEgupirwirin hane

절반굴송의방가 하여, 공여제량작후진 이라.

zErbangursoQUbaQga hayE, goQyEzeryaQzaghujin ira

지금 사람 가벼이 말고 옛 사람 좋아하여

zigIm saram gabyEi margo yEd saram zoahayE

맑고 고운 시는 본받아 이웃삼아야 하네

marKo gown sinIn bonbada iudsamaya hane

굴원과 송옥을 다잡고서 같은 수준이라 여겨

Gurongwa soQogIr dazabgosE gatIn suzunira yEgyE

제나라와 양나라 처럼 뒷 세상 티끌 될까 두렵네.

Zenzrzwa yaQnara cErEm dwid sesaQ tiKIr dwerKa duryEbne

诗 20: 戏为六绝句 (六)

未及前贤更勿疑，递相祖述复先谁？

仄仄平平仄仄平

去入平平平入平

仄平平仄仄平平

上去上入入去平

别裁伪体亲风雅，转益多师是汝师！

仄平平仄平平仄

入去去上去去上

仄仄平平仄仄平

去入平平上上平

미급전현갱물의 하라. 체상조술부선수 런가.

migIbzEnhyEngAQmurU hara. cesaQzosurbusEnsu rEnga.

별재위체친풍아 하니, 전익다사시여사 니라.

byErzAwicecinpuQa hani, zEnigdasasiyEsa nira.

앞 현인에게 미치지 못함을 의심하지 말아라

ab hyEninege micizi motamIr Usimhazi marara

저마다 서로 베끼니 누가 앞설 수 있겠는가

zEmada sEro bikini nuga absEr su idgennInga

거짓 문체를 가려내야 풍아와 가까워지나니

gEzid muncerir garyEnAya puQawa gaKawEzinani

더욱 보태어 스승이 많아지는 것, 이것이 곧 너희 스승이다.

dEug botAE sIsIQi manazinEn gEd igEsi god nEhU sIsIQida

韩语罗马字标记参考表

1. Consonants				2. Vowels		
Hangul	[IPA]	Roman letters		Hangul	[IPA]	Roman letters
		Onset	Coda			
ㅂ	[p]	b	B	아	[a]	a
ㄷ	[t]	d	D	어	[ə]	E
ㄱ	[k]	g	G	오	[o]	o
ㅈ	[tʃ]	z	D	우	[u]	u
ㅊ	[pʰ]	p	B	으	[i]	I
ㅌ	[tʰ]	t	D	이	[i]	i
ㅋ	[kʰ]	k	G	예	[e/ɛ]	e
ㅍ	[pʰ]	c	D	애	[ɛ]	A
ㅍ	[pʰ]	c	D	의	[wi]	U
ㅍ	[pʰ]	c	D	야	[ja]	ya
ㅍ	[pʰ]	c	D	여	[jə]	yE
ㅍ	[pʰ]	c	D	요	[jo]	yo
ㅍ	[pʰ]	c	D	유	[ju]	yu
ㅍ	[pʰ]	c	D	예	[je/jɛ]	ye
ㅍ	[pʰ]	c	D	애	[jɛ]	yA
ㅍ	[pʰ]	c	D	와	[wa]	wa
ㅍ	[pʰ]	c	D	워	[wə]	wE
ㅍ	[pʰ]	c	D	외	[we/wɛ]	we
ㅍ	[pʰ]	c	D	왜	[wɛ]	wA
ㅍ	[pʰ]	c	D	위	[wi]	wi

此表来自于 Sun-Ah, Jun(2000) “K-ToBI (Korean ToBI) Labelling Conventions”,
UCLA Working Papers in Phonetics 99

韩语汉字音系与《切韵》音系比较

声母表之一

		清		全 濁	
幫組		幫 補 po 布 p'ó	滂 浦 p'ó 普 po	並 菩 po 蒲 p'ó	
非組		非 封 poŋ 風 p'ũŋ	敷 峰 poŋ 豐 p'ũŋ	奉 逢 poŋ	
端泥組	一二	端 斗 tu 鬥 t'u	透 透 t'u 滔 to	定 頭 tu 投 t'u	
	三四	帝 tʃe 諦 tʃe	替 tʃe	弟 tʃe 遞 tʃe	
精組		精 箭 tʃən 薦 tʃən	清 淺 tʃən 切 tʃəl	從 錢 tʃən 踐 tʃən	
知組		知 中 tʃuŋ 忠 tʃuŋ 桌 t'ak	徹 寵 tʃoŋ	澄 重 tʃuŋ 蟲 tʃuŋ 濁 t'ak	
莊組		莊 莊 tʃaŋ 捉 tʃak	初 創 tʃaŋ	崇 牀 saŋ	
章組		章 蒸 tʃiŋ 彰 tʃaŋ	昌 稱 tʃiŋ 叱 tʃil	船 乘 siŋ	
見曉組		見 京 kiəŋ 荆 hiəŋ	溪 卿 kiəŋ 抗 haŋ	羣 鯨 kiəŋ	
影組		影 優 u 歐 ku			

声母表之二

		次 濁		清	全 濁
幫組		明 模 mo			
非組		微 無 mu			
端泥組	一二	泥 紐 niu 質 im	來 樓 lu		
	三四				
精組				心 仙 sən 硝 tʃo	邪 旋 sən
莊組				生 霜 saŋ	
章組				書 升 siŋ	禪 承 siŋ
日組		日 日 il			
見曉組		疑 迎 iəŋ 驗 həm		曉 亨 hiəŋ 呵 ka	匣 衡 hiəŋ 匣 kap 完 uan
影組		雲 尤 u	以 由 iu		

韵母比较表之一

	一 等			二 等		
	幫系	端系	見系	幫系 (泥組)	知莊組	見系
果開		多 ta 大 te 左 tʃua	我 a 個 ke			
果合	婆 p'a	妥 t'a 坐 tʃua 鎖 sue	火 hua 倭 ue			
假開				馬 ma 罵 me (拿茶 ta na)	①	家 ka
假合						瓜 kua
遇合	模 mo 部 pu 墓 mio	圖 to 杜 tu 賂 lə	古 ko			
蟹開	貝 p'ɛ	戴 tɛ 猜 si 賴 lə	改 ke	買 me 派 p'a	債 tʃɛ 差 tʃa	皆 ke 佳 ka 界 kie 矮 ue
蟹合	配 p'ɛ	內 ne 碎 sue 腿 t'ɔ	潰 kue 灰 hɔ			歪 ue 怪 kɔ 話 hua
效開	毛 mo	早 tʃo 套 t'u 腦 nɔ	高 ko	飽 p'ɔ 卯 mio ②	抄 tʃo	敲 ko 巧 kio
流開	茂 mu 某 mo	走 tʃu	狗 ku			

韵母比较表之二

		三 四 等						見 系	
		幫 系	端 系			知 系			
			端組	泥組	精組	莊組	知章組		日母
果開								茄 ka	
果合								靴 hua	
假開				寫 sa		遮 tʃa	惹 ia	夜 ia	
遇 魚 合			慮 lia	序 sə	詛 tʃə	書 sə	如 iə	許 hə	
					所 so	除 tʃe		與 iə 預 ie	
	虞	府 pu 甫 po	屢 lu	需 su 取 tʃy	數 su	住 tʃu	儒 iu	具 ku 裕 iu 娛 o	
蟹開	閉 pʃe 米 mi	帝 tʃe 低 tʃə	禮 lie 麗 lia 泥 ni	齊 tʃe 西 sə		製 tʃe 誓 sə		雞 kie 憩 ke 奚 he	
	蟹合	肺 pʃe		歲 se 脆 tʃy		稅 se ③	芮 ie	桂 kie 圭 kiu 衛 y	
止開	皮 pʃi 寐 me	地 tʃi	理 li	子 tʃa 梓 tʃe	師 sa 柿 si④	詩 si	而 i 兒 a	奇 ki 宜 ii	
止合	飛 pi		淚 lu 類 liu	隨 su 醉 tʃy ⑤	帥 su 衰 sə	水 su 吹 tʃy⑥		規 kiu 軌 kue 季 kie 貫 ky 畏 o	
效開	廟 mio 秒 tʃo	雕 tʃo	料 lio	笑 so		照 tʃo	擾 io	要 io 叫 kiu	
流 尤 開	否 pu 謀 mo		流 liu	秋 tʃu 就 tʃy	瘦 su	手 su⑦ 臭 tʃy 稠 tʃo	柔 iu	友 u 有 iu	
	幽	謬 liu						幽 iu	

