语音乐律研究报告

Status Report of Phonetic and Music Research

2017

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. Kong, Jiangping, and Ruifeng Zhang. "VAT of the lexical tones in Mandarin Chinese." Journal of Chinese Linguistics 45.2 (2017): 275-289.……3
2. 杨锋，孔江平.汉语普通话不同文体朗读时的胸腹呼吸特性,《清华大学学报（自然科学版)》,第57卷,第2期,,176-181页……18
3. 汪高武, 党建武, 孔江平，2017，基于磁共振成像的汉语普通话舌尖调音建模，清华大学学报（自然科学版)，第57卷，第2期，158-163页……24
4. 董理，孔江平.昆曲闺门旦颤音的嗓音特征,《清华大学学报（自然科学版)》，第57卷，第6期，625-630页……30
5. 姚云，吴西愉，孔江平（通讯作者）.矢量半径驱动的汉语普通话立体声道模型,《清华大学学报（自然科学版)》，第57卷，第9期，945-951页……36
6. 刘文，杨正辉，孔江平.2017，新寨苗语单字调及双字调的声学实验研究,《民族语文》，第2期，第12-24页……43
7. 刘文、张锐锋.鱼粮苗语低平调和低降调的声学感知研究,《语言学论丛》，第五十四辑 …… 56
8. 李英浩，孔江平，2017，语速对普通话音段产生的影响，清华大学学报（自然科学版)，第57卷，第9期，963-969页……72
9. 孔江平.藏语声调起源的多维视角…… 79

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

VAT OF THE LEXICAL TONES IN MANDARIN CHINESE

Kong Jiangping Zhang Ruifeng
Peking University

ABSTRACT

The purpose of this research was to investigate the association of vocal attack time (VAT) and tones in speakers of Mandarin Chinese, and to explore how tones initiated at different pitch levels affected VAT. SP and EGG signals were synchronously recorded from 72 young undergraduates or postgraduates (42 females and 30 males) while they were reading aloud a wordlist of 50 disyllabic words at their most comfortable pitch, loudness and rate. VAT measures revealed three findings. (1) Vocal attack time shows no significant difference between the common *yangping* and the *yangping* derived from *shangsheng*. This, from a physiological perspective, supports the argument that the tone sequence 3-3 in Mandarin is indeed converted into 2-3, nothing else. (2) The tones of Mandarin Chinese that start from low pitch levels (35, 21) tend to present significantly different VAT values from those that start from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This embodies the nonlinear contra-variant relationship between VAT and F0 at vowel onsets. (3) There are deviations or individual differences: a small number of people do not follow this pattern.

Acknowledgment This research was funded by the National Social Sciences Foundation of China and the grant number was 10&ZD125. The authors also thank all the experts who reviewed the article.

Authors claim no conflict of interests to publish the paper in *Journal of Chinese Linguistics*.

Kong Jiangping (author for correspondance) [jpkong@pku.edu.cn] is a professor at Department of Chinese Language and Literature and Center for Chinese Linguistics, Peking University, No. 5 Yiheyuan Road, Haidian District，Beijing 100871， P.R. China. He is also affiliated to Joint Research Center for Language and Human Complexity, Chinese University of Hong Kong. He designed this research and guided Zhang Ruifeng in recording and data analyses.

Zhang Ruifeng is PhD candidate at Department of Chinese Language and Literature, Peking University. Under the help of Prof. Kong, he analyzed the data and wrote the article in English.

KEYWORDS
Vocal attack time **L**exical tones **P**honation onset **N**onlinear contra-variant relationship

1. INTRODUCTION

1.1 Vocal Attack Time

Vocal attack time (VAT) is a concept proposed by Baken and Orlikoff (1998, 1999) based on the time delay between the rise of the sound pressure (SP) and the appearance of an evident electroglottographic (EGG) signal, when SP and EGG signals are recorded simultaneously. In the presence of a trans-glottal airflow, the vocal folds oscillate with small amplitudes before they arrive at the midline of the glottis. On their arriving at the glottal midline with periodic contact achieved and stabilized, the amplitude of their oscillations grows very quickly. Therefore, the SP signal begins its growth to large magnitude well before the vocal-fold contact occurs. However, the EGG signal, as a record of vocal-fold contact area, has nearly no amplitude until the vocal-fold contact is achieved, and only after that does its amplitude grow rapidly. The EGG and SP signals are thus offset with respect to each other, and VAT is taken to be the time lag between the rise of them measured at the onset of phonation. Positive VAT values indicate that the initiation of SP signals leads that of EGG signals while negative ones signify the latter preceding the former. When the two sorts of signals rise at the same point of time, VAT equals zero. So VAT provides a potentially useful measure that varies with vocal attack characteristics. Orlikoff et al. (2009) for example, have reported negative VATs for all attempts of their subjects to produce a hard glottal attack. A computer program was developed to automatically extract VAT measures from the EGG and SP signals simultaneously recorded, and the validity of this measurement was experimentally demonstrated by Orlikoff et al. (2009). In 2012, Roark Watson, and Baken proposed a figure of merit (FOM) for VAT measurement，which was actually Pearson's correlation coefficient determined from the amplitude features of SP and EGG. VAT measurement has been used for nonlinguistic research by Roark, Watson,

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

Baken, Brown, and Thomas (2012) to acquire normative data of VAT in healthy young adults. In 2012, VAT was also measured for linguistically constrained voice onsets during the production of the six Cantonese tones (MA et al. 2012).

1.2 Lexical Tones in Mandarin Chinese

As a well-known tone language in Asia, Mandarin Chinese has four distinctive lexical tones: The first one, named *yinping*, is a high level tone with pitch sustained high on pitch level 5; the second, named *yangping*, is a mid-rise with pitch climbing from level 3 up to level 5; the third, *shangsheng*, is a fall-rise that dips first from levels 2 to 1 and then rises to level 4; the last one, *qusheng*, is a full fall that starts from level 5 and glides all the way down to level 1; the values of these tones are consequently recorded as *yinping*(55), *yangping*(35), *shangsheng*(214), *qusheng*(51). Because these lexical tones distinguish meanings in Mandarin Chinese, the same morpheme may have different meanings when adopting different pitch contours, for example, /mi/ with a fall-rise (214) signifies 'rice' in English, but means 'honey' when its pitch contour is altered to a full fall (51).

Unlike in English where phonemic variation occurs frequently, in Mandarin Chinese there are often occurrences of Tone Sandhi. The fall-rise pitch contour of *shangsheng* (214) mentioned above is only seen on syllables before pauses or in citation form. However, when two such contours are juxtaposed in speech flow, the first of them is always definitely altered into a mid-rise (35), the pitch contour that *yangping* always adopts. And furthermore, in flowing speech, the fall-rise of *shangsheng* preceding *yinping*, *yangping* or *qusheng* is nearly unexceptionally modified into a low-fall (21), with pitch dipping slightly from level 2 to 1. All these have made the original contour of Tone 3 (214) very seldom heard in connected speech. There are also occasions when Tone Sandhi is optional. The original tone of *yi* 一，a Chinese word that means "one" in English, is 55, a high level pitch pattern, when it is in citation form or at the end of a sentence. But in flowing speech, this pattern can be modified into 35, when it precedes morphemes of *qusheng* (e.g., *yiyang* 一样 55+51→35+51), or into 51, when it goes before *yinping*,

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

yangping or *shangsheng* (e.g., *yiban* 一般 55+55→51+55, *yizhi* 一直 55+35→51+35, *yiqi* 一起 55+214→51+214). But not all Mandarin speakers follow suit, and there are people who still pronounce *yiban* 一般 as 55+55.

1.3 Purpose

The aforementioned linguistic features of Mandarin tones have rendered the most frequent contours of tones in flowing speech as four types: *yinping* (55), *yangping* (35), *shangsheng* (21) and *qusheng* (51). By comparing the pitch levels from which they start, it is possible to distinguish the four types as two categories: The first and fourth tones both start with the highest pitch and go to one category, while the second and third tones that start from low levels 3 and 2 go to the other. As is well-known, pitch is a very important perceptual correlate of F0, which is associated with the rate of vocal-fold oscillations. Since the tones with a high-pitch onset have a higher rate of vocal-fold oscillation than those with a low-pitch onset during the initial stage of phonation, they may adopt different mechanisms of laryngeal adjustment, and present dissimilar characteristics of vocal attack. The purpose of the present investigation is to examine the association of VAT and tone in speakers of Mandarin Chinese, and to explore how tones initiated at different pitch levels affect vocal attack time. This is an attempt to measure VAT for linguistically constrained voice onsets.

2 METHOD

2.1 Wordlist

Three considerations decided which disyllabic words to choose for the present study. (1) The first syllable of the word should start with a head vowel (Chao 1970, 18-25), namely, no initial consonant or semivowel medial should stand at the beginning, because the computer program designed for VAT extraction only works efficiently on syllables beginning with vowels, and a large part of a Chinese tone contour is spread on the head vowel of a syllable. (2) The three vertex vowels of Mandarin Chinese (/a/ /i/ /u/) should be chosen as the head vowels of the first syllables of words, because they occupy the utmost points on the vowel chart and represent the entire scope of tongue movement during

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

speech. (3) Each final of the first syllable should adopt four distinctive tone patterns (*yinping*, *yangping*, *shangsheng* and *qusheng*), with each pattern being further followed by *yinping*, *yangping*, *shangsheng* and *qusheng* respectively as the second syllables of words. Since VAT measurement would only be taken for the voice onset of the first syllables, there was no requirement on the composition of the second syllables. All these resulted in a wordlist of 50 double-syllable words with /ai/ /an/ /aŋ/ /au/ /i/ /u/ chosen to be the finals of the first syllables, as is seen Table 1.

Table 1 50 disyllabic words used in the present study

vowel of syllable1	tone combination	vowel of syllable1	tone combination	vowel of syllable1	tone combination	vowel of syllable1	tone combination	vowel of syllable1	tone combination
/ai/	ai ge 哀歌 55+55	/ai/	ai xiao 矮小 214+214	/u/	wu ji 屋脊 55+214	/u/	wu cha 误差 51+55	/i/	yi dian 疑点 35+214
	ai qiu 哀求 55+35		ai qi 嗳气 214+51		wu gou 污垢 55+51		wu shi 务实 51+35		yi han 遗憾 35+51
	ai ku 哀苦 55+214		ai xi 爱惜 51+55		wu bian 无边 35+55		wu chan 物产 51+214		yi jing 已经 214+55
	ai dao 哀悼 55+51		ai qing 爱情 51+35		wu chang 无常 35+35		wu bi 务必 51+51		yi wei 以为 214+35
	ai pi 挨批 35+55		ai shi 碍事 51+51		wu chi 无耻 35+214	/i/	yi ban 一般 55+55		yi mian 以免 214+214
	ai ai 皑皑 35+35	/an/	an pai 安排 55+35		wu gu 无故 35+51		yi cong 依从 55+35		yi hou 以后 214+51
	ai zai 挨宰 35+214	/aŋ/	ang zang 肮脏 55+55		wu can 午餐 214+55		yi fa 依法 55+214		yi xiang 异乡 51+55
	ai zheng 癌症 35+51	/au/	ao ban 凹版 55+214		wu chi 舞池 214+35		yi dao 医道 55+51		yi chang 异常 51+35
	ai xing 矮星 214+55	/u/	wu gui 乌龟 55+55		wu dao 舞蹈 214+214		yi xin 疑心 35+55		yi zhi 意旨 51+214
	ai ran 蔼然 214+35		wu po 巫婆 55+35		wu bi 舞弊 214+51		yi chuan 遗传 35+35		yi lun 议论 51+51

2.2 Subjects and Instrumentation

Recordings of the wordlist were obtained from 42 females (mean age = 24.0 years, standard deviation (SD) = 2.1) and 30 males (mean age = 22.7 years, SD = 1.9), all of whom were undergraduates or postgraduates in universities. They were able to speak standard Mandarin

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

Chinese and use it freely for daily communication. None of them had any voice or hearing problems, and they were all in sound health at the time of testing. The process of recording was accomplished in the sound-treated booth at the Language Lab of the Chinese Department, Beijing University, where the background noise was below 25 dBA. While recording, the Adobe Audition 1.5 was set at the double-channel interface with a sampling rate of 44100 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 personal computer through a sound console (Behringer XENYX502). With their lips about 10 cm away from the microphone, the subjects were asked to read aloud the disyllabic words using their most comfortable pitch, loudness and rate. Each subject read the wordlist twice, and the one reading with better quality was chosen for the present study.

Not counting the 13 bad-quality culls, the total number of speech samples eventually acquired was 3587 with 2095 from females and 1492 from males. Among the 3587 first syllables of words, 1036 were spoken with *yinping* (55), 1075 were spoken with *yangping* (35), 640 with *shangsheng* (21) and 836 with *qusheng* (51). The original tone contour 214 was absent from the database and the number of *shangsheng* tokens was smaller than others because of the Tone Sandhi described above: 214+55/35/51→21+55/35/51, 214+214→35+214. Moreover, some subjects produced the first syllable of "*yiban* 一般 (55+55)" as 51, while others did it as 55.

2.3 Parameter Extraction and Data Preprocessing

VAT measures were extracted largely automatically from the EGG and SP signals using the computer program developed by Roark, Watson, Baken, Brown, and Thomas (2012), the process of which consisted of four components. The second component was to automatically identify a 600-millisecond segment of the SP and EGG signals that was centered at the approximate time of vocal onset of the first syllable of the disyllabic word. This was based on two criteria that had to be simultaneously satisfied for the band pass-filtered EGG signal: Local energy had to be greater than 15% of the maximum energy and local cycle length must

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

have shown less than 15% variation. However, observation had shown that, for 107 speech samples from our database, the 600-millisecond segments thus identified were not centered at the voice onset of the first syllables, but somewhere else, for example at the onset of the second syllables, suggesting that inadequate EGG signal quality of these samples failed the two criteria above. VAT measures of these samples were marked "WS" (wrongly segmented) in the comments column of excel sheets.

From the 3587 VAT values obtained, the 107 "WS" measures were first taken away, and the remaining 3480 were divided into two groups: 2025 measures for females, and 1455 measures for males. Each group was then processed separately in the same way: Because of the large number of outliers among VAT values, measures that were beyond ±2 standard deviations from the mean VAT were cut out. Eventually, another 100 measures were deleted from the database, and among the 3380 that remained (skewness =-0.32, kurtosis =3.397), VAT ranged from -40.4 ms to 37.1 ms, the average and standard deviation of it being -0.32 ms and 7.16 ms respectively. SPSS 13.0 (SPSS. Inc. USA) was used for all the analyses below.

3 RESULTS

Among the preprocessed database, there are 1000 speech samples, whose first syllables carry mid-rise pitch contours of *yangping* (35), and of these contours, 195 are derived from *shangsheng* (214) by this pattern of Tone Sandhi: 214+214→35+214. A one-sample *t* test showed that the VAT values of these 195 derived *yangping* contours are not significantly different from those of the common *yangping* contours ($t_{(194)}$ =1.486, p = 0.139 > 0.05). Similarly, there were 43 subjects, who pronounced the first syllable of "*yiban* 一般 (55+55)" as 51, a full fall, and according to another one-sample *t* test, the VAT values of these 43 derived *qusheng* pitch contours (51) are not significantly different from those of the common *qusheng* contours either ($t_{(42)}$ = 0.822, p = 0.416 > 0.05). It is therefore reasonable to regard these derived pitch patterns of 35 and 51 as belonging to common *yangping* and *qusheng* categories respectively.

Since the final data corpus is composed of 1961 VAT measures for females, 1419 for males, 992 for *yingping* (55), 1000 for *yangping* (35) (including 195 derived ones), 599 for *shangsheng* (21) and 789 for

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

qusheng (51) (including 43 derived ones), a two-way analysis of variance with mixed measures on two factors was done with speaker gender (female vs. male) as the between-subject factor and tone (the 4 tones) as the within-subject factor. Results revealed a significant main effect of tone ($F_{(2.444)} = 33.59$, $p < 0.05$, $R^2 = 0.324$), a non-significant main effect of gender ($F_{(1,70)} = 0.179$, $p = 0.673 > 0.05$, $R^2 = 0.003$), and a non-significant tone by gender interaction effect ($F_{(2.444)} = 0.262$, $p = 0.813 > 0.05$, $R^2 = 0.004$). Means and SDs of VAT calculated among the 72 subjects are listed in Table 2 and plotted in Figure 1.