韵母表之三

	一 等			二 等		
	幫系	端系	見系	幫系（泥組）	知莊組	見系
咸舒開		三 sam	暗 am		衫 sam	鹹 ham
山舒開		單 tan ①	安 an 乾 kən	班 pan 扮 pun	山 san	眼 an
山舒合	半 pan	短 tan	換 huan		撰 tʃan	關 kuan 灣 man
臻舒開			痕 hin 恨 han			
臻舒合	本 pon 門 mun 悶 min	頓 ton 鈍 tun ②	困 kon			
宕舒開	忙 mar	塘 tar	康 kaŋ			
宕舒合			黃 huaŋ			
江舒開				邦 paŋ 棒 poŋ	窗 tʃaŋ	講 kaŋ
曾舒開	崩 puŋ	等 tiŋ	肯 kiŋ 恒 haŋ			
曾舒合			弘 hoŋ			
梗舒開				猛 meŋ（冷 leŋ ③）	爭 tʃeŋ	行 heŋ 衡 hiəŋ
梗舒合						礦 kuaŋ 橫 həŋ
通舒合	蒙 moŋ	冬 toŋ	公 koŋ			

韵母比较表之四

		三 四 等							
		幫系	端系			知系			見系
			端組	泥組	精組	莊組	知章組	日母	
咸舒開	𪛗 p'iam	點 tʃəm	念 niəm	尖 tʃəm 潛 tʃəm		閃 səm	染 iam	謙 kiəm 劍 kam 欠 him	
咸舒合	犯 pəm								
深舒開	品 p'um		林 lim	心 sim	參 sam	深 sim	任 im	金 kim 飲 im	
山舒開	變 piən 婉 man	天 tʃən	年 niən	淺 tʃən		展 tʃən	然 iən	牽 kiən 件 kam 咽 in	
山舒合	反 pan 煩 pən		戀 liən	全 tʃən		船 sən	軟 iən	阮 uan 篆 iən 速 uən	
臻舒開	真 臻	貧 pin	鄰 lin	新 sin		真 tʃin	人 in	引 in 僅 kin 巾 kən	
	殷							近 kin	
臻舒合	諄		輪 liun	筍 sun 迅 sin		春 tʃun	潤 iun	均 kiun	
	文 分 pun							軍 kun	
宕舒開			涼 lian 娘 naŋ	想 saŋ	霜 saŋ	長 tʃaŋ	讓 iaŋ	央 aŋ 羊 iaŋ	
宕舒合	方 paŋ							狂 kuaŋ	
曾舒開	冰 piŋ		陵 liŋ			稱 tʃiŋ 蒸 tʃiŋ		孕 iŋ 應 iŋ	
梗舒開	平 p'iaŋ 盟 meŋ 聘 piŋ	頂 tʃəŋ	領 liəŋ	清 tʃəŋ		正 tʃəŋ		輕 kiəŋ	
梗舒合								兄 hiəŋ	

通 舒 合	東	風 p'ung 鳳 pou	隆 liung	崇 sung	中 t'ung 終 t'ong	絨 iung 融 iung	弓 kung 融 iung
	鍾	奉 pou	濃 non 龍 lion	松 song	鍾 t'ong 重 t'ung	茸 ion	其 kong 胸 hiung

韵母比较表之五

	一 等			二 等		
	幫系	端系	見系	幫系	知莊組	見系
咸入開		蠟 lap	合 hap、hop		插 sap	匣 kap 狹 hiap 恰 hip
山入開		辣 lal	割 hal	拔 pal	殺 sal	轄 hal
山入合	末 mal	脫 t'al 撮 t'ual	活 hual		①	滑 hual
臻入合	沒 mol 勃 pal [不] pul	突 tol	忽 hol 窟 kul			
宕入開	薄 pak②	拓 t'ak、t'ək	鶴 hak			
宕入合			郭 kuak			
江入開				剝 pak	捉 t'ək	學 hak 殤 huak
曾入開	北 puk	得 tik 德 tək 塞 sek 則 t'ik、t'ik	黑 hik 刻 kak			
曾入合			國 kuk 或 hok			
梗入開				白 pek 迫 pak	澤 rək 摘 t'ək	客 kek 革 hiək
梗入合						獲 hək
通入合	卜 pok	毒 tok	屋 ok			

韵母比较表之六

		四 等					見 系				
		幫 系	端 系			知 系					
			端組	泥組	精組	莊組				知 章 組	日 母
咸 入		疊 tʃəp	淵 liəp	接 tʃəp		涉 səp		業 əp 葉 iəp			
咸 合	法 pəp 乏 pʰip										
深 入			立 lip	集 tʃip 習 sip		十 sip 拾 sip	入 ip	急 kip			
山 入	減 miəl	鐵 tʃəl	列 liəl	切 tʃəl		舌 səl	熱 iəl	子 hiəl 歌 həl 謁 al			
山 合	伐 pəl 發 pal		劣 liəl	絕 tʃəl		拙 tʃəl 說 səl		血 hiəl 月 uəl 曰 ual			
臻 入	質 pʰil		栗 liul	漆 tʃil 膝 sil	瑟 sil	失 sil	日 il	吉 kil 乙 il			
臻 入	迄							乞 kəl 契 kil			
臻 入	術		律 liul	戌 sul 恤hiul	率 sol	出 tʃʰul		橘 kiul			
臻 入	物 mul							屈 kul			
宕 入			略 liak	削 sak		酌 tʃʰak	弱 iak	約 iak 腳 kak			

宕 入	縛 pak							
曾 入	逼 pʰip		匿 nik 力 liək	息 sik 卽 tʃik	側 tʃik 色 sek	直 tʃik ①		翼 ik 極 kik 億 ək
曾 入								域 iək
梗 入	壁 piək	敵 tʃək	歷 liək 溺 nik	席 sək ②		尺 tʃək		譯 iək 益 ik 劇 kik 液 ek
梗 入								役 iək
通 入	屋 pok		六 liuk	宿 suk	縮 tʃuk	熟 suk	肉 iuk	菊 kuk 育 iuk
通 入	燭		綠 lok	足 tʃok		束 sok	辱 iok	玉 ok 欲 iok

此表来自于朴正俸（2009）韩国汉字音系与厦门方言音系比较研究，四川出版集团，p65-79.

新闻 20 篇

新闻 1: 《人民日报海外版》(2011 年 03 月 07 日 第 04 版)

신화사 베이징 3 월 6 일 보도: 11 기 전국인대 4 차 회의는 6 일 오후 인민 대회당에서 기자 간담회를 진행했다. 국가 발전 및 개혁위원회 장핑(張平)주임, 주즈신(朱之鑫), 쉬썬핑(徐憲平) 부주임이 '12 차 5 개년' 계획 강요와 관련해 기자들의 질문에 답했다. 장핑은 "11 차 5 개년 계획 시기 중국은 경제사회 발전에서 세계가 주목하는 성과를 이룩했다. 동시에 중국이 발전하는 과정에 불균형, 부조화, 발전이 지속적으로 이루어지지 않는 등 몇 가지 문제점이 여전히 존재하고 있다. '12 차 5 개년' 계획을 작성하는 과정에서 이런 문제와 관련하여 깊이 연구했다. 앞으로 5 년간의 발전에서 이와 같은 문제를 잘 해결해야 한다. 이는 '12 차 5 개년' 계획을 구성하는 특징이기도 하다"고 밝혔다.