Table 2 Means and SDs of VAT of different tones and genders among all the 72 subjects

	Sex	Mean	Std. Deviation	N
yinping(55)	female	-1.602	5.406	42
	male	-1.710	4.235	30
	Total	-1.647	4.920	72
yangping(35)	female	1.573	3.878	42
	male	2.176	3.675	30
	Total	1.824	3.781	72
shangsheng(21)	female	.903	4.826	42
	male	1.612	4.812	30
	Total	1.198	4.799	72
qusheng(51)	female	-2.462	6.595	42
	male	-2.017	3.576	30
	Total	-2.276	5.512	72

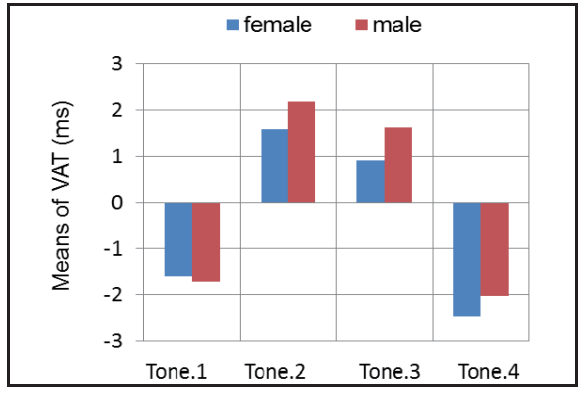


Figure 1 Means of VAT as a function of tone and gender (72 subjects)

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

Figure 1 indicates that the average VATs for three of the four tones are all smaller for females than for males, but those for tone1 are not, suggesting the necessity for analyses of simple effect. Thus a paired-samples *t* test has further shown that, for both males and females, VATs between *yinping* (55) and *yangping* (35), *shangsheng* (21) and *qusheng* (51) are significantly different from each other (For males: 55 vs.35: $t = -5.421$, $p = 0.00 < 0.05$; 21 vs. 51: $t = 4.479$, $p = 0.00 < 0.05$. For females: 55 vs.35: $t = -4.526$, $p = 0.00 < 0.05$; 21 vs. 51: $t = 4.291$, $p = 0.00 < 0.05$), while those between *yangping* (35) and *shangsheng* (21), *yinping* (55) and *qusheng* (51) are not (For males: 35 vs. 21: $t = 1.007$, $p = 0.322 > 0.05$; 55 vs. 51: $t = 0.702$, $p = 0.488 > 0.05$; For females: 35 vs. 21: $t = 1.222$, $p = 0.229 > 0.05$; 55 vs. 51: $t = 1.34$, $p = 0.188 > 0.05$). Namely, the means of VAT are much longer for *yangping* and *shangsheng* than for *yingping* and *qusheng* in both genders. This is why a cluster analysis has put *yinping* and *qusheng* into one category, but *yangping* and *shangsheng* into another depending on the VAT measures of their voice onsets (see Figure 2).

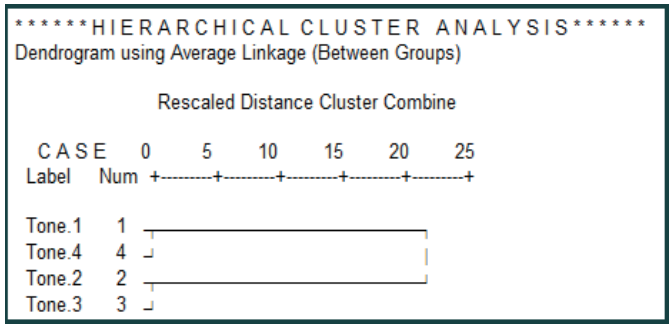


Figure 2 Result of a hierarchical cluster analysis

However, a close inspection on the mean VATs of the four tones (55, 35, 21, and 51) of each subject can also divide the 72 subjects (including males and females) into two groups: 46 of them (63.89%) have mean VATs of both *yangping* and *shangsheng* longer than those of *yinping* and *qusheng* (see Figure 3), but 26 of them (36.11%) display various patterns other than this (see Figure 4).

©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

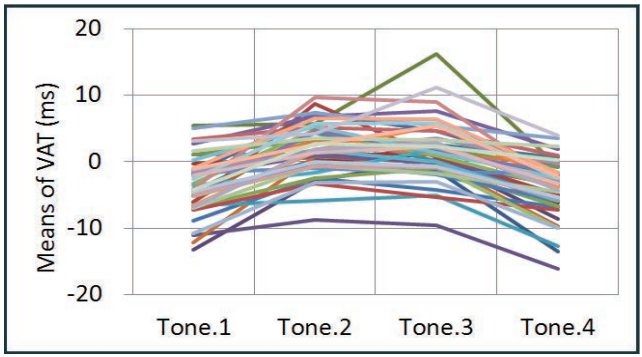


Figure 3 46 of the 72 subjects displayed mean VATs of four tones in the same pattern. Each line indicates the average VATs of one person.

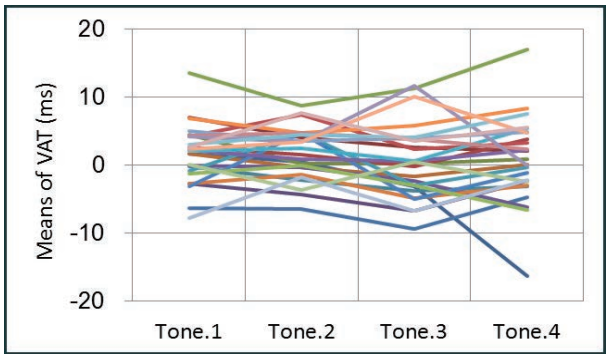


Figure 4 26 of the 72 subjects displayed mean VATs of four tones in miscellaneous patterns. Each line indicates the average VATs of one person.

A two-way analysis of variance with mixed measures on two factors test done on the 46 still presents a significant main effect of tone ($F_{(3)}$ = 87.644, $P < 0.05$, $R^2 = 0.666$), a non-significant main effect of gender ($F_{(1,44)} = 2.163$, $p = 0.149 > 0.05$, $R^2 = 0.047$) and a non-significant tone by gender interaction effect ($F_{(3)} = 0.632$, $P = 0.595 > 0.05$, $R^2 = 0.014$). In Table 3 and Figure 5 are indicated the means and SDs of VAT of the 46 subjects as a function of tone and gender. The exception in Figure 1 no longer shows up in Figure 5, suggesting that individual differences may slightly affect the result of the whole population.

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

Table 3 Means and SDs of VAT of different tones and genders among the 46 subjects

	Sex	Mean	Std. Deviation	N
yinping(55)	female	-3.847	4.641	26
	male	-2.992	3.746	20
	Total	-3.475	4.251	46
yangping(35)	female	1.199	4.193	26
	male	2.751	3.147	20
	Total	1.874	3.815	46
shangsheng(21)	female	.807	4.781	26
	male	2.778	3.630	20
	Total	1.664	4.386	46
qusheng(51)	female	-4.839	4.621	26
	male	-2.811	3.713	20
	Total	-3.957	4.326	46

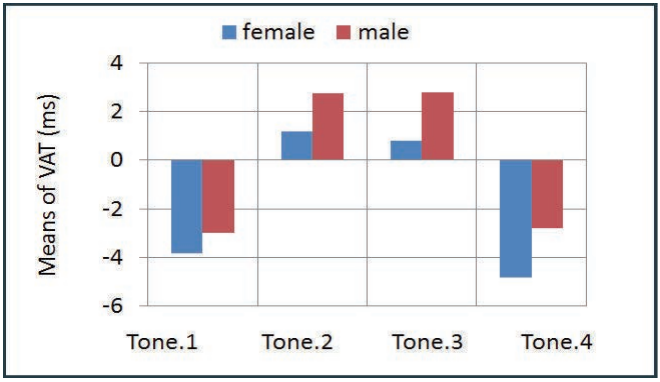


Figure 5 Means of VAT as a function of tone and gender (46 subjects)

4. DISCUSSION

4.1 Tone Sandhi

One debate concerning tone sandhi in Mandarin Chinese had been whether the tone sequence 3-3 is homophonous with the sequence 2-3. The issue was eventually settled by Wang and Li (1967), Wang and Peng 2006, 120-121) with a perception experiment. The 130 pairs of test items used in his research were thus designed: The two members of each pair shared the same phonological features except that of pitch contour; In

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

other words, one member carried the tone sequence 2–3 while the other carried 3–3. These items were recorded in random order and then presented randomly for native speakers of Mandarin to listen to and identify whether they were sequences 3-3 or 2-3. None of the listeners were able to present a rate of accuracy over 55%, and even the person from whom the test items had been recorded couldn't correctly identify over 60% of them, suggesting that the *yangping* pitch contour (35) derived from *shangsheng* (214) was perceptually no different from that of the common *yangping* (35). The result of the first one-sample *t* test done in the present study supports this argument from the perspective of physiology: VAT values of the second tone derived from the third tone are not significantly different from those of the common second tone, suggesting that all the *yangping* contours, whether original or derived, share similar laryngeal adjustments at their voice onsets, and therefore display approximate features of vocal attack. In a word, both perception and physiology point to one conclusion: The pitch contour 214 before another 214 is indeed phonemically the same as *yangping*. The result of the second one-sample *t* test leads to a similar judgment: The onsets of phonation of all the *qusheng* contours, whether the original 51 or the ones derived from 55, are physiologically alike, and the two kinds should be perceptually no different.

4.2 VAT and Lexical Tones

The second finding from the analyses above can be summarized as follows: In a large group of Mandarin speakers, the two lexical tones with high-pitch onsets, *yinping* (55) and *qusheng* (51), display smaller VAT values, but the other two with low-pitch onsets, *yangping* (35) and *shangsheng* (21), present much larger ones (see tables 2 and 3 and figures 1, 2, 3 and 5); In other words, a higher rate of vocal-fold oscillation tends to be associated with a shorter VAT value, and vice versa. This negative VAT-F0 correlation at the linguistically constrained voice onset is also seen in the three level tones of Cantonese (Ma et al., 2012): In females, mean VATs of high, mid and low level tones are respectively 0.72 ms, 1.70 ms, 1.78 ms; In males, mean VATs of high, mid, low level tones, although longer than those in females, are also thus lined up, 3.99ms,

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

4.64ms, 4.69ms. However, Tables 2 and 3 also indicate counterexamples: For both males and females, mean VATs of *shangsheng* (21) should always be larger than those of *yangping* (35), because the former has a lower initial pitch than the latter; but this is actually not the case. These conform to the finding of the VAT study on 5 linguistically unconstrained pitch levels in Mandarin (Zhang et al. 2015). In a large group of people, as pitch levels shift from one to five, there is a linear increase of pitch, but a nonlinear decrease of VAT: from Levels Two to Five, each mean value of VAT is not always larger than the one that follows; But, the average VAT at Level One is always the largest among the five pitch levels, and is much larger than that of all the others. Therefore, for both linguistically constrained and unconstrained vocal onsets, VAT and pitch tend to present a nonlinear contra-variant relationship in most mandarin speakers.

4.3 Individual Differences

46 of the 72 subjects produced the low-pitch onsets of the second and third tones (35, 21) with longer VAT means than they did the high-pitch onsets of the first and fourth tones (55, 51), while 26 of them showed inconsistent patterns of VAT means in pronouncing the four. This seems to support the findings by Zhang et al. (2015) that as pitch means increase linearly from Levels One to Five, mean VATs decrease nonlinearly in a large group of people but increase nonlinearly in a small group of them, and that different people incline to use different strategies in increasing pitch height. However, among the 26 subjects observed in the present study, mean VATs of four tones are ordered as *yangping* (35) 1.736 ms > *yinping* (55) 1.586ms > *qusheng* (51) 0.697 ms > *shangsheng* (21) 0.375 ms, and a positive VAT-F0 correlation is not seen at the phonation onsets of the four tones. What caused the individual differences needs to be further investigated.

5. CONCLUSION

Firstly, vocal attack time, as a measure of phonatory function of the vocal folds, shows no significant difference between the common *yangping* and the *yangping* derived from *shangsheng*, and between the common *qusheng* and the *qusheng* derived from *yinping*. This is

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

physiologically in support of the argument that the tone sequence 3-3 in Mandarin is indeed converted into 2-3, nothing else. Secondly, the tones of Mandarin Chinese that start from low pitch levels (35, 21) tend to present significantly different VAT values from those that start from high pitch levels (55, 51), with mean VATs of the former being much longer than those of the latter. This is with the nonlinear contra-variant relationship between VAT and F0 at the vowel onsets. Thirdly, there are deviations or individual differences: a small number of people do not follow this pattern.

REFERENCES

BAKEN，R. J., and R. F. Orlikoff 1998. Vocal fold adduction time estimated from glottographic signals. Poster presented at: the 25th Mid Winter Meeting of the Association for Research in Otolaryngology, St. Petersburg, FL (February 1998).

_____. 1999. Estimating vocal fold adduction time from EGG and acoustic records. In *Programme and Abstract Book : Xxivth World Congress of the International Association of Logopedics and Phoniatrics (IALP)*, ed. H.K. Schutte, P. Dejonckere, H. Leezenberg, B Mondelaers, and H. F. Peters, 15. Nijmegen: Uitgeverij Katholieke Universiteit Nijmegen. (24th World Congress of IALP was held in 1998)

CHAO, Yuen Ren. 1970. *A Grammar of Spoken Chinese*. Berkeley (CA): University of California Press.

MA, E., R. J. Baken, and R. M. Roark. 2012. Effect of tones on vocal attack time in Cantonese speakers. *Journal of Voice* 26(5): 670. e1-670. e 6.

ORLIKOFF, R.F., D.D. Deliyski, R.J. Baken, and B.C. Watson. 2009. Validation of a glottographic measure of vocal attack. *Journal of Voice* 23 (2): 164 -168.

ROARK, R.M., B.C. Watson, and R.J. Baken. 2012. A figure of merit for vocal attack time measurement. *Journal of Voice* 26(1):8-11.

ROARK, R.M., B.C. Watson, R.J. Baken, D.J. Brown and J.M. Thomas. 2012. Measures of vocal attack time for healthy young adults. *Journal of Voice* 26(1):12-17.

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

WANG, S-Y. William, and K.P. Li. 1967. Tone 3 in Pekingese. *Journal of Speech Hearing Research* 10(3):629-36.

WANG, S-Y. William (Wang Shiyuan) and Peng, Gang. 2006. *Yu yan, yu yin **yu** ji shu* 语言，语音与技术 (Language, speech and technology). Shanghai: Shanghai Jiaoyu Chuban She 上海教育出版社.

ZHANG, Ruifeng, R. J. Baken,, and Kong Jiangping. 2015. Vocal attack time of different pitch levels and vowels in Mandarin. *Journal of Voice* 29(5):542-547.