新闻 2: 《人民日报海外版》(2011 年 03 月 21 日 第 04 版)

본보 베이징 3 월 20 일 보도: 천더밍 중국 상무부장은 오늘 베이징 '2011 년 중국 발전고위층 포럼'에서 '12 차 5 개년' 기간 중국은 수출입과 외자유치·대외투자를 각각 동등하게 중시하고 대외개방수준을 높여 경제발전방식 전환의 원동력으로 삼겠다고 발표했다. 천더밍은 지난 10 년간 중국은 세계무역기구 가입국으로서 규정을 배우고 이행하던 수준에서 점차 규정을 제정하는 참여국이 되었다고 밝혔다. 이어 중국정부는 각고의 노력으로 법률·법규와 부문 규정 3,000 개를 정리했으며 평균관세를 15%에서 10%로 낮추고 100 여개 서비스 무역부문을 개방했다고 말했다.

新闻 3: 《人民日报海外版》(2011 年 04 月 14 日 第 02 版)

국가 에너지국 국장 류데난은 "중국과 브라질 간 석유가스와 전력 및 신에너지자원 등 각 분야의 협력에 큰 전망이 있다. 우선 중국은 안정적이고 끊임없이 발전하는 시장이 있다. 둘째, 중국은 탐사개발과 가공 방면에서 자금력과 기술력 및 관리 능력이 있다."고 말했다. 류데난은 "브라질은 이미 중국 석유수입 10 대국가에 속한다. 양국간에 특히 중화공분야와 석유분야에서의 교류횟수가 늘어나고 무역량의 균형이 잡히면, 양국간의 보조적역할이 제대로 드러날 것이다"라고 하면서 "12 차 5 개년 기간 중국경제발전과 에너지 발전은 커다란 기회이다. 앞으로 브라질 기업과 중국 기업이 각종 방식으로 협력을 진행하길 바란다"고 밝혔다.

新闻 4: 《人民日报海外版》(2011 年 04 月 13 日 第 04 版)

4 월 12 일 오전, 국가 관광국은 베이징 뉴스브리핑에서 『국무원의 '중국 관광의 날' 설립을 동의하는 비준』에 근거하여 2011 년부터 매년 5 월 19 일을 '중국 관광의 날'로 정했다고 발표했다. 이날은 『서하객 여행기(徐霞客遊記)』 출간일이기도 하다. 국가 관광국 주산중(祝善忠) 부국장은 『서하객 여행기』 출간일을 '중국 관광의 날'로 정한 것은 문화와 관광의 밀접한 연관을 보여 주는 것이라고 밝혔다. 서하객은 중국 명나라 때의 위대한 여행가이자 지리학자이며, 사학가이자 문학가이다. 『서하객 여행기』는 중국의 지형·지질구조를 계통적으로 고찰한 지리서의 명작이며 중국의 풍경과 자원을 서술한 여행답사기로서 국내외에 밀접한 영향력을 가지고 있다.

新闻 5: 《人民日报海外版》(2011 年 02 月 18 日 第 01 版)

신화사 베이징 2 월 17 일 보도: 국가외환관리국의 17 일 발표에 따르면 중국에 유입되는 대규모 해외자금이 실물경제활동과 기본적으로 부합된다고 밝혔다. 비록 개미 무리가 이사 가듯 핫머니가 불법으로 유입되고 있지만, 정규 금융기구의 대규모

집중유입상황은 발견하지 못 했다. 국가외환관리국 국제수지 분석팀이 발표한 『2010 년 중국 해외자금유동감사보고』에 따르면 2010 년 중국 핫머니 유입이 355 억 달러로 외환비축 증가량의 7.6%를 차지한다. 하지만 이 데이터 모두가 불법적인 해외자금의 유동은 아니라고 분석했다.

新闻 6: 《人民日报海外版》(2011 年 11 月 21 日 第 02 版)

본보 베이징 보도 : 기자가 중국-아세안 상무 이사회로부터 알아본 바에 의하면 올해 1 월부터 10 월까지 중국과 아세안 무역액이 2959 억 달러로 지난해 동기 대비 25.7% 성장했다. 그 중 중국이 아세안으로부터의 수입은 27.5%, 중국의 아세안에 대한수출은 23.7% 성장했다. 중국과 아세안과의 무역 랭킹 3 위권 나라들로는 말레이시아, 태국, 싱가포르 등이 있다. 양국 무역 상품 구조에서 전자기계 품목이 50%를 차지하고 있으며, 그 가운데 전자제품이 양국 최대 무역아이템이다.

올해 8 월까지 중국과 아세안 쌍무 투자는 이미 900 억 달러에 이르렀다. 그 중 중국 기업의 아세안 투자는 총 223 억 달러이다. 중국 기업의 아세안 투자는 아세안 기업의 중국투자에 비해 늦게 시작이 되고 액수도 적으나 발전이 빠르다.

新闻 7: 《人民日报海外版》(2011 年 11 月 10 日 第 02 版)

신화사 유엔 11 월 8 일 보도 : 반기문 유엔사무총장은 8 일, 뉴욕 유엔 총부에서 중국 기업이 자신의 사회적 사명을 이행하고 사회의 지속가능한 발전을 공동으로 추진하는 동시에 녹색 성장의 목표를 실현해야 한다고 말했다.

반기문 총장은 이날 ‘중국 기업가 클럽’ 대표단과의 오찬회에서 다음과 같이 말했다. 경제의 지속가능한 발전과 사회 이익을 공동으로 제고하는 문제에 관한 유엔 ‘글로벌 계약’계획에 이미 200 여개 중국기업이 참가했다. 반총장은 ‘글로벌 계약’ 참가 기업들은 세계의 글로벌 환경 속에서 멀리 내다보고 단기적인 재무 수익을 낼 뿐만 아니라 사회, 환경과 윤리 방면에서 각 기업의 가치를 실현해야 한다고 말했다.

新闻 8: 《人民日报海外版》(2011 年 11 月 07 日 第 04 版)

신화사 워싱턴 보도 : 미국 주재 중국대사관은 최근 행사를 개최해 미국 측이 제기한 ‘10 만 명 중국유학계획’을 지지했다.

장예수이(張業遂)주미 대사는 올해 4 월 류연동(劉延東)중국 국무원 부총리 와 힐러리 미 국무장관이 진행한 중미 인문 고위층 협상에서 중국 측은 ‘10 만 명 미국인 중국 유학 계획’을 지지할 것을 밝혔다. 또한 중국에서 공부하는 미국 유학생들에게 보다 많은 장학금을 제공하기로 약속했으며, 이 계획은 이미 진전이 됐다고 말했다. 그는 중국 측은 계속적으로 조치를 취해 ‘10 만 명 중국 유학 계획’을 지지하겠다고 덧붙였다.

‘10 만 명 중국 유학계획’은 오바마 미국대통령이 2009 년에 먼저 제안했다. 이 방안에 따라 미국은 4 년 내에 10 만 명의 미국 학생들을 중국에 보내 학습시킬 계획이다.

新闻 9: 《人民日报海外版》(2011 年 11 月 18 日 第 01 版)

본보 네이멍구 쓰즈왕치(四子王旗) 11 월 17 일 보도: 17 일 19 시 32 분, 텐궁 1 호 목표 비행선과 두 차례의 도킹을 원만하게 완성한 선저우(神舟) 8 호 우주선 귀환창이 내몽골 중부의 쓰즈왕치(四子王旗) 낙하장에 안전하게 착륙했다.

이 우주선은 11 월 1 일 5 시 58 분 주취안(酒泉) 위성발사센터에서 발사돼 16 일 13 시간 동안 1100 만 킬로미터 우주를 돌아 중국 우주비행사상 최장시간, 최장거리의 비행 기록을 남겼다.

선저우 8 호 귀환창이 안전하게 착륙한 후 중공중앙 국무원 중앙군사위원회는 텐궁

1 호와 선저우 8 호 도킹의 원만한 성공을 축하하며 축전을 보냈다.

新闻 10: 《人民日报海外版》(2011年11月10日 第01版)

본보 베이징 11 월 9 일 보도 : 경제협력개발기구(OECD)가 최근 발표한 ‘2011 년 과학, 기술과 산업 점수표: 지식경제 중의 혁신과 성장’ 보고에 의하면 2000 년부터 2010 년까지 글로벌 주요 국가의 특허가 1990 년부터 2000 년 기간에 비해 평균 20% 줄었다. 반대로 중국, 인도와 한국을 대표로 한 아시아 국가의 특허는 크게 증가했다. 지난 10 년 동안 중국은 정보 통신 기술 혁신의 3 위권에 진입했다.

보고에 의하면 미국, 독일과 일본의 발명가들이 제출한 특허와 사용률이 가장 높았으며, 아시아 국가인 중국, 인도와 한국의 특허의 질이 분명하게 높아지고 있다고 한다.

新闻 11:

지난달 22 일 뉴질랜드 크라이스트처치에서 발생한 강진으로 실종됐던 남매 가운데 오빠 유길환(24)씨의 시신이 발견됐다. 외교부는 4 일 “노광일 뉴질랜드 주재 한국 대사가 뉴질랜드 경찰로부터 실종된 오빠 유씨의 시신 발견 사실을 전달받았다”고 말했다. 유씨의 시신은 캔터베리 TV 건물 더미에서 발견됐다. 함께 실종된 유길환씨의 여동생 유나온(21)씨의 시신은 찾지 못했다. 유씨 남매는 지난달 22 일 낮 12 시 51 분쯤 발생한 6.2 규모의 강진으로 CTV 내 어학원에서 수업을 받다 빌딩이 무너져 내리면서 매몰됐다.

新闻 12:

소식에 따르면 연합군이 리비아에 연일 공습을 하고 있는 가운데 카다피 정권은 미국 등과 물밑 접촉을 통해 리비아 사태를 해결하려는 ‘출구전략’을 찾고 있는 것으로 알려졌다. 또 카다피 정권의 핵심인사들이 위기에 빠진 리비아를 구하기 위해 카다피를 배반하거나 망명을 모색하는 분위기도 감지됐다고 영국 일간지 ‘더 타임스’가 22 일 보도했다. 하지만 이 신문은 이와 연관된 리비아 고위 인사들이 누구지에 대해선 구체적으로 언급하지 않았다. 이날 AP 통신 등 외신들은 “카다피 측이 장기화되고 있는 리비아 사태를 해결하기 위해 미 고위 관리들과 협의하고 있다는 정황이 속속 드러나고 있다”며 “최근 카다피군에 억류됐던 ‘뉴욕 타임스’ 기자 4 명이 풀려난 것도 이 같은 외교적 노력의 결과”라고 보도했다.

新闻 13:

정부는 독도 영유권 기술이 강화된 일본의 중학교 교과서 검정 결과가 30 일께 발표되는 데 따라 28 일 대책 마련에 착수했다. 국무총리실과 외교통상부·교육과학기술부 등 관계부처는 이날 ‘독도영토관리대책단’ 회의를 열고 독도 헬기 이착륙장 보수공사 착수와 문제 교과서의 채택 확산 저지 방안 등을 검토했다. 정부 당국자는 “정부는 다양한 경로로 일본 정부에 신중한 대응을 촉구해왔다”며 “그럼에도 일본 측이 독도 영유권 기술을 강행할 경우 말이 아니라 행동으로 보여줄 것”이라고 말했다.

新闻 14:

일본 후쿠시마(福島) 원자력발전소에서 누출된 요오드·세슘 등 방사성 물질이 서울 등 한반도에서도 검출됐다. 한국원자력안전기술원은 “서울 등 국내 12 곳에서 지난 일주일 동안 채집한 공기 중의 먼지를 분석 중인데 몇 곳에서 방사성 요오드가 검출됐다”고 3 월 28 일 자정쯤 긴급 발표했다. 한국원자력안전기술원측은 그러나 검출된 방사성 요오드가 인체에는 영향을 미칠 수 없는 극미량이라고 밝히면서 정확한 분석

결과는 29 일 나을 예정이라고 말했다. 이와 관련, 한 원자력 전문가는 “방사성 요오드가 검출되면 세슘은 거의 함께 나온다”며 세슘의 한반도 확산도 기정사실화 했다.

新闻 15:

남북한은 3 월 29 일 백두산 화산 문제를 의제로 민간 전문가 회의를 열어 화산 관련 연구 및 대책 마련의 필요성에 대한 의견 접근을 봤다. 경기도 문산 도라산 남북출입사무소에서 열린 회의에서 북측 단장인 윤영근 화산연구소 부소장은 모두발언을 통해 기상이변 등을 언급한 뒤 “기상 현상도 잘 모르겠고 지진 또한 잘 모르는 일”이라며 백두산 화산 협의의 필요성을 거론했다. 윤 부소장은 “일본에서 지진이 있는 다음에 우리 백두산 지하수 관측공에서 물이 약 60 cm 출렁거리고 샘물에서 흙탕물이 나오는 현상이 많았다”고 전했다.

新闻 16:

한국과 중국 정부가 올해 안에 자유무역협정(FTA) 협상 개시를 선언하기로 합의했다. 협상 개시를 선언하면서 민감하지 않은 공산품에 대해 곧바로 무관세나 파격적으로 낮은 관세를 매기는 ‘조기수확(early harvest) 조항’이 적용된다.

베이징의 한 소식통은 “한·중 양국이 7 월 8 일 베이징에서 연내 FTA 개시를 선언하기로 최종합의하고 양해각서(MOU)를 체결할 예정”이라며 “이를 위해 이태호 외교부 통상교섭본부 FTA 국장이 이번 달 7 일 베이징에 도착했다”고 말했다. 그는 “8 일 열리는 중국 상무부와 의 국장급 회담에서 MOU 에 양국이 서명할 예정”이라고 전했다.

新闻 17:

연간 물가상승률 4% 달성이 사실상 물 건너 갔다. 10 월 4 일 통계청에 따르면 9 월 소비자물가는 지난해 같은 달보다 4.3% 올랐다. 8 월에 비해 상승폭을 1%포인트 줄였지만 9 개월 연속 4%대를 유지했다. 금반지·전셋값이 물가를 끌어올렸다.

여름철 크게 오른 채소·과실류 가격이 떨어진 덕에 지난달 신선식품 물가는 지난해 동월 대비 7.4% 내렸다. SK 텔레콤의 이동전화 통화료 1000 원 인하도 물가 상승세를 둔화시키는 데 일조했다. 기획재정부 최상목 경제정책국장은 “추석 이후 수요 감소와 기후여건 개선 등으로 채소·과실류 수요가 안정되면서 물가 상승세가 꺾이는 데 도움이 됐다”고 말했다.

新闻 18:

서울시는 시 비정규직 근로자 2800 명을 단계적으로 정규직으로 전환하기로 했다고 밝혔다. 서울시는 정규직으로 전환할 인력의 규모, 들어갈 예산, 일정 등에 관한 연구용역을 곧 발주할 예정이다. 비정규직 근로자의 정규직 전환은 박원순 서울시장의 공약이다.

이날 주용태 일자리정책과장은 “내년 초에 연구용역을 발주해 7 월께 결과가 나오면 부서 협의를 통해 최종안을 만들 계획”이라고 말했다. 서울시 본청 직원 1 만 6228 명 중 비정규직 근로자 1500 여 명과 서울메트로, SH 등 투자·출연기관과 사업소 비정규직 근로자 1300 명이 정규직 전환 검토 대상자다.

新闻 19:

무소속 박원순 서울시장 후보의 승리는 ‘정권심판론’과 ‘안철수 바람’이 빚어낸 합작품이다. 여기에 야권이 ‘단일후보’로 승부수를 던지면서 10 년 만에 서울시를 탈환할 수 있었다.

이명박 정권 임기 4 년 차를 맞아 정부·여당에 대한 민심이 크게 나빠진 상황에서

불거진 이명박 대통령의 ‘내곡동 사저’ 문제는 선거에 결정적인 영향을 미쳤다는 게 박 후보와 민주당 측 분석이다. 박 후보 캠프의 우상호 대변인은 26 일 “내곡동 사저 파문이 경제실정 등으로 열 받고 있는 국민들을 더욱 자극했다”며 “이는 정권과 한나라당에 대한 심판”이라고 말했다.