普通话四声的声门碰撞时间研究

孔江平　张锐锋
北京大学

提要

汉语作为典型的声调语言，有着四个能区分意义的词调：阴平(55)，阳平(35)，上声（214）和去声（51）。但由于连读变调，在语流中经常出现的四声调值为：阴平（55），阳平（35），上声（21）和去声（51）。本文旨在研究汉语普通话者的声门碰撞时间（Vocal Attack Time, or, VAT）与声调的关系，以探索起始于不同音高层级的声调对该参数的影响情况。我们从 72 位发音人（男 30 人和女 42 人）同时录制了 50 个双音节词的语音和喉头仪信号，所有发音人均为 20 多岁的在校大学生或研究生。单独样本的 T 检验表明：阳平以及由上声变来的阳平，去声以及由阴平变来的去声在声门碰撞时间上不存在显著性差异。这就从生理上支持了两上声相连，前上变阳平的观点。二因素重复测量的方差分析表明：起始于低音高层级的声调与起始于高音高层级的声调在声门碰撞时间上存在着显著性差异，前者的值明显大于后者。但还存在有个体差异，72 位发音人中有 26 位不符合这一模式。

关键词

声门碰撞时间 **字**调 **发**声起始 **非**线性反变关系

The Journal of Chinese Linguistics vol.45, no.2 (June 2017): 275-289
©2017 by The Journal of Chinese Linguistics. All rights reserved.
0091-3723/2017/4502-01: VAT of the lexical tones in Mandarin Chinese

ISSN 1000-0054 清华大学学报（自然科学版） 2017 年 第 57 卷 第 2 期 11/18
CN 11-2223/N J Tsinghua Univ (Sci & Technol), 2017, Vol. 57, No. 2 176-181

汉语普通话不同文体朗读时的胸腹呼吸特性

杨 锋[1]， 孔江平[2]

（1. 暨南大学 华文学院，应用语言学研究院，广州 510610；2. 北京大学 中国语言文学系，中国语言学研究中心，北京 100871）

摘 要：该文主要研究汉语普通话不同文体朗读时的胸腹呼吸特性。实验同步录制了 4 位播音员朗读的语音和胸腹呼吸信号，选取近体诗、词、新闻和散文这 4 种文体语料 90 篇，使用自主编写的呼吸分析程序，提取了吸气相和呼气相的呼吸重置幅度、时长、斜率、面积等参数。实验结果表明：言语状态下的胸腹呼吸与自然呼吸不同；不同文体朗读时的胸腹呼吸特点不同，诗、词朗读中的各级呼吸重置分布更具规律性，新闻和散文朗读中胸腹呼吸重置可分为 3 级，对应出现在段首、复句首和分句首位置；朗读中主要以胸腹联合式呼吸为主，胸腹呼吸二者的作用不同。该研究对汉语韵律研究和言语产生具有重要意义，为言语产生的呼吸生理建模提供参考。

关键词：实验语音学；言语产生；呼吸控制

中图分类号：H 017 **文献标志码**：A

文章编号：1000-0054(2017)02-0176-06

DOI：10.16511/j.cnki.qhdxxb.2017.22.011

Chest and abdomen breathing while reading literature in Mandarin

YANG Feng[1], KONG Jiangping[2]

(1. Institute of Applied Linguistics, College of Chinese Language and Culture, Jinan University, Guangzhou 510610, China;
2. Research Center for Chinese Linguistics, Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

Abstract: Chest and abdomen breathing characteristics were studied while reading Mandarin Chinese literature. The speech and breathing signals were recorded from four annoucers while reading 90 passages with different styles, including modern poems, ci, news, and essays. The extracted breathing parameters were the breathing reset amplitude, duration, slope, and area during the inhalation and exhalation phases. The results show that the breathing strategy during speech differs from that during natural breathing with different chest and abdomen breathing patterns for different styles of prose. Systematic breathing resets occur while reading poems and ci. Three levels of breathing resets are identified while reading news and essays at the paragraph-initial, main clause-initial and subordinate clause-initial positions. The chest and abdomen breathing are coordinated during all the reading activities, but each has its distinct functions. These results are of significance to prosody and speech production and provide a basis for articulatory modeling of breathing in speech production.

Key words: experimental phonetics; speech production; breathing control

呼吸是言语产生的动力，语言的进化得益于呼吸器官的进化和人类对呼吸器官的灵活控制能力[1-2]。对呼吸的研究是研究言语产生的一个重要方面，但由于言语过程中呼吸器官运动的复杂性和研究设备所限，对言语呼吸特性的研究并未深入展开。

Ladefoged 等的研究表明呼吸与韵律相关，具有实现重音和感知语调的重要作用[3-9]。不同的言语状态中，呼吸器官运动方式不同[10]。朗读中存在呼吸规划，与篇章语法结构相关[11]。

国内吴宗济等最早利用气流气压信号对汉语进行研究[12]。近些年不少学者对汉语普通话朗读时的腹呼吸重置特点进行了研究[13-15]，但对普通话朗读呼吸的分析只使用到了腹呼吸信号，未研究胸腹呼吸二者的特点、作用和关系。也有文献对言语胸腹呼吸机制进行了初步探索[16-17]，但对不同文体朗读的胸腹呼吸缺少大数据量的分析，尚未建立汉语言语呼吸模型。

本文同步录制了大量汉语普通话近体诗、词、新闻和散文 4 种文体语料朗读时的胸呼吸、腹呼吸和语音信号，对胸、腹呼吸信号和语音信号进行同步分析，研究不同文体朗读时的胸腹呼吸规律、胸腹呼吸二者的作用和相互关系。

收稿日期：2016-06-20

基金项目：国家社会科学基金重大项目(10&ZD125,16ZDA211)

作者简介：杨锋(1978—)，男，副教授。

通信作者：孔江平，教授，E-mail：jpkong@pku.edu.cn

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1　实验说明

1.1　语料和发音人

本文同步录制 4 名播音员朗读的语音、胸呼吸和腹呼吸信号，语料包括近体诗 30 篇、词 20 篇、新闻 20 篇、散文 20 篇，语料尽量挑选发音人熟悉的、200 字左右的篇章。4 位播音员年龄在 25～35 岁之间，2 男 2 女，受过良好的播音训练。

1.2　信号采集

本实验的录音工作在北京大学中国语言文学系语言学实验室录音室内进行。录音使用的采集设备是 AD Instruments PowerLab PL3516 16 通道高速记录仪，使用 Chart5 同步采录 4 个通道的信号：第 1 通道为通过麦克风和调音台采集的语音信号，第 2 通道是通过电子声门仪(EGG)采集的嗓音信号，第 3 通道为通过呼吸带传感器采集的胸呼吸信号，第 4 通道为腹呼吸信号。采样频率均为 44 kHz。胸呼吸带系在腋下 10 cm 处，腹呼吸带系在肚脐上方。

录音使用的话筒是 Sony ECM-44B，调音台是 Behringer XENYX502。电子声门仪是 KAY6103。胸腹两根呼吸带是 AD Instrument MLT1132。

2　参数提取

本文使用 Matlab 编写程序建立分析平台，如图 1 所示，对语音、嗓音、胸呼吸、腹呼吸 4 路信号进行同步分析，提取语音、嗓音和呼吸 3 类参数(本文主要使用呼吸参数)。

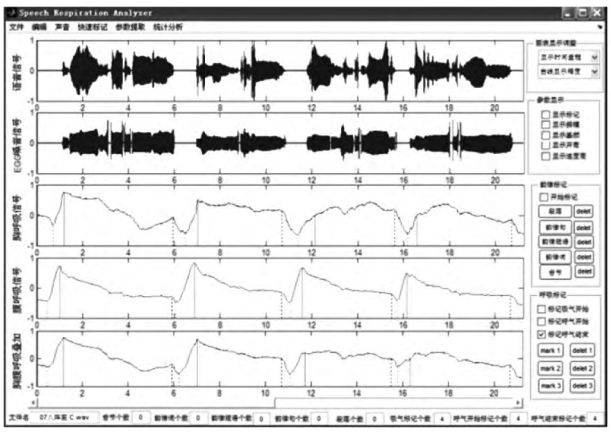


图 1　用 Matlab 编写的数据分析平台界面

呼吸参数定义如图 2 所示，以腹呼吸为例，横轴是时间，纵轴是归一化呼吸幅度，曲线是腹呼吸信号，灰度波形背景是同步的语音信号，曲线上升表示吸气，曲线下降表示呼气。t_A、t_B、t_C 分别是吸气开始时间、呼气开始时间、呼气结束时间，吸气段时长 $TI=t_B-t_A$，呼气段时长 $TE=t_C-t_B$，AI 为吸气重置幅度，AE 为呼气重置幅度。吸气段斜率、呼气段斜率、吸气相面积和呼气相面积通过以上参数求得。

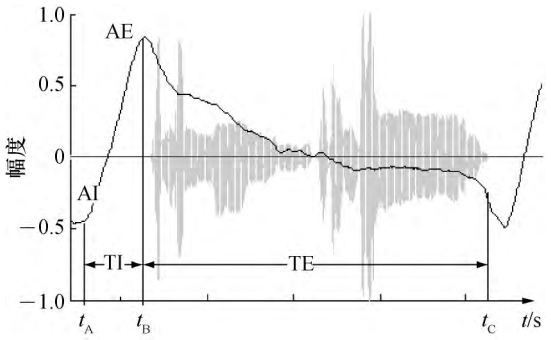


图 2　呼吸参数定义

首先，对呼吸信号进行低通滤波，去除高频干扰信号。然后，对呼吸信号进行归一化处理，接着标记胸腹呼吸每个周期的吸气开始时间点、呼气开始时间点、呼气结束时间点，程序由此计算出其他呼吸参数，将标记后的语音文件和参数一起保存为 mat 格式的文件，并输出呼吸参数到 Excel 中进行统计分析。

3　实验结果分析

3.1　自然呼吸胸腹呼吸特性

自然呼吸状态下胸腹呼吸每个周期的吸气相和呼气相时长、斜率和面积都接近，显著区别于言语状态，言语状态下呼气相的时长和面积远大于吸气相，主要原因是呼气相承载着语音。非言语自然呼吸状态下，使用胸腹联合式呼吸的发音人居多。

图 3 所示为自然呼吸状态下的胸腹呼吸信号，共 4 个呼吸周期，平均每个呼吸周期为 3 s，周期性规律显著。如表 1 所示，胸呼吸吸气相和呼气相时

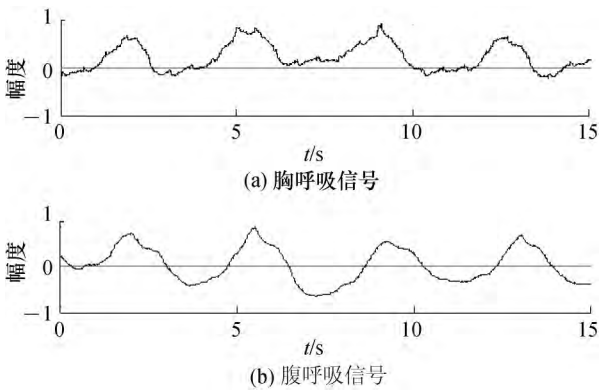


图 3　自然呼吸状态下的胸腹呼吸信号

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

长均值分别为 0.96 和 0.89，面积均值为 0.41 和 0.37，上升斜率为 1.04，下降斜率为−1.72，表明胸呼吸每个周期的吸气相和呼吸相近似。每个周期呼气重置幅度均值为 0.67，呼吸间隔均值为 1.8 s。

表 1 自然呼吸状态下的胸腹呼吸参数均值

吸气相	幅度 AI	时长 TI/s	斜率	面积
胸呼吸	−0.19	0.96	1.04	0.41
腹呼吸	−0.15	0.83	1.36	0.40
呼气相	**幅度 AE**	**时长 TE/s**	**斜率**	**面积**
胸呼吸	0.67	0.89	−1.72	0.37
腹呼吸	0.80	0.85	−1.22	0.36

自然呼吸状态下腹呼吸呈现同样特征，吸气相和呼气相近似，腹呼吸吸气相和呼气相时长均值分别为 0.83 和 0.85，面积均值为 0.40 和 0.36，上升斜率为 1.36，下降斜率为−1.22。每个腹呼吸周期呼气重置幅度均值为 0.80，每个呼吸周期间隔 1.6 s。

3.2 近体诗朗读胸腹呼吸特性

本实验共选取了五言律诗、七言律诗、五言绝句和七言绝句共 30 首。由于近体诗格律规整，结构简单，因此胸腹呼吸类型也比较简单。其胸腹呼吸特点是：胸腹呼吸重置可以分为 2 级，每联(上下两小句为一联)起始处的呼吸重置较大，每小句起始处的胸腹呼吸重置较小，呼气相时长与语音时长一致。图 4 所示是七言律诗《登高》前 4 句朗读的胸腹呼吸信号，灰色波形背景是同步的语音信号，汉字是对应的文本。图中 A、B、C、D 是 4 句的起始位置，有 4 个呼吸重置，A 和 C 处是每联的起始位置，胸腹重置幅度都较大，胸呼吸重置幅度均值为 0.89，腹呼吸重置幅度均值为 0.82，大于 B 和 D 处的对应均值 0.56 和 0.55。语音段与呼气相重合，但 4 处呼吸重置时间早于语音起始时间平均 0.21 s，与语音生理机制相符，胸腹呼吸器官开始呼气，气流进入喉部触动声带振动产生语音，中间有短暂的时间间隔。

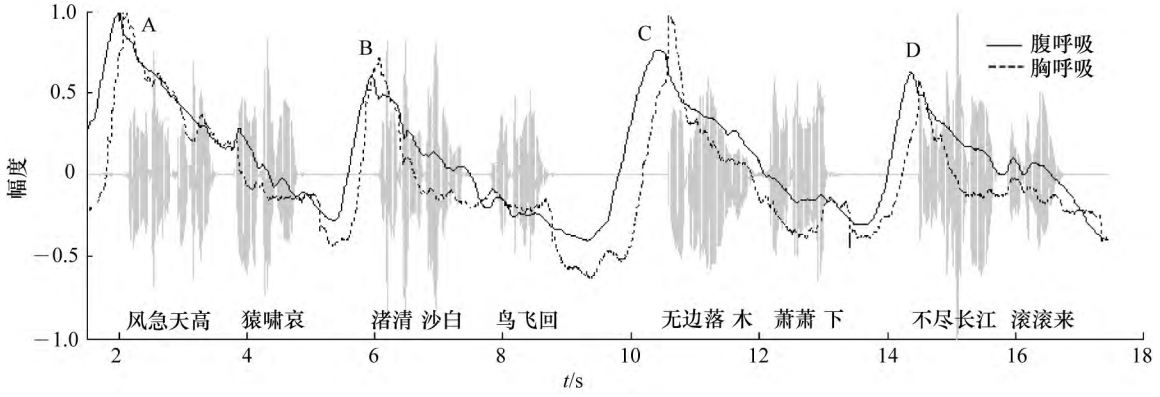


图 4 七言律诗《登高》前 4 句朗读的胸腹呼吸信号

3.3 词朗读胸腹呼吸特性

本次实验选取了 20 首常见的词，均是发音人熟悉的词。图 5 所示是词《赤壁怀古》朗读的腹呼吸信号，图中 A 和 B 标记处是上阕和下阕的起始位置，呼吸重置幅度分别是 0.95 和 0.96，显著大于其他位置的呼吸重置幅度。对其中一位发音人朗读的 20 首词的胸腹呼吸呼吸重置幅度参数进行 K-means 聚类分析，结果如图 6 所示，得到 3 类，即 3 级呼吸重置：1 级呼吸胸腹重置幅度为 0.83 和 0.92，2 级为 0.37 和 0.46，3 级为 0.22 和 0.17。从 3 级呼吸重置所在的时间点可知，1 级呼吸重置出现在上阕和下阕起始处，2 级呼吸重置大多出现在每句起始处，3 级出现在句内停顿处。

3.4 新闻朗读胸腹呼吸特性

新闻朗读的胸腹呼吸特点是：胸腹呼吸重置分为 3 级，每段段首都有一个 1 级呼吸重置，复句起始处为 2 级呼吸重置，分句起始处为 3 级呼吸重置。本次实验选取了 20 条 200 字左右的短篇新闻，均是发音人熟悉的篇章。图 7 所示是一篇新闻朗读的腹呼吸和同步的语音信号，图中 A、B、C 3 处峰值所在位置，是 3 段的段首位置，根据图 8 中聚类分析结果显示，这 3 处可视为 1 级呼吸重置。2 级胸腹呼吸重置的均值分别为 0.41 和 0.46，出现在复句起始端，3 级胸腹呼吸重置的均值分别为 0.06 和 0.17，出现在分句起始端。由于 4 位发音人均是专业的播音员，受过专业播音训练，同时新闻短文具有相对固定的格式，因此每位发音人的呼吸信号规律性都较强。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

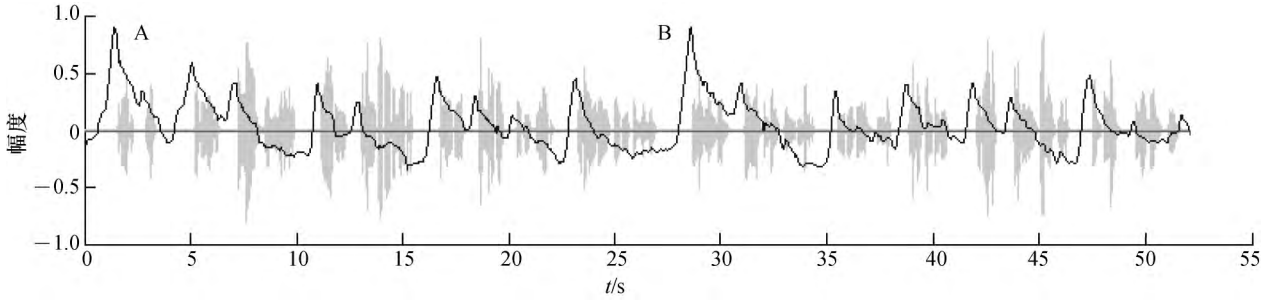


图 5　词《赤壁怀古》朗读的腹呼吸和语音信号

最终聚类中心

	聚类		
	1	2	3
胸呼吸幅度	0.83	0.37	0.22
腹呼吸幅度	0.92	0.46	0.17

每个聚类中的案例数

聚类	1	41.000
	2	237.000
	3	74.000
有效		352.000
缺失		0.000

图 6　对 20 首词朗读的胸腹呼吸呼吸重置幅度的 *K*-means 聚类分析结果

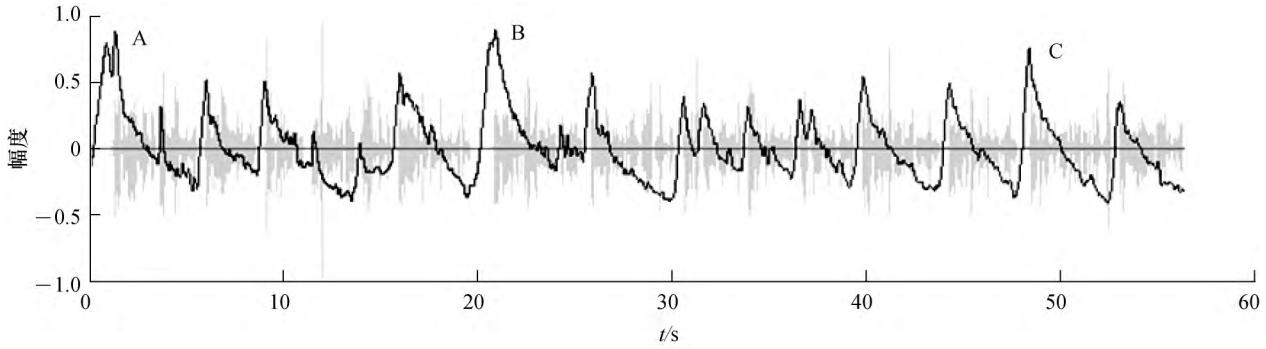


图 7　新闻朗读的腹呼吸和语音信号

最终聚类中心

	聚类		
	1	2	3
胸呼吸幅度	0.79	0.06	0.41
腹呼吸幅度	0.78	0.17	0.46

每个聚类中的案例数

聚类	1	67.000
	2	73.000
	3	246.000
有效		386.000
缺失		0.000

图 8　对 20 条新闻朗读的胸腹呼吸呼吸重置幅度的 *K*-means 聚类分析结果

3.5　散文朗读胸腹呼吸特性

散文朗读的胸腹呼吸特点是：胸腹呼吸重置也分为 3 级，但 3 级呼吸重置出现的位置不如新闻朗读中固定，段中或句中也可出现 1 级和 2 级呼吸重置。本次实验选取了 20 篇 200 字左右的散文，均是发音人熟悉的篇章或段落。图 9 所示的是一篇散文朗读的腹呼吸和同步的语音信号，1 级呼吸重置出现的次数明显增多，而且出现的位置并不仅限于段首，根据图 10 中聚类分析结果显示，1 级和 2 级呼吸重置数量比新闻朗读中明显增加，主要原因可能是散文不同于新闻，发音人需有感情的朗读，由此带来 1 级和 2 级呼吸重置次数的增多。

3.6　言语中胸腹呼吸模式

对本次实验中所录制的胸腹呼吸信号根据呼吸参数进行统计分析，可得到胸腹呼吸的模式，如图 11 和 12 所示。在朗读中主要使用胸腹联合式呼吸，

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

胸呼吸曲线呈现梯形，表明胸腔在言语时保存扩张状态，而腹呼吸信号呈现斜三角形，在言语过程中腹部持续收缩，直至语音结束。由此可知，在胸腹联合式呼吸朗读中，胸呼吸主要作用是在发音时使胸腔保持扩张状态至发音结束，为发音提供足够的气息和胸腔共鸣；腹呼吸主要作用是通过腹肌和膈肌的稳健收缩，保证稳定的声门下压，控制气流持续释放，以获得连续的语音。

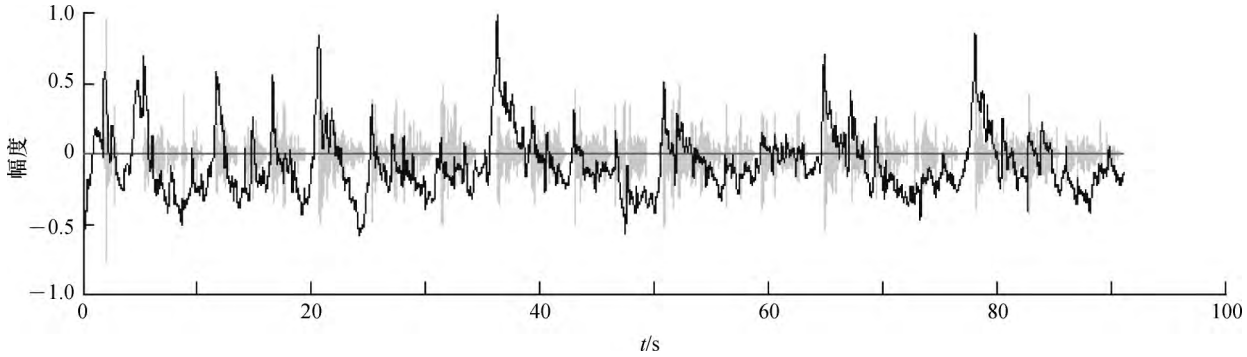


图 9 散文朗读的腹呼吸和语音信号

最终聚类中心

	聚类		
	1	2	3
胸呼吸幅度	0.55	0.76	0.13
腹呼吸幅度	0.53	0.83	0.15

每个聚类中的案例数

聚类	1	247.000
	2	85.000
	3	310.000
有效		642.000
缺失		0.000

图 10 对 20 篇散文朗读的胸腹呼吸呼吸重置幅度的 *K*-means 聚类分析结果

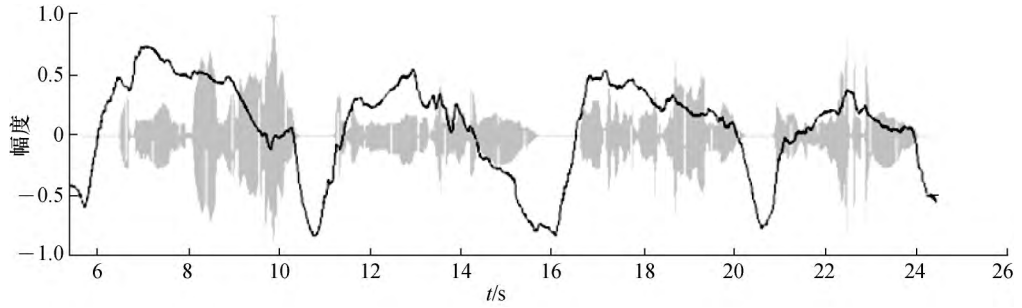


图 11 典型的胸呼吸模式与语音

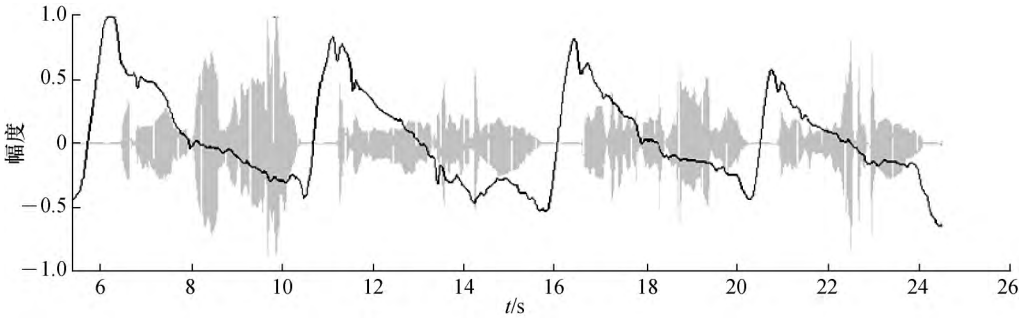


图 12 典型的腹呼吸模式与语音

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4　结论与展望

本文主要通过分析汉语普通话中诗、词、新闻和散文4种文体语料朗读时的胸腹呼吸信号，研究不同文体朗读时的胸腹呼吸信号特点。使用自主编写的呼吸分析程序，分析同步录制的语音和胸腹呼吸信号，提取了吸气相和呼气相的呼吸重置幅度、时长、斜率、面积等参数，进行了统计分析。实验结果表明：

1）言语状态下的胸腹呼吸与自然呼吸不同，自然呼吸状态下胸腹呼吸每个周期的吸气相和呼气相时长、斜率和面积都接近，言语状态下呼气相的时长和面积远大于吸气相，呼气相承载着语音信号，语音时长和呼气相时长相等。

2）不同文体朗读时的胸腹呼吸特点不同，朗读诗、词的各级呼吸重置分布更具规律性。新闻朗读中胸腹呼吸重置可分为3级，3级呼吸重置对应出现的位置分别是段首、复句首和分句首，散文朗读中呼吸重置大小可因情感语音需要而发生变化。

3）朗读中主要以胸腹联合式呼吸为主，在发音过程中胸腔保持扩张状态至发音结束，胸呼吸曲线呈现梯形，为发音提供足够的气息和胸腔共鸣；腹呼吸信号呈现斜三角形，腹部持续收缩至语音结束，通过腹肌和膈肌的稳健收缩，保证稳定的声门下压，控制气流持续释放，以获得连续的语音。

不同级别呼吸重置出现的位置并不相同，与语法结构、韵律单元边界存在着一定的对应关系，下一步的研究应从呼吸角度探索汉语的韵律特点，同时研究不同发音人呼吸的个性特征，以及多人实时对话时话语轮回的呼吸特点。由于本次实验中胸腹呼吸信号录制的是胸腔和腹腔的变化，所得数据是相对值，要获得更为精准的呼吸数据就需要同步录制气流气压信号，这样可得到气流量、声门下压力、肺的容积变化等更加精确的数据。同时，结合螺旋CT拍摄的肺部图像，建立肺部的3维动态几何模型，可更加深入地研究言语产生的呼吸机制。

参考文献　(References)

[1] MacLarnon A, Hewitt G. Increased breathing control: Another factor in the evolution of human language [J]. *Evolutionary Anthropology: Issues, News, and Reviews*, 2004, **13**(5): 181-197.

[2] Gea J. The evolution of the human species: A long journey for the respiratory system [J]. *Archivos de Bronconeumología: English Edition*, 2008, **44**(5): 263-270.

[3] Ladefoged P, Loeb G. Preliminary studies on respiratory activity in speech [J]. *UCLA Working Papers in Phonetics*, 2002, **101**: 50-60.

[4] Draper M H, Ladefoged P, Whitteridge D. Respiratory muscles in speech [J]. *Journal of Speech, Language and Hearing Research*, 1959, **2**(1): 16-25.

[5] Ohala J J. Respiratory activity in speech [M]// Speech Production and Speech Modelling. Berlin: Springer Netherlands Press, 1990: 23-53.

[6] Hixon T J, Goldman M D, Mead J. 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*, 1973, **16**(1): 78-90.

[7] Stathopoulos E T, Hoit J D, Hixon T J, et al. Respiratory and laryngeal function during whispering [J]. *Journal of Speech, Language and Hearing Research*, 1991, **34**(4): 761-780.

[8] Slifka J L K. Respiratory Constraints on Speech Production at Prosodic Boundaries [D]. Boston, MA: Massachusetts Institute of Technology, 2000.

[9] Baken R J, Cavallo S A, Weissman K L. Chest wall movements prior to phonation [J]. *Journal of Speech and Hearing Research*, 1979, **22**(4): 862-875.

[10] Huber J E, Chandrasekaran B, Wolstencroft J J. Changes to respiratory mechanisms during speech as a result of different cues to increase loudness [J]. *Journal of Applied Physiology*, 2005, **98**(6): 2177-2184.

[11] Fuchs S, Petrone C, Krivokapić J, et al. Acoustic and respiratory evidence for utterance planning in German [J]. *Journal of Phonetics*, 2013, **41**(1): 29-47.

[12] 吴宗济，林茂灿. 实验语音学概要 [M]. 北京：高等教育出版社，1989：33-34.
WU Zongji, LIN Maocan. A Prime of Experimental Phonetics [M]. Beijing: Higher Education Press, 1989: 33-34. (in Chinese)

[13] 谭晶晶，李永宏，孔江平. 汉语普通话不同文体朗读时的呼吸重置特征[J]. 清华大学学报：自然科学版，2008，**48**(4)：613-620.
TAN Jingjing, LI Yonghong, KONG Jiangping. Breathing-reset when reading literature in Mandarin [J]. *Journal of Tsinghua University: Science and Technology*, 2008, **48**(4): 613-620. (in Chinese)

[14] 张锦玉，石锋，白学军. 讲述与朗读状态下呼吸差异的初步分析 [J]. 南开语言学刊，2012(1)：56-63.
ZHANG Jinyu, SHI Feng, BAI Xuejun. Preliminary analysis of respiratory diversity between the states of narrating and reading [J]. *Nankai Linguistics*, 2012(1): 56-63. (in Chinese)

[15] Shi F, Zhang J, Bai X, et al. Intonation and respiration: A preliminary analysis [J]. *Journal of Chinese Linguistics*, 2010, **38**(2): 323-335.