新闻 20:

국립환경과학원은 9 일 2004 년부터 진행 중인 국가 장기생태연구사업의 중간 조사 결과 발표를 통해 “한반도의 연평균 기온이 점차 상승하면서 국내 자연생태계에서도 변화 추세가 점차 뚜렷해지고 있다”고 밝혔다.

환경과학원은 특히 제주도와 남해안 일부에서만 나타나고 있는 아열대 기후가 2071 년 이후에는 백두대간의 일부 고산지대를 제외하고는 남한 전역으로 확대될 것으로 예측했다.

연구팀은 이 같은 온난화의 영향으로 지리산의 해발 400m 지점에 위치한 전남 구례군 토지면의 소나무 숲에서는 온대 수종인 소나무의 밀도가 줄어드는 대신 난대수종이 증가하고 있다고 지적했다.

汉语和韩语近体诗的呼吸节奏特征对比研究*

吴韩娜¹, 孔江平²

(1.2. 北京大学 中国语言文学系, 北京 100871)

文 摘: 本文从生理语音学的角度着手, 以汉语诗和与它内容一致的韩语诗为研究对象, 分别采集并分析了语音信号和呼吸信号, 从而探讨两种不同语言的近体诗呼吸节奏的特点。研究发现: (1) 汉语近体诗朗读中胸腹呼吸分工明确, 胸呼吸与生理信号对应, 腹呼吸与韵律单位对应, 而朗读韩语诗中, 其功能比较模糊; (2) 不同语言的呼吸表现和呼吸重置幅度和时长均有关系; (3) 汉语近体诗朗读中呼吸重置幅度和时长有显著性相关, 而韩语呼吸重置幅度和时长没有相关。

关键词: 语音生理学; 近体诗格律; 呼吸节奏; 韵律; 韩语

中图分类号:

由于在语言中的重要地位, 韵律特征研究受到语言学、心理学、声学以及计算机科学、人工智能等诸多领域的重视。语言教学领域也一直特别关注韵律特征研究。由于汉语有声调, 汉语和其他语言进行对比时, 韵律成分可以说比音段成分更重要, 所谓的“洋腔洋调”在很大程度上也取决于声调和比声调更高的语音层次, 即韵律的掌握程度(林涛, 1996)。

近十几年以来, 汉韩语言对比研究日益增加, 而语篇节奏对比研究很少见。由于语流的韵律构造是一个复杂的过程, 它受到生理、心理、语法等各个方面的制约, 因此研究者的角度也各不相同。而目前的研究大多集中在音系学、声学和心理学方面, 从生理角度进行研究的并不多。

呼吸运动为言语提供动力, 满足言语交际的需求, 尤其在语流中的停顿与呼吸运动有直接的关系, 与此同时, 言语也影响着呼吸的形式和特点, 因此对言语呼吸进行探索非常具有意义。

在生理方面, 根据 Liberman (1967) 曾提出的“呼吸群(breath-group)”理论, 在国内研究中, 谭晶晶(2008)利用机电电脑仪和呼吸带传感器记录了呼吸信号, 发现呼吸结构可以反映人们对表达内容的整体认知规划和生理机能的制约。其中生理机制主要指的是呼吸调节, 呼吸是划分韵律层级的重要线索(Tseng C. et al, 2005)。

众所周知, 汉语和韩语是不同语系的两种语言, 包括语音和句法结构等语言特征差别很大, 汉语诗歌主要采用音位律和音声律, 相比, 韩语

诗歌的遵守音步律。

由于两种语言具有不同语言特征, 如果对语篇韵律特征进行对比研究的话, 不仅把汉语和韩语两种语言的韵律特征更加明确, 而且还能够从多角度显示语言韵律结构特征的共性和对韵律特征影响的种种因素。

鉴于以上几点, 本文打算从呼吸节奏这一新的角度着手, 以对比研究的方法, 探讨汉韩两种语言的近体诗呼吸节奏的特点, 并试图寻找呼吸方式和声学特征怎么对应、呼吸节奏和韵律结构的关系以及不同句法结构对呼吸节奏的影响。

1 实验方法

1.1 发音人

本实验汉语和韩语母语者各选两名女声, 共选择了四名年龄范围为 25-35 岁的发音人, 她们语音纯正, 分别汉语普通话和首尔话标准, 有丰富的播音或朗读经验, 在朗读速度以及调整情感方面上也非常稳定, 便于除去韵律的相对性和可变性特征。

1.2 实验材料

本研究选取了五言绝句、七言绝句、五言律诗和七言律诗各五首总共20首, 由于近体诗的格律严格, 结构简单, 即和其他文体相比, 近体诗的呼吸信号类型比较简单。

录音材料为内容一致, 语言和形式不同的三种

*基金项目: 中国哲学社会科学重大投标项目 (批准号: 10&ZD125); 国家自然科学基金面上项目 (批准号: 61073085)。

作者简介: 吴韩娜 (1979-), 女 (韩), 韩国首尔, 博士生。

通讯联系人: 孔江平, 北京大学中文系, 教授, jpkong@pku.edu.cn。

文本：1) 汉语近体诗版；2) 韩语汉字音版；3) 韩语翻译版。由汉语母语者来朗读汉语近体诗版，韩语母语者则朗读韩语汉字音版和韩语翻译版两种文本。

在韩国所使用的汉字借字标记法中，在汉语诗歌采用的是“口诀”。它是指朗读汉文时，为了了解意思或便于诵读，而在每个句子下方记录的语法要素的统称，可以写为诸如“隐(은,는)”，“伊(이)”之类的汉字(黄京洙，2009)。

汉文和汉字在韩语体系和文化中占有特殊的地位，迄今为止韩国学生在课堂上仍然以口诀的方式来第一次尝试阅读汉语的诗歌。因此，本文为了将不同语言结构和呼吸韵律之间的关系加以明确，而添加了汉字版的语料进行对比。

1.3 录音和数据处理

每个发音人在北京大学中文系语音乐律实验室单独进行录音。通过 PowerLab 采集器和 Chart 软件采集为*.adicht 格式的文件，然后转换为*.wav 格式的语音文件，最后读入到言语呼吸韵律分析系统中，对韵律层级、胸腹呼吸重置进行标记，提取韵律层级参数、胸腹呼吸参数，以 SPSS 进行统计分析(杨峰，2012)。

2 实验结果

2.1 基本呼吸重置特征对比

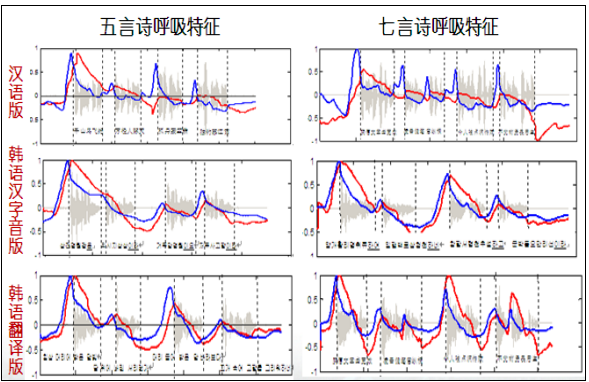


图 1 近体诗朗读中胸、腹呼吸与语音波形

图 1 显示的是三种不同形式的近体诗呼吸重置特征。朗读近体诗中胸、腹呼吸基本分成二级，大小呼吸交替出现，每联的第一句一定会出现大呼吸，在大多数情况下，第二句和第四句开始处具有小呼吸，第二联开始处有时具有大呼吸，有时会出现小呼吸。与之前的呼吸研究结果最显著的不同点是，不管是哪种语言的呼吸曲线，韵律短语边界处不仅会出现抖动，还会出现明显的小

呼吸，尤其在七言诗中更加明显。

从功能上看，腹呼吸比胸呼吸早，而比胸呼吸结束得快。腹呼吸先给言语活动提供较大的能量，另外腹呼吸也会出现在韵律短语边界处，此时胸呼吸会持平，可以说腹呼吸对韵律单位边界比较敏感。相比，胸呼吸比腹呼吸来得晚，而语音信号结束点和胸呼吸呼气结束点比较一致，可见胸呼吸与言语活动的开始、保持和结束更有关系。