[16] 杨锋，侯兴泉，孔江平. 言语产生的胸腹呼吸机制 [J]. 清华大学学报：自然科学版，2013，**53**(6)：852-855.
YANG Feng, HOU Xingquan, KONG Jiangping. The chest and belly breathing control in speech production [J]. *Journal of Tsinghua University: Science and Technology*, 2013, **53**(6): 852-855. (in Chinese)

[17] YANG Feng. A study on the features of chest and abdominal breathing between reciting and chanting Chinese poetry [J]. *Journal of Chinese Linguistics*, 2015, **43**(1B): 399-410.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

ISSN 1000-0054 清华大学学报（自然科学版） 2017 年 第 57 卷 第 2 期 8/18
CN 11-2223/N J Tsinghua Univ (Sci & Technol), 2017, Vol. 57, No. 2 158-163

基于磁共振成像的汉语普通话舌尖调音建模

汪高武[1]， 党建武[2]， 孔江平[3]

（1. 北京师范大学 文学院，北京 100875；2. 天津大学 计算机科学与技术学院，天津 300072；
3. 北京大学 中国语言文学系，北京 100871）

摘 要：通过对汉语普通话磁共振成像数据的分析，对舌尖的形状和运动进行调音建模。建立了汉语普通话磁共振成像调音数据库，包括 9 个单元音和 75 个辅音变体。提取了发音器官在正中矢状面上的形状边缘；对舌头的形状边缘进行主成分分析，发现舌尖和舌体分开建模更为简洁；针对舌尖调音动作，用舌尖前伸（TTP）和舌尖上翘（TTR）两个调音参数来控制舌尖形状和动作，建立了舌尖的调音模型。

关键词：磁共振成像；汉语普通话；舌尖；调音模型

中图分类号：H 017 **文献标志码**：A

文章编号：1000-0054(2017)02-0158-06

DOI：10.16511/j.cnki.qhdxxb.2017.22.008

Modeling of the tongue tip in Standard Chinese using MRI

WANG Gaowu[1], DANG Jianwu[2], KONG Jiangping[3]

(1. School of Chinese Language and Literature, Beijing Normal University, Beijing 100875, China;
2. School of Computer Science and Technology, Tianjin University, Tianjin 300072, China;
3. Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

Abstract: The tongue tip motion in Standard Chinese was modeled based on articulatory data from magnetic resonance imaging (MRI) images. An MRI articulatory database was developed for Standard Chinese, including 9 vowels and 75 consonant variants. Principle component analysis (PCA) of the tongue shape was then used to find articulatory factors. The results show that the tongue should be divided as the tongue tip and tongue body and modeled separately for more precise results. The tongue tip motion is modeled with two articulatory parameters for tongue tip protrude and tongue tip raise which represent the protruding/advancing and raising/retroflexing movements of the tongue tip.

Key words: magnetic resonance imaging (MRI); Standard Chinese; tongue tip; articulatory model

为了更好地理解言语产生的过程，即如何从音位输入到调音器官的动作，再到声学输出，需要建立相应的声道模型。目前对于言语产生的声学建模，多采用把嗓音发声和声道调音分开的声源-滤波模型[1-2]。声带振动产生的声源信号经过声道的调制，再经过唇的辐射，成为语音的声波信号。在此理论基础上，针对发音器官的形状和运动，人们建立了很多声道模型，可以大致分为以下 3 类：面积参数模型、几何调音模型和生理调音模型。

面积参数模型直接以声管的截面积为变化参数，而不用关注各个发音器官的形状[3-4]。这类模型的缺点在于只关注声管面积的变化，难以跟生理结构和调音动作对应起来，也难以进一步研究发音器官各自的运动规律和相互之间的协调动作。生理调音模型则深入到生理层面，对发音器官的骨骼肌肉等组织进行建模，并关注肌肉结构和生理机能的限制（如体积守恒、组织形变）[5-6]，是最为深入的模型；但由于技术限制，目前尚缺乏建模所需的肌肉力量和肌电信号的经验数据。几何调音模型则介于这两类模型之间，把声道分解成各个调音器官，分别研究其形状变化和运动规律，但又不需要获取肌肉结构组织的生理数据，因此成为目前研究最多的模型。

在调音模型里，舌头是最为灵活的器官，在语音产生中的地位最为关键。因此，对其进行建模，找出其在发声过程中运动的基本规律，有利于理解协同发音和音系学现象的内在机制[7]。目前已经有很多对舌头建模的分析，有解析的、统计的，有 2 维的、3 维的，有实时的和非实时的[8-9]。

对舌头建模的难点在于，如何用最精简的参数

收稿日期：2016-06-23

基金项目：北京师范大学青年基金项目（SKXJS2014020）；
教育部社科青年基金项目（16YJC740065）

作者简介：汪高武（1970—），男，讲师。E-mail：wgw@bnu.edu.cn

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

去产生和控制灵活的舌头模型，因为从舌头的言语任务来看，只需要较少自由度便能发出各种语音[7]。舌面曲线虽然复杂，但有着系统的特定模式，需要关注如何去提取哪些参量来表征舌面形状。为了寻找这些参量，一般可从两个角度进行：先验和后验。先验是指预先指定一套参数来描述舌头的形状，例如舌体轮廓线圆心坐标及弧线、舌体最高点的坐标等，再用数据来训练模型参数[10-11]。后验是指用统计的方法分析舌头形状的数据，找到影响运动的因子[12]，把这些因子作为参量控制舌头的形状，常见的方法有平行因子分析(parallel factor analysis，PFA)[13]、主成分分析(principal component analysis，PCA)[14]、线性成分分析(linear component analysis，LCA)[9]等。

本文将结合先验和后验两种角度，先用主成分法分析舌头的形状，再参考得到的因子来设置舌体形状曲线的参数，建立相应的调音模型。以前的研究[13-14]都是把舌头作为一个整体分析，而本文发现把舌头分为舌尖和舌体两部分，可以使得控制参数更为精简，调音意义更为明显。

本文的研究目的在于最终建立汉语普通话声道调音模型。在理论方面，加深对汉语普通话的言语产生和调音生理特性的认识；在应用方面，模型对普通话的语言合成参数设定、虚拟说话人和普通话辅助教学，特别是聋哑儿童的可视化辅助教学，都有着实用价值。

1 汉语普通话磁共振成像发音数据库

磁共振成像(magnetic resonance imaging，MRI)是近来应用到声道研究领域的先进技术，已用于研究英语、法语、日语等多种语言，但目前还缺乏针对汉语普通话的系统研究[15-16]。为了进行汉语普通话声道生理和调音模型的研究，本文建立了相应的普通话声道 MRI 发音数据库。

实验中使用的 MRI 设备是 Shimadzu-Marconi ECLIPSE 1.5 T PowerDrive 250 Scanner。主要参数设置为：图像实际尺寸 256 mm×256 mm，分辨率 256×256 像素，每个像素点的大小为 1 mm^2。

语料选取了普通话中的 9 个单元音，例如 a、o、e、i、u 等；75 个辅音变体，例如 b(a)、b(i)、b(u)、b(o)、d(a)、d(i)、d(u)、d(e)等(先考虑所有声母后接单韵母的情况)。在拍摄声母的辅音变体时，要求发音人假想辅音后面接着元音，但这个元音不读出声音。例如b(a)，就让发音人假想要发出“疤”字，但并不发出元音的声响，保持成阻状态。本实验的发音人为一名成年男性。

图 1 所示为普通话元音“a”的 MRI 图像和处理后所得到的正中矢状面上的发音器官形状。其中舌头的边缘是用 14 个点插值生成的曲线。

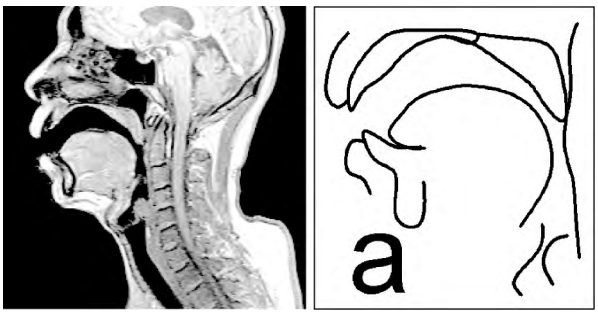
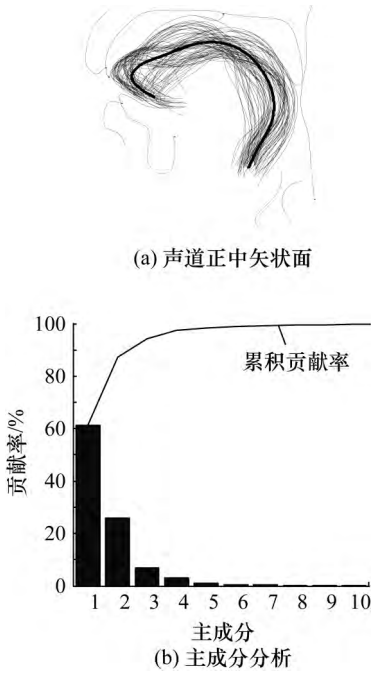


图 1 普通话元音“a”的 MRI 图像和处理后得到的发音器官形状

2 舌头形状的主成分分析

2.1 对舌头整体的因子分析

本文对 MRI 数据库的舌形数据进行主成分分析，得到的结果如图 2 所示。图 2a 是 MRI 的声道正中矢状面图。图 2b 则是主成分分析的贡献图，以柱状图的形式列出了各个主成分的贡献率(图 2b 中只画出了前 10 个主成分)。



(a) 声道正中矢状面

(b) 主成分分析

图 2 MRI 静态数据的舌形分布和前 10 个主成分的贡献率

本文用 14 个点来描述舌形，每个点有横纵 2 个坐标，数据空间为 28 维。为方便讨论，把每个点

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的坐标用复数表示，例如(123，200)转成 123＋200i，这样数据就转为复数空间中的 14 维，然后再进行主成分分析。由于舌形复杂，难以无损降维，因此得到 14 个主成分。各个成分的贡献率依次是：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%；累积贡献率是：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%，如表 1 所示。可见，用前 4 个主成分就可以描述 97.51%的舌形变化，平均误差为 1.0 mm。用前 6 个主成分就可以描述 98.99%的舌形变化，平均误差为 0.6 mm。仅从误差角度来看，若要求舌形误差不超过 1 mm(在数据图像中对应于 1 个像素点的大小)，则至少要用 4 个主成分来描述舌形。

图 3 是前 4 个主成分分量的变化图。以图 3a 为例，实线是舌形的平均值，是舌头的初始形状，也就是主成分 1 分量为 0 时舌头的形状。虚线是主成分 1 分量变化为 5 个单位，也即 5 mm(＋5 mm 和－5 mm)的时候舌头的形状，点线是变化为 10 mm(＋10 mm 和－10 mm)的时候舌头的形状，可以看出主成分 1 代表的是舌头的前后运动。同样，主成分 2 代表的主要是舌体不动、舌尖前伸的运动，主成分 3 代表的是舌根前移同时舌尖前伸并上翘的运动，主成分 4 代表是舌体拱起、舌尖上翘的运动。虽然本文已经尽量寻找这些主成分动作的调音意义，但有些地方仍不太尽如人意。例如，后 3 个主成分都有舌尖前伸的运动成分，这样改变主成分分量调节舌形的时候，就难以独立控制。本文希望各个主成分之间的相关性尽量小，对应的调音运动也比较独立和容易操作。

最后，用这 4 个主成分来重构舌形，并考察重构的误差。对误差要求最高的是塞音和塞擦音，因为对于成阻的地方，误差稍微大一些，重构的舌形就无法形成阻塞，就达不到语音学上区别对立的要求。这也是目前不少主成分研究不足的地方，这些研究只注意前几个主成分重构在统计上的误差大小，较少讨论这些误差分布在什么部位，对语音的区别有没有关键的影响。例如对于 p(i)、t(e)、z(a)、r(u)等，虽然重构舌形的整体误差不大，但不该成阻的地方与硬腭紧贴有了阻塞，而在该与硬腭形成阻塞的地方没有阻塞，这就说明用 4 个主成分来描述舌形还是不够。经过计算，只有用 6 个主成分才可以满足要求，使得重构的舌形能区分所有阻塞细微差别的音素，例如 b(i)和 p(i)、z(a)和 s(a)等。

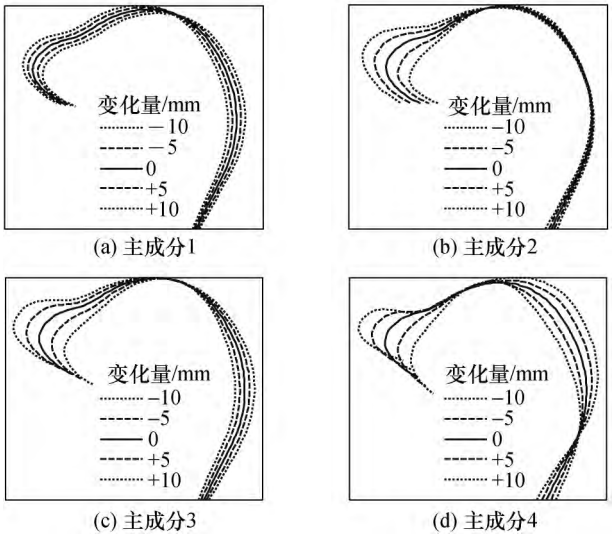


图 3 舌形的前 4 个主成分的运动趋势示意图

2.2 对舌尖和舌体分开的因子分析

用 6 个主成分描述舌形，可以满足区别普通话不同音素的要求，但为了简化模型，在控制重构舌形误差的前提下，应尽量减少主成分(也即控制参数)的数量。可以观察到，舌体和舌尖的运动具有相对的独立性。舌体是由颏舌肌、下纵肌、舌垂直肌、舌横肌和上纵肌组成，而舌尖主要是上纵肌延伸而成，两者的结构不同。舌尖附着在舌体上，运动特别灵活，很多音素特别是舌尖元音和辅音都是由舌尖的运动产生。这就提示把舌尖和舌体分开来分析可以简化参数。如图 4 所示，小圈所示为舌尖部分，大圈所示为舌体部分。

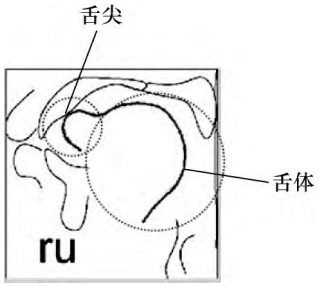
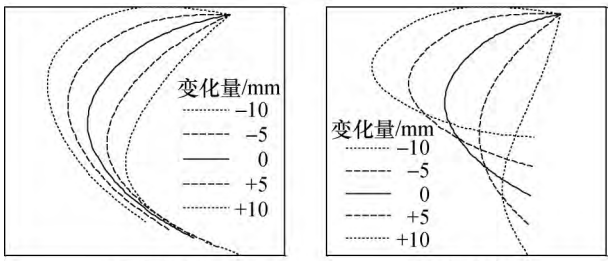


图 4 口腔示意图(把舌尖和舌体分开来分析)

首先对舌体进行主成分分析，得到各个成分的累积贡献率分别是：78.59%，90.40%，95.87%，98.55%，99.21%，99.55%，99.72%，99.86%，99.94%，100.00%，如表 1 所示。如果用前 3 个主成分来重构舌体，其平均误差为 1.2 mm。然后，再

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

对舌尖进行主成分分析，得到各个成分的累积贡献率分别是：92.66%，98.84%，99.69%，100.00%。如果用前 2 个主成分来重构舌体，平均误差为 0.3 mm。图 5 所示为舌尖运动的前两个主成分，可以看出，舌尖的这两个主成分的调音意义是很明显的：主成分 1 代表的是舌尖上翘，对应着翘舌音和卷舌音的动作；主成分 2 代表的是舌尖前突。



(a) 主成分1 (b) 主成分2

图 5 舌尖的前 2 个主成分的运动趋势示意图

下面根据这两个主成分重构舌尖，重点关注重构出的舌尖曲线能否满足同部位塞音与擦音的区分。图 6 是舌体重构误差最大的几个音位变体（限于篇幅，只列出 4 个辅音）。可以看到，对于舌尖前塞擦音 z，其重构的舌形能够成阻。对于舌尖前擦音 s，重构的舌形也能够留出通道，与 z 形成对立。舌尖后塞擦音 zh 和擦音 sh 也是如此。因此，用这 2 个主成分来描述舌尖是足够的。

图 6 舌尖重构误差最大的 4 个辅音的口腔示意图

本文把舌头分为舌体和舌尖两个部分，用 3 个主成分来描述舌体，用 2 个主成分来描述舌尖。这样做有 3 点进步：1）基于解剖生理，把舌头分为舌体和舌尖，使得各自主成分对应的动作具有更明确的调音意义。2）与用 6 个主成分重构舌形相比，用 3 个主成分重构舌体、2 个主成分重构舌尖，在保证能够区别各个音素变体的前提下，降低了参数的维度，使模型更加简洁。3）从表 1 可以看出，这样做减小了重点部位也就是舌尖的重构误差。