从图 1 所示，不管是哪种语言，五言还是七言诗，呼吸重置位置以及其曲线大体相似，这说明朗读者按照严格的近体诗格律来安排呼吸方式，体现近体诗的情感以及近体诗格律带来的节奏感，该特征具有一定的普遍性。

在不同版本之间最显著的不同点是，汉语七言近体诗朗读中腹呼吸存在三级呼吸重置，一级呼吸处于每首诗的每联开始位置，二级呼吸处于每长句的开始位置，三级呼吸处于长句末尾小句。每长句对应一个胸呼吸重置，韵律边界处出现呼吸中段抖动。相比，朗读韩语翻译版七言近体诗中胸呼吸存在三级呼吸重置，一级呼吸处于每首诗的每联开始位置，二级呼吸处于每长句的开始位置，三级呼吸处于长句末尾小句。每长句对应一个胸呼吸重置，韵律边界处出现呼吸中段抖动。

2.2 汉语版近体诗呼吸重置特征

在本次实验中，为了探讨大小不同的呼吸重置幅度的频度分布，我们提取 20 首汉语近体诗中所有的呼吸重置幅度，在 SPSS 中将他们画出直方图并进行聚类分析，如下图所示。

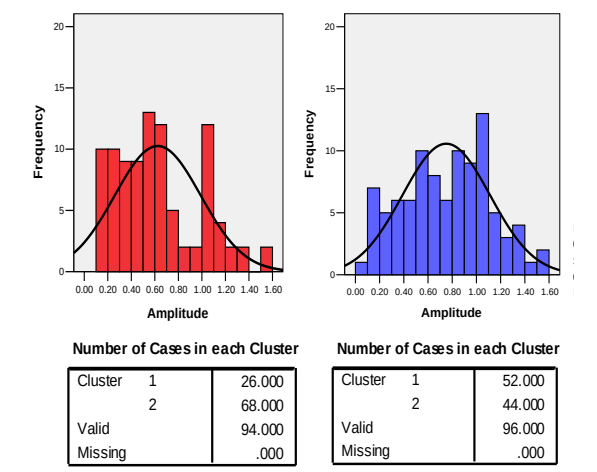


图 2 汉语五言近体诗呼吸重置分布图和聚类分析结果(左图：胸呼吸，右图：腹呼吸)

汉语五言近体诗的胸腹呼吸重置频度分布基本符合正态分布，胸呼吸重置幅度的平均值为 0.62，腹呼吸平均值为 0.75，腹呼吸重置幅度的

平均值比胸呼吸的稍微大一些。胸腹呼吸重置次数分别为胸呼吸 94 次(每首律诗平均呼吸次数为 6.5 次、绝句平均为 3 次)、腹呼吸 96 次(每首律诗平均为 6.4 次、绝句平均为 3.3 次)，就是说汉语五言近体诗的每诗行基本出现一次呼吸，胸腹呼吸出现的次数基本一致。

从上图 2.看，无论胸呼吸还是腹呼吸重置幅度的分布在 1.0-1.1 和 0.5-0.6 处分别出现了两个比较明显的峰值，这说明汉语五言近体诗的呼吸重置大致可以分成两级。

胸腹呼吸都存在二级呼吸重置，腹呼吸一级呼吸基本上处于每首诗的每联开始位置，二级呼吸处于每长句的开始位置，因此，二者的呼吸重置比较有规律地交替出现，其频度也比较接近。与此相比，胸呼吸一级呼吸处于每首诗的开始位置，不一定每联都出现，经常由二级呼吸重置替代。结果，二级呼吸重置次数比一级呼吸多。可见，朗读汉语五言近体诗中腹呼吸重置更有规律，胸呼吸重置方式除了每联开始部分之外，比较自由地调节呼吸重置幅度。

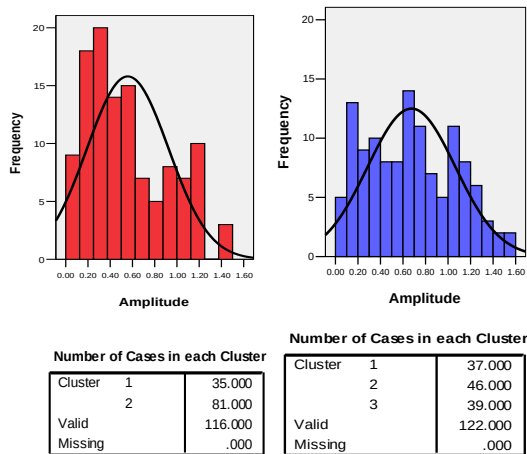


图 3 汉语七言近体诗呼吸重置分布图和聚类分析结果（左图：胸呼吸，右图：腹呼吸）

与五言近体诗相比，朗读汉语七言近体诗的时候，胸腹呼吸的重置幅度都会比五言诗的小，而重置次数会变多。如图 3 所示，胸呼吸重置幅度的平均值从 0.62 调到 0.56，腹呼吸平均值从 0.75 调到 0.67，而胸腹呼吸重置次数分别为胸呼吸 94 次变成 116 次、腹呼吸 96 次变成 122 次，无论是重置幅度还是次数，胸腹呼吸的变化程度比较一致，胸腹呼吸的幅度都降低了 10%，相反呼吸次数两个都增加了 25%，很有规律，无个人差异。

从上图我们比较容易发现，七言近体诗的胸腹呼吸分布与五言近体诗有两个明显不同的地方是，在朗读七言近体诗中，腹呼吸重置幅度会分

成三个级别，在 0.2 和 0.6-0.8 和 1.0-1.3 处分别出现了三个比较明显的峰值，该特征没有个人差异。

2.3 韩语汉字音版近体诗呼吸重置特征

汉字版韵律结构比较特殊，因为韩语的音步由 3-6 个音节组成，朗读汉字版五言近体诗中，安排的呼吸幅度比汉字版的低，也足以完成一小句，而体现出不同语言结构的不同节奏特征。

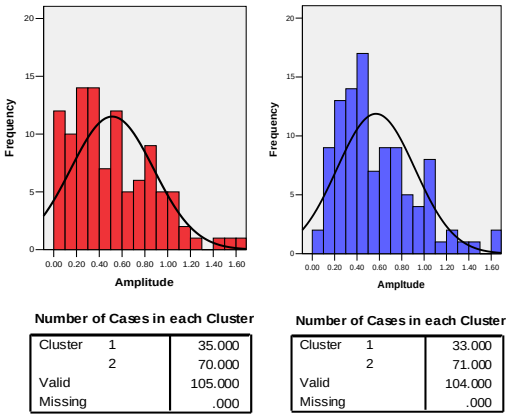


图 4 韩语汉字音五言近体诗呼吸重置分布图和聚类分析结果（左图：胸呼吸，右图：腹呼吸）

从上图 4 看出，韩语汉字音版近体诗朗读中胸腹呼吸重置明显分为两级，胸腹呼吸重置的平均值分别为胸呼吸 0.52、腹呼吸 0.57，腹呼吸的重置幅度比胸呼吸的稍微高一些，这些特征与汉语五言近体诗朗读中的呼吸信号特征比较一致。

韩语汉字版五言近体诗朗读中胸呼吸重置聚类分析结果显示重置次数分别为 35 次和 70 次，可见无论是胸呼吸还是腹呼吸，二级呼吸次数比一级呼吸的多一倍。如图 3 直方图给出的分布所示，幅度集中在了 1.0-1.1 上，比较稳定。与此相比，二级呼吸的次数比一级呼吸的两倍以上。腹呼吸呼吸重置分布和胸呼吸的情况比较相似，二者的呼吸次数也基本一致。