表 1 整个舌头、舌体、舌尖的主成分重构误差

整个舌头			舌体			舌尖		
主成分	累积贡献率/%	重构误差/mm	主成分	累积贡献率/%	重构误差/mm	主成分	累积贡献率/%	重构误差/mm
1	61.36	4.1	1	78.59	2.8	1	92.66	0.5
2	87.34	2.3	2	90.40	1.8	2	98.84	0.3
3	94.31	1.5	3	95.87	1.2	3	99.69	0.0
4	97.51	1.0	4	98.55	0.7	4	100.00	0.0
5	98.52	0.8	5	99.21	0.5	5	100.00	0.0
6	98.99	0.6	6	99.55	0.4			
7	99.31	0.5	7	99.72	0.3			
8	99.54	0.4	8	99.86	0.2			
9	99.68	0.4	9	99.94	0.1			
10	99.78	0.3	10	100.00	0.0			
11	99.86	0.2						
12	99.93	0.2						
13	99.97	0.1						
14	100.00	0.0						

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

3 舌尖的调音建模结果

从先验角度预定义舌头模型，往往用一段弧线或直线表示舌体或舌尖，舌形跟真实数据差距较大，而且往往缺少对舌尖下表面的描述。例如，把舌体建模为半径为 20 mm 的圆弧，把舌面和舌尖上表面建模为一条顺着舌体圆弧的弧形切线[10]。

本文先根据普通话发音数据的主成分分析，发现舌头分为舌体和舌尖两部分建模会更准确简洁，而且舌尖下表面也包含在内。对于舌尖部位，由于包含了舌尖上下表面，先初步选择用半边椭圆曲线来模拟舌尖的原始形状，较为直观方便。然后，参考主成分分析得到的因子(主成分 1 和主成分 2)，用两个更为直观的调音参数(舌尖前伸和舌尖上翘)来控制舌尖椭圆曲线。

设椭圆长轴 a，短轴 b，则舌尖边缘曲线方程为：

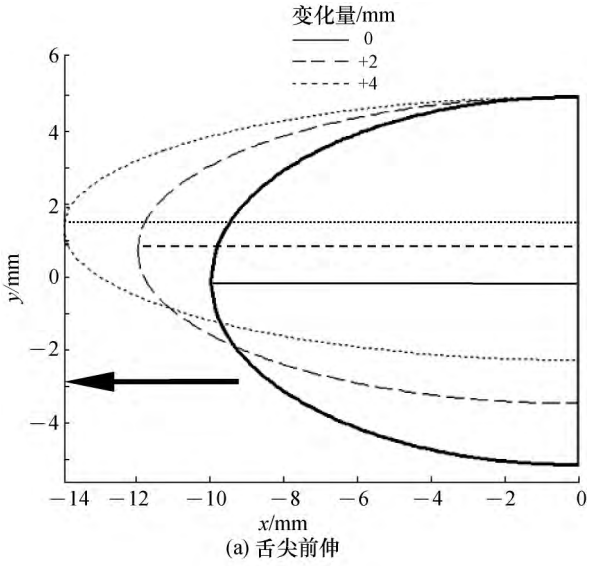
$$(x/a)^2+(y/b)^2=1,$$
$$-a<x<0,\ -b<y<b. \tag{1}$$

经测量，舌尖在休息状态下，其长度为 8～20 mm，厚度为 6～12 mm。这里不妨假设舌尖模型椭圆曲线的长轴 $a=10$ mm，短轴 $b=5$ mm。从生理上看，由于舌尖肌肉体积保持不变，因此假设在正中矢状面上舌尖的面积也保持不变，则方程(1)还要加上一个限制条件，即椭圆面积保持不变 $\pi ab=\pi\times 10\ \text{mm}\times 5\ \text{mm}=50\pi\ \text{mm}^2$。

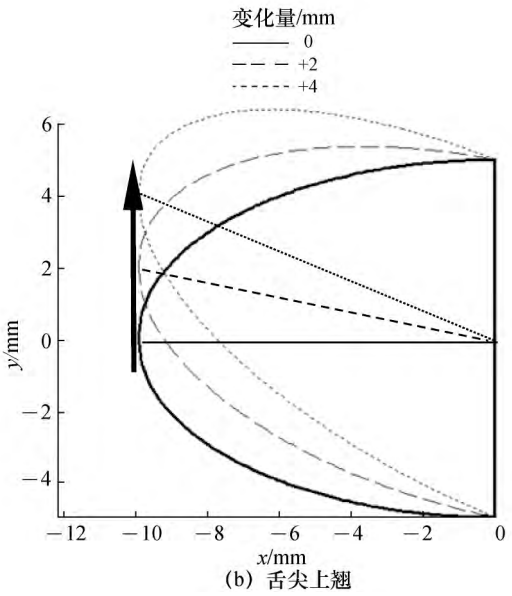
如图 7 所示，从身体左侧看过去，以舌尖椭圆曲线的中心为坐标原点，以 x 轴表示舌尖的长度，以 y 轴表示舌尖的厚度。设置舌尖前伸调音参数为 TTP (tongue tip protrude)，当其数值为正，则舌尖椭圆长轴增加，舌尖变扁前伸，负值则舌尖变厚缩回。舌尖上翘调音参数为 TTR (tongue tip raise)，若其数值为正，则椭圆变斜，舌尖上翘。根据祖暅原理(等幂等积定理)，舌尖中剖面面积不变。两个调音参数可以同时变化调节，使得舌尖前伸的同时也可以上翘，能够产生更大的舌尖运动范围。得到舌尖曲线的最终方程为

$$\left(\frac{x}{a+\text{TTP}}\right)^2+\left(\frac{y-\dfrac{\text{TTP}}{a+\text{TTP}}x-\dfrac{b\cdot\text{TTP}}{a+\text{TTP}}}{\dfrac{ab}{a+\text{TTP}}}\right)^2=1. \tag{2}$$

其中，$-a-\text{TTP}<x<0$。



(a) 舌尖前伸



(b) 舌尖上翘

图 7 正中矢状面上(从身体左侧看舌头)舌尖前伸、舌尖上翘参数对舌尖形状的影响

4 结 论

首先，本文发现把舌尖和舌体分开建模更为简洁，控制因子(调音参数)从 6 个降为 5 个(舌体 3 个，舌尖 2 个)，舌尖敏感部位的重构误差从 0.6 mm 降为 0.3 mm。其次，利用椭圆曲线初步建立了舌头模型的舌尖部分，由两个调音参数控制曲线方程。下一步的工作需要对舌体部分进行类似的建模，从而得到整个舌头的模型，以最终建立整个的汉语普通话声道调音模型，并根据整个舌头的重

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

构效果、声道的声学特性以及合成语音的质量来衡量建模质量，以改进或选择更好的曲线方程和控制参数。另外，本文使用的汉语普通话MRI数据库目前只有静态的发音数据，若要更加全面地建立和衡量发音模型，还需要补充动态的发音数据。

参考文献 (References)

[1] Fant G. Acoustic Theory of Speech Production [M]. 2nd Ed. Hague: Mouton, 1970: 328.

[2] Hardcastle W J, Laver J. The Handbook of Phonetic Sciences [M]. Oxford: Blackwell Publishing, 1999.

[3] Story B H. A parametric model of the vocal tract area function for vowel and consonant simulation [J]. *J Acoust Soc Am*, 2005, **117**(5): 3231-3254.

[4] Flanagan J. Speech Analysis Synthesis and Perception [M]. New York: Spinger, 1972.

[5] Wilhelms-Tricarico R. A biomechanical and physiologically-based vocal tract model and its control [J]. *J Phonetics*, 1996, **24**(1): 23-38.

[6] Dang J W, Honda K. Construction and control of a physiological articulatory model [J]. *J Acoust Soc Am*, 2004, **115**(2): 853-870.

[7] Iskarous K. Patterns of tongue movement [J]. *J Phonetics*, 2005, **33**(4): 363-381.

[8] Badin P, Bailly G, Reveret L, et al. Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images [J]. *J Phonetics*, 2002, **30**(3): 533-553.

[9] Engwall O. Combining MRI, EMA and EPG measurements in a three-dimensional tongue model [J]. *Speech Comm*, 2003, **41**(2/3): 303-329.

[10] Mermelstein P. Articulatory model for the study of speech production [J]. *J Acoust Soc Am*, 1973, **53**(4): 1070-1082.

[11] Coker C H. A model of articulatory dynamics and control [J]. *Proceedings of the IEEE*, 1976, **64**(4): 452-460.

[12] Lindblom B, Sundberg J. Acoustical consequences of lip, tongue, jaw, and larynx movement [J]. *J Acoust Soc Am*, 1971, **50**(4): 1166-1179.

[13] Harshman R, Ladefoged P, Goldstein L. Factor analysis of tongue shapes [J]. *J Acoust Soc Am*, 1977, **62**(3): 693-707.

[14] Beautemps D, Badin P, Bailly G. Linear degrees of freedom in speech production: Analysis of cineradio-and labio-film data and articulatory-acoustic modeling [J]. *J Acoust Soc Am*, 2001, **109**(5): 2165-2180.

[15] Wang G, Kitamura T, Lu X G, et al. MRI-based study of morphological and acoustical properties of Mandarin sustained steady vowels [J]. *J Signal Process*, 2008, **12**(4): 311-314.

[16] Wang Y, Wang H, Gao J, et al. Detailed morphological analysis of mandarin sustained steady vowels [C]// International Symposium on Chinese Spoken Language Processing (ISCSLP). Hong Kong, 2012: 413-416.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

ISSN 1000-0054 清华大学学报（自然科学版） 2017 年 第 57 卷 第 6 期 12/19
CN 11-2223/N J Tsinghua Univ (Sci & Technol), 2017, Vol. 57, No. 6 625-630

昆曲闺门旦颤音的嗓音特征

董 理[1]， 孔江平[2]

（1. 深圳大学 人文学院，深圳 518060；2. 北京大学 中国语言文学系，北京 100871）

摘 要：为揭示昆曲闺门旦演唱颤音的波动特征以及发声方式的变化规律，该文对 2 名专业的昆曲闺门旦演员演唱颤音时的喉头仪信号进行分析，从颤音周期、幅度、发声类型、生理动作等角度研究该行当颤音的特性。结果表明：闺门旦的颤音基频波动频率较低、幅度较小，颤音波动周期中伴随着一定程度的嗓音发声变化；在大部分颤音片段中，颤音接触商变化频率、颤音速度商变化频率与颤音基频变化频率一致，相位相同或相反，各参数周期变化的相位关系与嗓音类型有关；颤音的产生伴随着喉头的上下抖动。该研究揭示了闺门旦颤音的产生有其独特的发声规律。

关键词：闺门旦；颤音；喉头仪；基频；接触商；速度商

中图分类号：H017；J825 **文献标志码**：A

文章编号：1000-0054(2017)06-0625-06

DOI：10.16511/j.cnki.qhdxxb.2017.26.030

Phonatory characteristics of the vibrato voice in Young woman roles in Kunqu Opera

DONG Li[1], KONG Jiangping[2]

(1. School of Humanities, Shenzhen University, Shenzhen 518060, China;
2. Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

Abstract: The vibrato characteristics of two professional Kunqu Opera singers playing Young woman roles were measured using three EGG parameters for the fundamental frequency (F0), contact quotient (CQ) and speed quotient (SQ). The F0 in the Young woman role's singing showed a slow vibrato rate and a low vibrato amplitude as with their gentle voice. CQ and SQ changed periodically during vibrato singing. For most vibrato periods, the CQ and SQ vibrato rates were equal to the F0 vibrato rate. The vibrato phases of the three parameters showed no differences or a 180°difference, depending on the phonation type. The singers vibrate their larynx to producing the vibrato.

Key words: Young woman role; vibrato; electroglottography; fundamental frequency; contact quotient; speed quotient

在演唱时，纯粹的拉长会使声音听起来直而呆板，给人以紧张僵硬之感。若配以颤音，则可令声音松弛柔和，音色更圆润悦耳，达到余音袅袅，绕梁三日的效果，极大丰富了歌曲的表现色彩。

从物理上分析，颤音是一种周期性的、接近正弦波的发音频率的调节模式[1]。其调节模式的规律程度与歌手的演唱技巧高低有关，技巧越高的歌手，其颤音规律性越强。同时，对于受过训练的歌手而言，其颤音不受听觉反馈的影响，非常稳定[2]。歌手自身的颤音频率和幅度相对稳定，颤音特征可以通过训练得以改变，但比较困难。颤音的幅度和频率会影响声音的听感。一般而言，每秒小于 5.5 个波动的颤音让人感觉很慢，而大于 7.5 个波动的颤音让人觉得紧张。虽然歌手自身的颤音相对稳定，但其个数也可能会受演唱时情绪复杂变化的影响。演唱中颤音基频的变化幅度通常为±1 或±2 个半音，小于±0.5 个半音的波动常见于管乐器，而大于±2 个半音的波动不好听，这种波动常出现在老龄歌手的演唱中，其演唱特色为颤音基频变化频率低、幅度大。颤音的波动幅度变化通常伴随着音强的变化。颤音改变了基频，同时改变了第一共振峰附近谐波的频率。能量最强的谐波决定了音强，而这一谐波与共振峰的关系会影响颤音的音强与波动幅度的关系。如果能量最强的谐波在第一共振峰以下，音强与基频的变化呈正相关；如果在第一共振峰以上，二者则成反比。因为谐波在接近共振峰的过程中振幅会明显增大。但事实上，颤音振幅的变化对感知的影响不大[3]。

昆曲闺门旦行通常饰演名门闺秀或端庄少妇，务求音色含蓄柔和。演员演唱时使用了大量的颤

收稿日期：2016-05-05

基金项目：国家自然科学基金面上项目(61073085)；
中国哲学社会科学重大投标项目(10 & ZD125)

作者简介：董理(1986—)，女，讲师。
E-mail: dongli_pku@aliyun.com

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

音，声音的柔和颤动不仅存在于句末的长拖腔中，还融入到字里行间，使音色更加柔美隽永，感情饱满。听感上，闺门旦的颤音与美声唱法和通俗唱法的颤音差异很大，其基频波动的周期和幅度特点值得深入研究。同时，颤音可能不仅表现为基频的周期性变化，还可能伴随有发声方式或发声类型的转变。

闺门旦丰富的音色变化[4-5]和发声类型[6]为研究颤音的发声方式提供了不可多得的素材，为合成高自然度颤音奠定了基础。为了揭示昆曲闺门旦演唱颤音的波动特征以及发声方式的变化规律，本文对2名专业的昆曲闺门旦演员演唱颤音时的喉头仪信号进行分析，从颤音周期、幅度、发声类型、生理动作等角度来研究该行当颤音的特性。

1 实验方法

本文选了江苏省昆剧院的2名职业闺门旦演员为发音人，演员1和2录音时年龄分别为47岁和27岁，从业时间分别为27年和8年。演员较长的从业生涯保证了其稳定的音质能反映出行当的特性，2人年龄及从业时间的差异可以反映出两代闺门旦演员的发声差异，由此加强了实验的效度，在一定程度上避免了将个人特征作为行当的特征。录音在安静的普通房间完成，采集了语音和喉头仪(electroglottography，EGG)2路信号，采样频率为20 kHz，分辨率为16 b。2名演员分别演唱了一首南曲和一首北曲。2路信号中，语音信号包含了调音和发声的全部信息，喉头仪信号则仅反映出发声的特征。虽然也可以通过逆滤波技术从语音信号中得到声源信息，进而研究嗓音特征，但对于女性演唱嗓音来说，当前逆滤波技术还不足以实现将调音信息完全滤除，因此本文仅使用了喉头仪信号。

参数计算过程中，对喉头仪信号先进行了预处理，用二阶小波(dyadic wavelet)变换的方法分离出低频抖动，滤掉了高频噪声。鉴于闺门旦EGG信号较为复杂，本文在3种常用的确定EGG信号声门接触段的方法[7-10]中选定了稳定性最强的尺度法，以35%为尺度计算出基频(fundamental frequency，F0)、接触商(contact quotient，CQ)和速度商(speed quotient，SQ)3个参数。图1显示了如何用35%的尺度算法标记出周期、接触段、不接触段、关闭段和分离段，这对结果的分析非常重要。其中，CQ定义为接触段/周期×100%，SQ定义为分离段/关闭段×100%[11]。