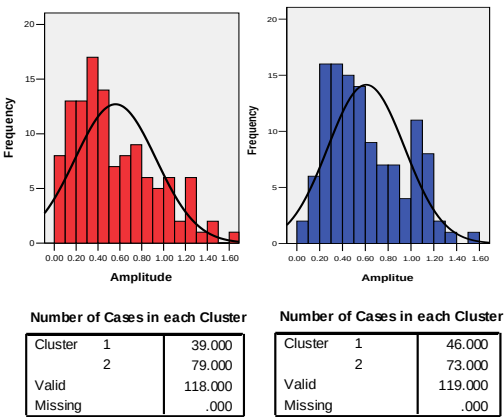


图 5 韩语汉字音七言近体诗呼吸重置分布图和聚类分析结果（左图：胸呼吸，右图：腹呼吸）

汉语七言近体诗的胸、腹呼吸重置频度分布曲线高峰呈偏左，因为胸呼吸幅度的平均值比腹呼吸的低，分别为 0.56 和 0.61。如图 5. 所示，朗读韩语汉字音版七言近体诗中胸、腹呼吸重置次数分别为 118 次和 119 次，每首律诗平均呼吸次数呼吸为 7.7 次、绝句平均为 4.1，就是说七言近体诗的每诗行基本出现一次呼吸，胸、腹呼吸出现的次数基本一致。

与韩语汉字音版五言近体诗朗读中的呼吸特征相比，朗读韩语汉字音版七言近体诗中胸、腹呼吸重置幅度会变高，重置次数也会变多，腹呼吸平均幅度比胸呼吸的高。无论是重置幅度还是次数，胸、腹呼吸的变化方式与程度比较一致，非常有规律，这是汉语和韩语汉字音版呼吸方式的相同点。

2.4 韩语翻译版近体诗呼吸重置特征

本次实验总共 20 首汉语近体诗的韩语翻译版诗歌，虽然韩语翻译版诗歌也是每首由两联四句或四联八句组成，而韩语翻译版五言和七言近体诗每句的平均音节数分别为 11 个和 14 个（汉语五言和七言近体诗音节数分别为 5 个和 7 个）。可见与汉语近体诗相比，韩语翻译版近体诗的音节数远远超过汉语近体诗的音节数，近似达到汉语的两倍以上。从韵律结构来看，韩语翻译版近体诗五言诗每句由 2-6 个音步，七言诗由 3-6 个音步来组成一小句（汉语分别为 2 个和 3 个），显然汉韩两种版本的差异比较明显。那不同韵律结构如何影响到生理呼吸的安排，如何显示不同语言的特征？

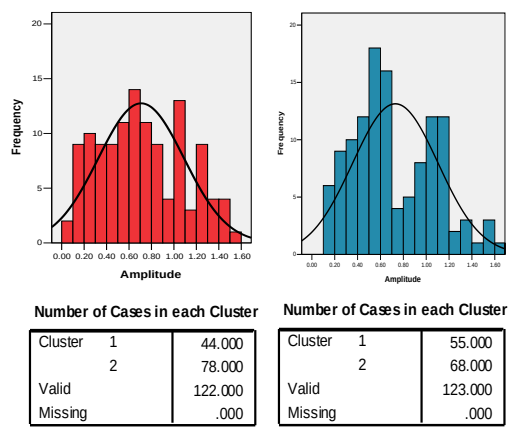


图 6 韩语翻译版五言近体诗呼吸重置分布图和聚类分析（左图：胸呼吸，右图：腹呼吸）

上图 6 所示，韩语翻译版五言近体诗的胸、腹呼吸重置频度分布基本符合正态分布，腹呼吸重置幅度的平均值比胸呼吸的高一些，分别为 0.70 和 0.73。胸、腹呼吸重置分别出现了 122 处（每首

律诗平均呼吸次数为 8 次、绝句平均为 4.2 次）和 123 处（每首律诗平均为 8.1 次、绝句平均为 4.2 次），等于韩语翻译版五言近体诗朗读中每诗行基本出现一次呼吸，如上文给出的呼吸曲线所示，胸、腹呼吸出现的次数基本一致。

从具体数据分析显示，三种不同版本中，韩语翻译版平均幅度最高，音节数与呼吸次数也最多，而严格遵守汉语近体诗格律，呼吸安排颇有规律。这说明，汉语近体诗格律的严格性制约不同结构语言的节奏特征。

另外，有趣的是，如下图 7 所示呼吸信号分析结果表明，韵律短语边界处也会出现小呼吸，与声学分析中的停顿时长无关，值得进一步研究。

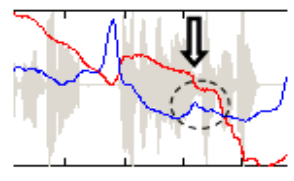


图 7 汉语近体诗朗读中的小呼吸例外

上文已提到，由于语言结构不同，在诗歌翻译时，两个汉字基本对应一两个音步。韩语翻译版五言近体诗呼吸特征分析显示，不管每一个小句包含几个音步，韩语朗读者会尽量遵守汉语近体诗格律，表现出颇有规律的呼吸模式。那由此再加一两个音步，等于加了一个韩语诗歌中的一个韵律句的话，其呼吸方式会有什么变化？

从图 8 给出的聚类分析中可以看出，和汉语版最大的不同点是，朗读韩语翻译版七言近体诗中胸呼吸重置大体分成三级，腹呼吸重置分成二级，二中不同语言的胸腹呼吸安排方式完全相反。

如图 8 所示，就像韩语翻译版五言近体诗呼吸重置分析结果一样，腹呼吸平均幅度还是高于胸呼吸，从五言到七言，胸、腹呼吸次数都增加的很多，随之平均幅度也会变低了些，胸呼吸从 0.70 降到 0.69，基本一致，而腹呼吸从 0.73 降到 0.70，比胸呼吸的变化大一些。

而胸腹呼吸重置次数分别为胸呼吸 122 次变成 145 次（每首律诗平均呼吸次数为 9.5 次、绝句平均为 5.1 次）、腹呼吸 123 次变成 146 次（每首律诗平均为 9.4 次、绝句平均为 5.2 次），基本上比每首五言近体诗中出现的呼吸重置次数要多了一次。可以说，汉语和韩语版呼吸次数变化趋势是相似的，只是韩语近体诗朗读中，文体或者不同结构的变化对呼吸幅度和次数变化程度要比汉语近体诗低一些。尤其对韩语翻译版胸呼吸重置平均幅度几乎没有影响。

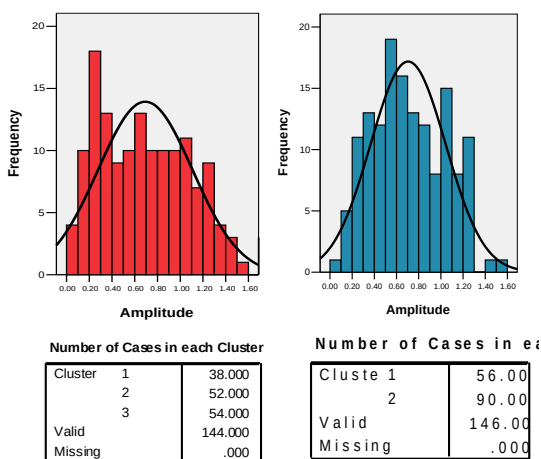


图 8 韩语翻译版七言近体诗呼吸重置分布图和聚类分析（左图：胸呼吸，右图：腹呼吸）

由此可见，朗读韩语近体诗中，增加了音节而韵律结构出现变化时，根据不同结构，胸呼吸会安排不同大小的呼吸幅度体现近体诗的节奏，结果胸呼吸就分成三级呼吸。与此相比，随着小呼吸的增加，腹呼吸平均幅度也会变低，而大小呼吸的固定比例不会改变。

2.5 呼吸重置幅度与时长之间的关系

我们运用 SPSS 对呼吸重置时长和重置幅度值作了相关分析，得到如下结果其显著性水平为 0.01。从呼吸幅度与时长的相关关系上，汉语和韩语之间存在显著性差异：汉语版近体诗和汉字版近体诗朗读中，呼吸幅度与时长之间的相关分析结果比较一致，即五言近体诗朗读中，幅度与时长之间成正比，存在显著性相关。相比，韩语版近体诗朗读中，不管是五言还是七言近体诗，不同级别的呼吸时长之间差别不明显，朗读者按照固定的呼吸模式，调整整个呼吸安排体现汉语近体诗格律，结果相关分析显示，幅度与时长之间不存在显著性相关。