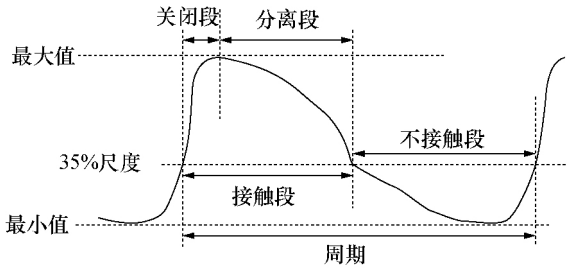


图1 35%的尺度算法

颤音的主要表现是基频的类正弦波波动，因此本文以基频为分析依据，将2首曲目中所有出现的基频类正弦波波动片段截取出来，进而分析闺门旦颤音的嗓音特征。考虑到半音与基频的非线性关系，所有基频数据被转换为半音表示，半音值“24”对应着基频220 Hz、A3音阶，半音值相差1代表实际音高相差一个半音。基于提取出的原始嗓音参数，采用SPSS对各参数的分布情况进行统计。为了更好地反映嗓音参数的周期变化规律，对嗓音参数结果进行了低通滤波(局部平滑)，平滑后计算出F0、CQ和SQ的波动周期和幅度，并用SPSS18对以上数据的分布情况进行统计。采用若干术语描述颤音的特性：颤音基频变化频率(vibrato frequency of F0，F0_VF)代表颤音片段中F0值随时间做类正弦波变化的频率；颤音基频变化幅度(vibrato amplitude of F0，F0_VA)代表F0的类正弦波变化的幅度，与正弦波的振幅相类似；颤音接触商变化频率(CQ_VF)、颤音接触商变化幅度(CQ_VA)、颤音速度商变化频率(SQ_VF)和颤音速度商变化幅度(SQ_VA)分别代表颤音片段中CQ和SQ随时间发生类正弦波变化的频率和幅度。

2 实验结果

2.1 颤音段嗓音参数的变化频率

图2显示了颤音的F0、CQ和SQ的变化频率，即1 s内进行类正弦波变化的次数。图2中参数后的数字对应演员的编号，如“F0_VF1”表示闺门旦1的F0在1 s内颤动的次数。图2中用箱线图的形式显示了各数据的分布情况，其中箱中的十字对应中位数，箱的两端对应四分之一和四分之三位点(25%和75%数据的分布位置)，线的两端对应10%和90%的分布位置。可以看出，2位演员F0_VF的分布都比较集中，波动频率在3.5～7次/s之间，数据主要分布在4.5～6次/s之间，中位数分别为5.1或5.2次/s。参照已有的感知研究[3]，

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

昆曲闺门旦的 F0_VF 大部分都分布在让人感觉很慢的范围内。CQ_VF 和 SQ_VF 的分布比 F0_VF 分散，有 75%左右的数据与 F0_VF 相重叠，还存在 25%以上的数据比 F0_VF 大。可以推测，在大部分颤音片段中，CQ 和 SQ 的变化与 F0 的变化是同步的，在小部分情况下，二者与 F0 的变化是不同步的。与 F0 相比，CQ 和 SQ 的周期性波动是比较同步的。

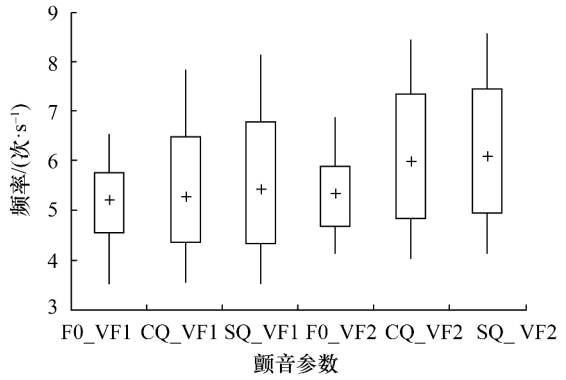


图 2　颤音 F0、CQ 和 SQ 变化频率

2.2　颤音段嗓音参数的分布

衡量嗓音参数的变化幅度大小，需要先考虑嗓音参数的分布范围。通过比较发现，颤音片段和整首曲目的嗓音参数差异很小。图 3 为不同曲目颤音片段的 F0、CQ 和 SQ 的分布情况。其中，箱中的十字对应中位数，箱的两端对应四分之一和四分之三位点，线的两端对应 10%和 90%的分布位置。比较同一演员的 2 首曲目可以发现，颤音区域的 F0 差异不大，F0 基本分布在 D4-C5 之间，主要集中在 E4-A♯4 之间。闺门旦 2 的 CQ 分布相对集中，基本分布在 34%～52%之间，SQ 主要分布在 70%～180%之间；2 首曲目的 CQ 和 SQ 略有差异，北曲的 CQ 略小、SQ 略大。闺门旦 1 的 CQ 分布范围比闺门旦 2 要宽得多，变化范围在 26%～66%之间；其北曲的 CQ 和 SQ 与其他曲目的差异较大，其 CQ 分布范围明显大于其他 3 首曲目，SQ 的值域明显低于其他 3 首曲目。这种差异是由闺门旦复杂的嗓音类型决定的。

2.3　颤音段嗓音参数的变化幅度

表 1 显示了颤音的 F0、CQ 和 SQ 的波动幅度，分别给出了 25%、50%和 75%的数据分布的位置。F0 的波动幅度大部分都在 0.5 个半音以上，同一演员演唱南北曲的差异不大。与西洋美声唱法的颤音相比，闺门旦的颤音 F0 波动幅度较小，有一半

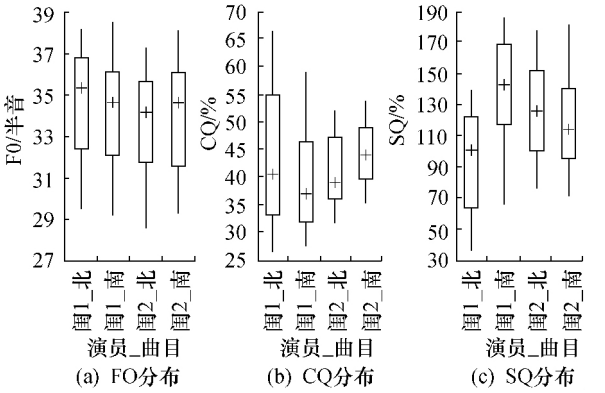


图 3　不同曲目颤音片段的 F0、CQ 和 SQ 的分布

颤音的波动幅度都小于 1 个半音，闺门旦 1 在演唱南曲时甚至有四分之一的颤音波动幅度小于 0.5 个半音，这在演唱中是很少见的。较小的 F0 波动幅度可能与保持柔和的音色相关。闺门旦 1 的颤音 F0 波动幅度小于闺门旦 2，这与闺门旦 1 声音更柔和是一致的。从 CQ 的波动幅度来看，2 名演员 CQ 的波动幅度都不大，这意味着在一个颤音周期内，发声类型并没有发生很大的变化，仅做了有限的调整。SQ 的波动幅度也较小，再次印证了这一推测。由此可见，闺门旦的颤音并不是发声类型的剧烈转换形成的，而是通过周期性的细微调节 F0 及嗓音的松紧程度来实现的。

表 1　颤音 F0、CQ 和 SQ 的波动幅度

演员_曲目	位点	F0_VA / 半音	CQ_VA / %	SQ_VA / %
闺 1_北	1/4	0.6	1.5	5.6
	1/2	0.7	2.7	13.2
	3/4	0.9	7.3	24.6
闺 1_南	1/4	0.5	1.3	5.7
	1/2	0.7	2.5	12.5
	3/4	0.9	4.7	27.1
闺 2_北	1/4	0.8	1	8.4
	1/2	1	2.1	15.6
	3/4	1.2	3.9	23.8
闺 2_南	1/4	0.8	1.4	9.8
	1/2	1	2.4	15.6
	3/4	1.2	3.7	25.6

2.4　颤音段嗓音参数变化的相位关系

从图 2 中可以看出，大部分的颤音片段中 F0__

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

VF、CQ_VF 和 SQ_VF 是一致的，但其相位关系还有待探究。图 4-7 显示了颤音片段中 4 种典型的 F0、CQ 和 SQ 波动的相位关系。图 4 和 5 中 F0 与 CQ 的相位相同，二者与 SQ 的相位相反，即相差半个周期；图 6 中 CQ 与 SQ 的相位相同，二者与 F0 的相位相反；图 7 中 F0、CQ 和 SQ 三者相位相同。从图 7 的第一个 F0 变化周期与 CQ 和 SQ 的对应关系还可以发现，存在 CQ_VF 和 SQ_VF 大于 F0_VF 的情况，此种情况下，CQ_VF 和 SQ_VF 是相同的，这与图 2 中观察到的分布也是一致的。由此可见，3 个参数的变化基本上是正相或反相关系，几乎不存在正交关系。

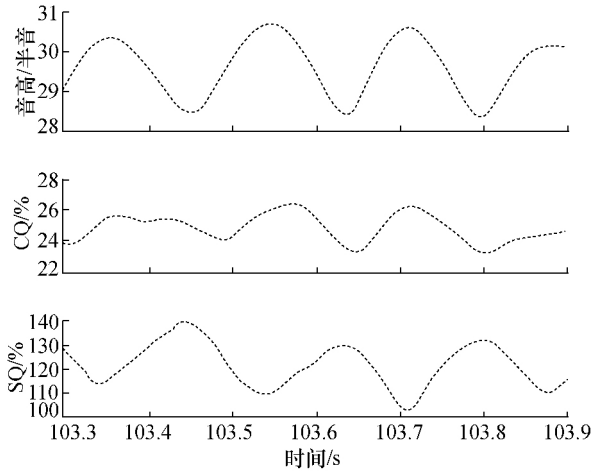


图 4 颤音中 F0、CQ 和 SQ 的相位关系 A

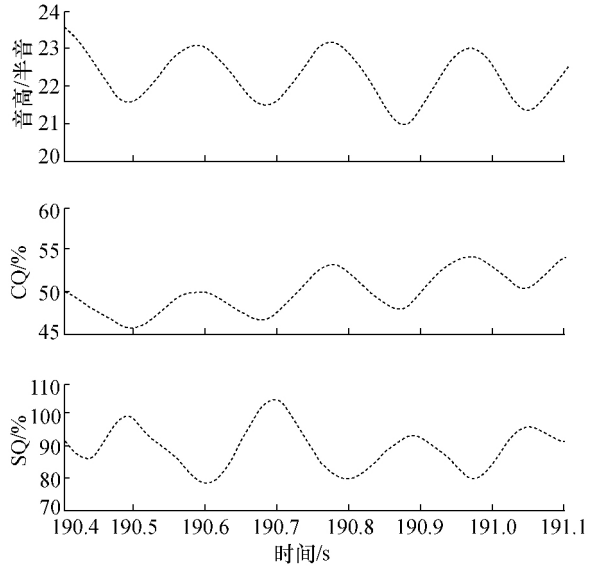


图 5 颤音中 F0、CQ 和 SQ 的相位关系 B

3 讨 论

为了准确解释颤音的嗓音变化情况，必须首先了解闺门旦嗓音的复杂变化情况。图 8 截取了 2 个极端的 EGG 例子，分别显示了声门关闭段和分离段出现双峰的情况，本文暂且将这 2 段出现双峰的位置分别称为第一关闭段和第二分离段，如图 8 中虚线方框所示。第一关闭段和第二分离段的变化会明显影响到 CQ 和 SQ 的结果。

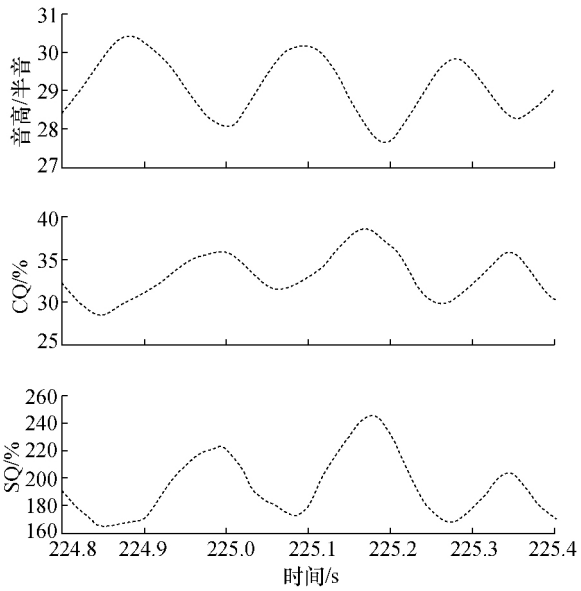


图 6 颤音中 F0、CQ 和 SQ 的相位关系 C

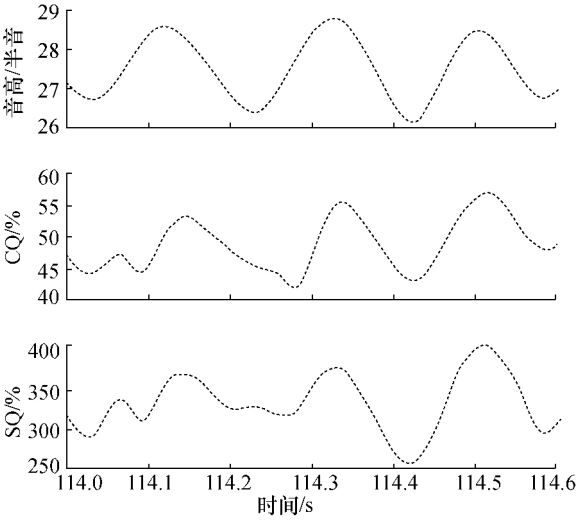


图 7 颤音中 F0、CQ 和 SQ 的相位关系 D

在正常嗓音中，第一关闭段对应着声带下部的接触，在闺门旦的演唱中，此段变化可以分为 3 种情况：1）当第一关闭段出现双峰时声门关闭段的时长增加，其占接触段的比例也大大增加，因此使 SQ 有很大一部分值落在 100%以下，双峰越显著，CQ 越大、SQ 越小，此种情况下计算出的嗓音参数特点是 CQ 较大、SQ 远小于 100%；2）第一关闭段内 EGG 信号上升斜率大且不出现双峰，如图 8a 方

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

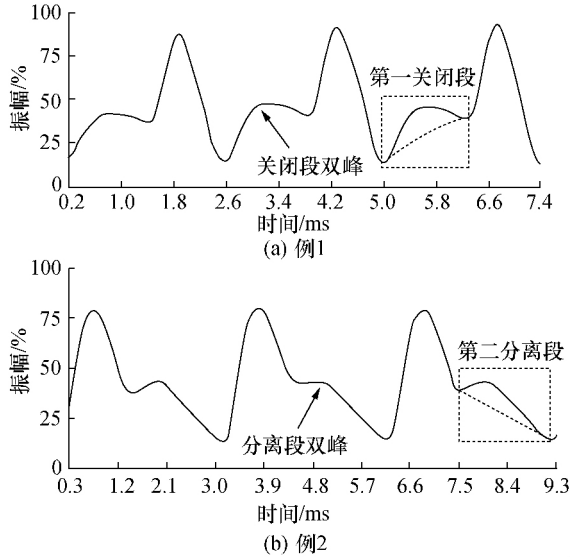


图 8　双峰信号示例

框内虚线所示，此时 EGG 上升的斜率越大，CQ 越大、SQ 越小，此种情况下 CQ 略小于前一种情况，SQ 比前一种情况更接近 100%；3）EGG 此段的波形与前 2 种情况相似，即上升或出现双峰，但 EGG 信号的最大能量小于 35%的尺度，此时该段被作为不接触段处理，计算得出的 CQ 较小，SQ 通常在 100%以上，大部分情况下都不会超过 250%。第一关闭段的产生是声门后部收紧一种表现[12]，CQ 越大，SQ 越小，表明声门后部收紧程度越大。