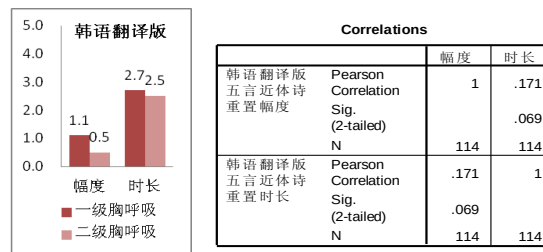
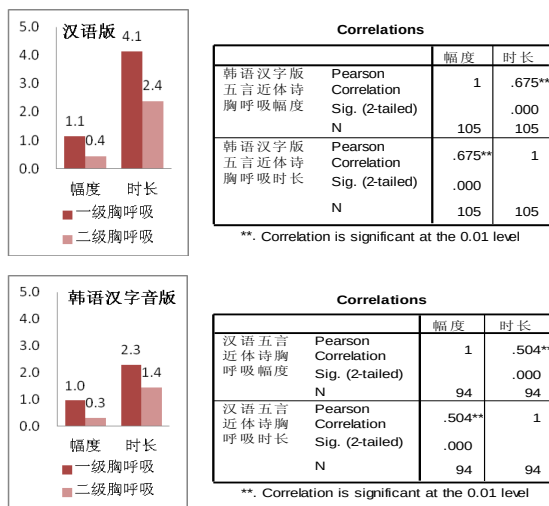


图9 汉语与韩语近体诗朗读中呼吸重置幅度与时长相关分析

可见，此结果既可以显示汉语近体诗格律的特征，也可以表现韩语音节结构的特征“等时性”。从整体呼吸安排来说，汉语近体诗格律优先于不同语言节奏特征，不管是哪种语言，五言还是七言，朗读者按照严格的近体诗格律来安排呼吸方式，体现近体诗的情感以及近体诗格律带来的节奏感，该特征具有一定的普遍性。这还说明，呼吸安排决定于格律和朗读者对诗歌内容的不同解释而产生的情感，基本上不在于生理上的需求。

3 结论

本章基于呼吸信号主要从呼吸节奏与韵律表现的角度考查具有不同语言结构的汉语和韩语近体诗朗读中的呼吸节奏特征的不同，并试图探讨呼吸节奏和韵律单位的对应以及不同语言结构对呼吸节奏的影响。

通过本次的分析，我们发现从宏观上看，因为汉语近体诗的格律严格，形式固定，从而不管哪种语言哪种形式，朗读近体诗中整体呼吸格局有一定的一致性：朗读近体诗中胸、腹呼吸基本分成二级，大小呼吸交替出现，每联的第一句一定会出现大呼吸，在大多数情况下，第二句和第四句开始处具有小呼吸，第二联开始处有时具有大呼吸，有时会出现小呼吸。与之前的呼吸研究结果最显著的不同点是，不管是哪种语言的呼吸曲线，韵律短语边界处不仅会出现抖动，还会出现明显的小呼吸。有趣的是，这些小呼吸的出现的位置往往与声学分析中的停顿的位置和时长无关，有待进一步考查。

从胸、腹呼吸配合的角度看，汉语版和韩语版朗读中腹呼吸起始点都早于胸呼吸，重置幅度高于胸呼吸，在开始朗读时给整个言语活动提供能量，有辅助胸呼吸的作用，而提前结束并准备下一次呼吸和言语活动；一开始朗读，胸呼吸晚于腹呼吸，胸呼气开始点与终点与言语信号时间点比较一致。可以说，这些特征具有普遍性。

不同语言版本之间主要差别在于：朗读汉语

近体诗时，胸、腹呼吸分工更加明确，就是说在开始朗读时给整个言语活动提供能量，比胸呼吸早点结束而朗读中对韵律单位的界限更敏感，韵律短语边界处常出现腹呼吸，从而呼吸次数比胸呼吸多一些；呼气之后保持一定时间使言语活动持续下去，因此胸呼吸的终点和言语信号终点基本吻合，胸呼吸与整个言语活动的完成有关。相比，朗读韩语版近体诗中，韩语版近体诗朗读中胸、腹呼吸下降趋势比较一致，二者的分工比汉语版呼吸特征模糊。

从五言近体诗到七言近体诗，汉语版和韩语版呼吸变化趋势也截然不同：汉语版五言近体诗朗读中胸、腹呼吸都存在二级呼吸，而七言近体诗朗读中，因小呼吸次数增加而平均幅度偏低，腹呼吸具有三级呼吸；汉字版五言近体诗朗读中的呼吸特征与汉语版的比较一致，而七言近体诗朗读中，增加的音节数对韩语朗读者对呼吸安排的影响不够大，结果呼吸次数增加，幅度也增加，而并没有出现显著变化；韩语翻译版近体诗的音步数已远远超出韩语诗歌本身结构中的音步数和音节数，但朗读五言近体诗中，胸、腹呼吸的格局颇有规律，均分成二级呼吸。韩语版七言诗呼吸重置特征像汉语版的变化一样，随着小呼吸次数的增加，平均呼吸幅度也降低了一些。与汉语版呼吸变化不同的是，韩语翻译版胸呼吸重置大小不一而变成三级呼吸，腹呼吸比较稳定。

由此可见，汉语近体诗格律的严格性制约不同结构语言的节奏特征。虽然呼吸配合方式不同，但韩语朗读者调整呼吸重置幅度与次数以及朗读速度而体现汉语近体诗格律特征。

从呼吸幅度与时长的相关关系上，汉语和韩语之间存在显著性差异：汉语版近体诗和汉字版近体诗朗读中，呼吸幅度与时长之间的相关分析结果比较一致，即五言近体诗朗读中，幅度与时长之间成正比，存在显著性相关。与此相比，韩语版近体诗朗读中，不管是五言还是七言近体诗，不同级别的呼吸时长之间差别不明显，朗读者按照固定的呼吸模式，调整整个呼吸安排体现汉语近体诗格律，结果相关分析显示，幅度与时长之间不存在显著性相关。

参 考 文 献

- [1] 林涛. 语音研究和对外汉语教学[J]. 世界汉语教学, 1996, 3: 18-21.
LIN tao. Phonetic study and Chinese teaching as a second language[J]. J Chinese Teaching in the World, 1996, 3: 18-21. (in Chinese)
- [2] Liberman P. Intonation, Perception, and Language [M]. Cambridge. Massachusettes. The MIT Press. 1967
- [3] 谭晶晶. 汉语普通话朗读时的呼吸节奏研究[D]. 北京: 北京大学, 2008.
TAN jingjing. A study on breathing-reset in different types of literature-reciting [D]. Beijing: Peking University, 2008. (in Chinese)
- [4] Tseng C, Pin S, Lee Y, et al. Fluent speech prosody: Framework and modeling[J]. Speech Communication. 2005, 46: 284-309
- [5] 黄京洙. 韩国语语言学概论[M]. 辽宁民族出版社, 2009
HWANG kyoung joo. Introduction of Korean linguistics[M]. Liaoning Minzu Press. 2009
- [6] 杨峰. 中国传统吟诵研究-从节奏、嗓音和呼吸[D]. 北京: 北京大学, 2012.
YANG feng. Studies on Chinese Traditional Chanting – from perspectives of prosody, phonation and respiration[D]. Beijing: Peking University, 2012. (in Chinese).

A Comparative Study on Breathing Rhythm Properties of Poem in Chinese and Korean

OH Hanna¹, KONG Jiangping²

1. 2. Department of Chinese Language and Literature, Peking University, Beijing 100871, China

Abstract: This study is a preliminary study on breathing rhythm properties of poetry in different languages of different prosodic construction: Chinese and Korean. Duration and amplitude of breathing-reset were detected and results showed that (1) The prosody and breathing rhythm varies according to the types of languages but it is very similar in metric poetry with strict metrics (2) In Chinese, the duration and amplitude of breathing-reset are highly correlated. But In Korean, there is no correlation. (3) The duration, amplitude and way of breath are all useful cues in differences languages and the hierarchies of breathing-resets. (4) The small breathing in the middle of poem is not related to physiological requirement in poem.

Key words: phonetic physiology; metric poem; breathing rhythm; prosody; Korean