在正常嗓音中，第二分离段对应着声带上部的分离，此段同样可以分为 3 种情况：1）当第二分离段出现双峰时，会导致分离段的时长显著增加，其占接触段的比例也相应增大，由此计算出的 CQ 和 SQ 都比较大，SQ 明显大于 100%，甚至会达到 500%以上，双峰越明显，CQ 和 SQ 都越大，此种情况也是声门后部收紧度增加的一种表现[12]；2）第二分离段内 EGG 信号下降斜率大且不出现双峰（如图 8b 方框内虚线所示），此种情况下，EGG 下降的斜率越大，CQ 和 SQ 越大，CQ 和 SQ 均小于前种情况；3）EGG 此段的波形与前 2 种情况相似，但 EGG 信号的最大能量小于 35%的尺度，此时该段也被作为不接触段处理，CQ 和 SQ 都明显小于前 2 种情况，对应的声门后部收紧程度也小于前 2 种情况。

将 EGG 信号的特征分析与图 3 的结果结合起来可以发现，闺门旦 1 的北曲演唱中第一关闭段出现情况 1 和 2 的比例较高，在其南曲演唱中，此种发声方式较少，造成了 2 首曲子 CQ 和 SQ 分布的显著差异。在闺门旦 2 的 EGG 信号中，第一关闭段也会出现情况 1 和 2，但其此种现象明显少于闺门旦 1。从 SQ 的分布还可以发现，90%以上的数据都分布在 200%以下，可以看出，分离段双峰的情况在闺门旦的演唱中所占比例较小。

考虑到不同的 CQ 和 SQ 所反映的声门振动方式截然不同，在探讨颤音周期内的嗓音变化时需要分别进行讨论。SQ 明显大于 100%时，第二分离段的变化对 SQ 的变化起主要作用；当 SQ 小于 100%时，第一关闭段的变化对 SQ 的变化起主要作用；当 SQ 接近 100%时，它对关闭段和分离段的变化都比较敏感。图 4 对应的是第一关闭段的第 3 种情况，即 EGG 信号上升或出现双峰，但最大能量小于 35%的尺度。图 5 对应着第一关闭段的第 1 和第 2 种情况，即声门接触始于第一关闭段的上升区域或双峰位置。图 6 对应着第二分离段的第 3 种情况以及第 1 种情况或第 2 种情况与第 3 种情况交替出现的情形。图 7 对应着第 1 种情况和第 2 种情况单独出现或交替出现的情形。结合前文对 EGG 信号形状的分析可以发现，图 4、5 和 7 都是在保持声门后部收紧的情况下产生的，F0 的提高伴随着声门后部内收程度的增加。图 6 对应的情况正好相反，在 F0 上升时，声门后部的收紧程度降低了，一种可能的解释是：F0 上升时，声带被拉得更长，声带张力增加，在声门后部内收程度不强的情况下，可能会降低声门后部的内收程度，加快声带分离的速度，进而导致 CQ 和 SQ 都下降。在进行颤音合成时，要结合发声动作与发声类型来确定各参数的相位情况。

以上的讨论表明：颤音不仅表现为 F0 的周期性调节，还伴随着声带振动方式的周期性改变。关于颤音的成因，可以推测如下：从颤音的出现位置看，不仅在长拖腔中可以观察到连续出现的多个周期的颤音，也可以在短的音高片段中、在起音之后紧跟着一个颤音。颤音可能是喉部的特殊状态，不用刻意去启动，条件一旦允许，如同一音高持续超过一定时长，如 3 倍于颤音周期或是更短的时间，就会触发颤音的产生；联系前文关于声门后部收紧程度的周期性变化，颤音的产生可能是喉部肌肉，尤其是控制勺状软骨的若干肌肉周期性放松的结果。此外，闺门旦在发颤音时还伴随着喉头的上下抖动（见图 9）。喉头的运动对声带的张力也有一定影响，可能是嗓音松紧变化的成因之一。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

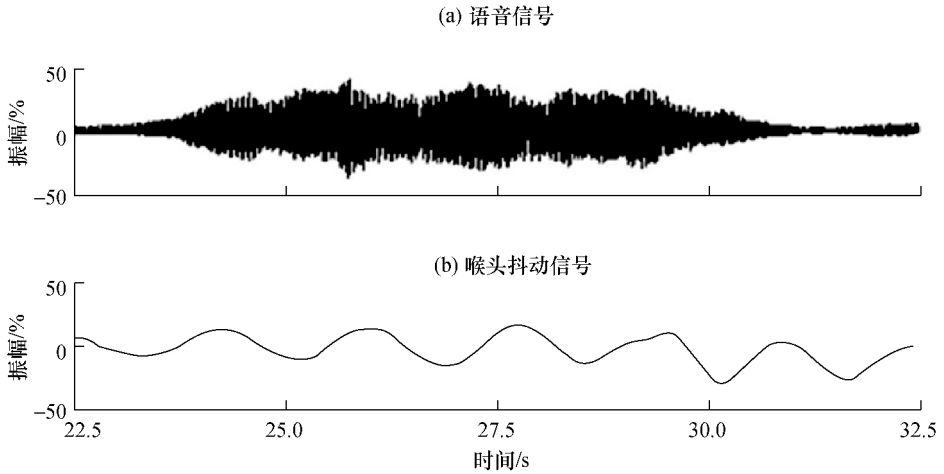


图 9 颤音对应的喉头抖动

4 结 论

本文对闺门旦演员演唱颤音时的喉头仪参数进行了分析，得出如下结论：闺门旦的颤音 F0 波动频率较低、幅度较小，与其柔和的音色相吻合；颤音不仅表现为 F0 的周期性波动，还伴随着嗓音松紧的周期性变化；颤音波动周期中嗓音松紧调节的幅度较小，并未伴随发声类型的剧烈转换；在大部分颤音片段中，CQ、SQ 和 F0 的变化周期一致，相位相同或相反；各参数周期变化的相位关系与嗓音类型有关；闺门旦的颤音伴随着喉头的上下抖动。

参考文献 (References)

[1] Schultz-Coulon H J, Battmer R D. Quantitative evaluation of the vibrato of singers [J]. *Folia Phoniatrica*, 1981, **33**(1): 1-14.

[2] Schultz-Coulon H J. The neuromuscular phonatory control system and vocal function [J]. *Acta Oto-laryngologica*, 1978, **86**(1-6): 142-153.

[3] Sundberg J. The Science of the Singing Voice [M]. DeKalb: Northern Illinois University Press, 1987.

[4] Dong L, Kong J P, Sundberg J. Long-term-average spectrum characteristics of Kunqu Opera singers' speaking, singing and stage speech[J]. *Logopedics Phoniatrics Vocology*, 2014, **39**(2): 72-80.

[5] Dong L, Sundberg J, Kong J P. Loudness and pitch of Kunqu Opera[J]. *Journal of Voice*, 2014, **28**: 14-19.

[6] **董理. 昆曲闺门旦行嗓音特征探究**[J]. **语言学论丛**, 2016, **2**: 292-304.
DONG Li. Voice characteristics of Young woman role in Kunqu Opera [J]. *Essays on Linguistics*, 2016, **2**: 292-304. (in Chinese)

[7] Childers D G, Hicks D M, Moore G P, et al. A model for vocal fold vibratory motion, contact area, and the electroglottogram[J]. *Acoustical Society of America Journal*, 1986, **80** (5): 1309-1320.

[8] Henrich N, D'Alessandro C, Doval B, et al. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation[J]. *Acoustical Society of America Journal*, 2004, **115**(3): 1321-1332.

[9] Rothenberg M, Mahshie J J. Monitoring vocal fold abduction through vocal fold contact area [J]. *Journal of Speech and Hearing Research*, 1988, **31**: 338-351.

[10] Howard D M, Lindsey G A, Allen B. Toward the quantification of vocal efficiency[J]. *Journal of Voice*, 1990, **4**: 205-212.

[11] Dong L, Sundberg J, Kong J P. Formant and voice source properties in two male Kunqu Opera roles: A pilot study[J]. *Folia Phoniatrica et Logopaedica*, 2014, **65**: 294-302.

[12] Herbst C T, Howard D, Schlömicher-Their J. Using electroglottographic real-time feedback to control posterior glottal adduction during phonation [J]. *Journal of Voice*, 2010, **24**(1): 72-85.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

ISSN 1000-0054 清华大学学报（自然科学版） 2017 年 第 57 卷 第 9 期 9/19
CN 11-2223/N J Tsinghua Univ (Sci & Technol), 2017, Vol. 57, No. 9 945-951

矢量半径驱动的汉语普通话立体声道模型

姚 云[1]， 吴西愉[2]， 孔江平[2]

（1. 河南大学 文学院，开封 475001；2. 北京大学 中国语言文学系，北京 100871）

摘 要：为得到更加精确的声道共鸣特性，人们越来越关心说话时声道的结构及其形状变化。该文从 MRI 3-D 图像数据中提取了汉语普通话 7 个单元音[a]、[o]、[ɤ]、[i]、[u]、[y]、[e]的声道边缘、中线和矢量半径数据，并沿着声道中线从唇到声门等间距截取声道的 36 个截面，对于每个截面，根据矢量半径的大小使同一个位置的截面形状作线性过渡，从而建立用矢量半径驱动的汉语普通话立体声道模型。计算模型的共振峰并合成语音样本，在与自然语音对比的听辨实验中，该模型取得了较好的语音合成效果。

关键词：汉语普通话；声道建模；矢量半径

中图分类号：H017 **文献标志码**：A

文章编号：1000-0054(2017)09-0945-07

DOI：10.16511/j.cnki.qhdxxb.2017.26.045

Radius vector-driven 3-D Mandarin vocal tract model

YAO Yun[1], WU Xiyu[2], KONG Jiangping[2]

(1. College of Chinese Language and Literature, Henan University, Kaifeng 475001, China;
2. Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

Abstract: Analyses of the vocal tract resonant characteristics need accurate models of the vocal tract shape. This article presents a three-dimensional Mandarin vocal tract model using vocal tract shape data and midsagittal radius vector data from MRI images for seven Mandarin sustained vowels [a], [o], [ɤ], [i], [u], [y] and [e]. The vocal tracts images were cut into 36 sections of equal distances along the midline of the vocal tract. The Mandarin vocal tract model for each section is then driven by the length of the radius vector in the cross-sectional images. The sound synthesized by this model sounds very much like natural speech.

Key words: Mandarin; vocal tract model; radius vector

声道形状决定了语音的共鸣特性（音色），尝试建立人类声道的模型并模拟人自然说话成为众多科学家和工程技术专家不懈追求的目标。18 世纪末 Kratzenstein 和 Von Kempelen 设计的模型已经能够准确模拟人类语言中几个最基本的单元音[1]，然而受当时技术水平的限制，人们很难获得精确的声道数据合成语音。文[2-3]将声道抽象为若干段横截面面积不同的均匀声管，从研究这种理想的均匀腔体出发来研究人类的声道。文[4-5]利用 α-β 模型推测声道截面面积，计算声道的传递函数，利用有限元方法估计声道的声学性质。

使用 α-β 模型，需要采集真实的声道中央矢状面并提取矢量半径来提高精度，众多学者利用医学成像设备获取发音过程中的声道形状。起初 X 光断层扫描技术被广泛应用，Chiba 等[6]利用 X 光照相技术获得口腔若干截面信息，研究元音的共振峰参数。由于 X 光对身体有害，不能大剂量长时间地多次扫描，渐渐被磁共振成像（magnetic resonance imaging，MRI）所代替。Baer 等[7]用 MRI 数据重建声道形状，获得声道面积函数，计算元音共振峰，并研究声道不同位置的矢状面宽度和截面面积之间的关系。文[8-11]利用 MRI 数据进一步分析了元音、辅音以及卷舌音。Story 等[12-13]建立了声道面积正交系数与元音共振峰之间的映射关系，并构造了一种基于声道面积函数的立体声道模型，可以动态地演示声道面积的变化。Dang 等[14-15]建立了 2.5 维生理发音模型并利用语音共振峰 F1、F2 的（对数）差异与舌面位置在日语中成负相关关系这一特点，用语音来逆推声道形状。

立体 MRI 声道数据显示，发音时口腔的腔体结构不只是单一的 1 个管道，还包括梨状窝、牙缝、喉管、舌下腔等这些随发音而变化的小腔体。Dang 等[16-18]发现鼻腔旁支的不对称性在共振峰的谱包

收稿日期：2016-05-06
基金项目：中国哲学社会科学重大投标项目（10&ZD125）
作者简介：姚云（1982—），男，讲师。
通信作者：孔江平，教授，E-mail：jpkong@pku.edu.cn

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

络上会产生额外的零极点对，元音由[i]到[a]动态调音过程中，牙缝会影响高次共振峰的位置并导致共振峰的谱包络中出现零点。以上研究表明：发音模型研究应该考虑鼻腔旁支、梨状窝和牙缝等声道旁支对共振峰的影响。

汉语的声道模型研究始于上世纪中叶，但发展缓慢，直到近期才逐渐得到学界的重视。利用 α-β 模型合成语音只考虑声道面积而忽略声道形状，因此很难精确计算出声道的共鸣特性。为了更好地研究人类声道特性及其对语音的影响，建立一个形状更加精确的声道模型十分必要。此外，鉴于目前磁共振设备已经能够实时采集声道中央矢状面的数据，用中央矢状面的数据驱动立体声道模型已成为可能，因此该文尝试建立保留声道形状细节特征的汉语普通话静态立体声道模型，并用截面半径参数驱动该声道模型，使该模型能够较好地模拟言语发音时的声道变化情况。

1 立体声道数据与建模

1.1 实验设计

本文使用的 MRI 图像是在日本国际电气通信基础技术研究所脑活动成像中心扫描获得。磁共振设备是 Shimadzu-Marconi ECLIPSE 1.5 T Power-Drive 250 scanner，设定扫描区域为 256 mm×256 mm 的正方形区域，每 1.5 mm 厚度生成一张图像，图像间距为 1.5 mm，用同步采样法获得 MRI 图像，所成图像为 DICOM 格式，图像高 512 像素，宽 512 像素，详细参数见表 1。录制的元音包括汉语普通话 7 个单元音[a]、[o]、[ɤ]、[i]、[u]、[y]、[e]。发音人年龄在 50 岁左右，普通话纯正，发音时固定发音姿态，每遍持续 3 s，中间有 1 s 呼吸时间。重复发音时，口腔形状尽量保持不变，必要时可用嘴呼吸来减少对软腭的影响。

由于 MRI 不能准确显示牙齿的形状，为了得到完整的声道形状，应把牙齿部分补充完整。具体做法是让发音人口含一种含铁量高的果汁或者果冻，确保口中的果冻把口腔中的空间都填满，不留气泡，并且上下牙齿不能接触，再进行磁共振成像，这样就可以得到牙齿边缘的轮廓。把牙齿的图像取出，用图像处理软件平移旋转之后叠加到原始的图像上。经过上述处理后的图像，声道轮廓就变得更加准确、清晰。

表 1 磁共振设备扫描参数列表

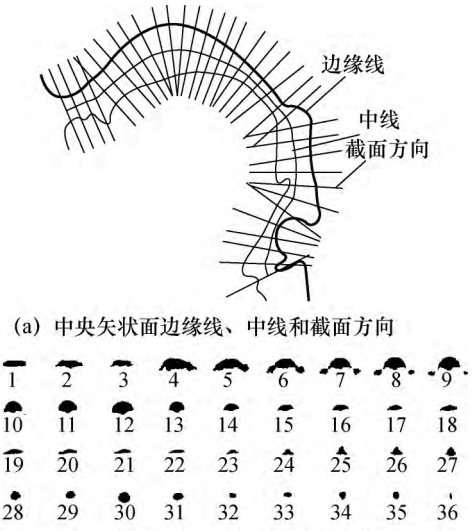
参数	数值
回波时间/ms	3.4
弛豫时间/ms	10
切片数目	44～51 矢状截面
切片厚度/mm	1.5
切片间隔/mm	1.5
视域/mm×mm	256×256
图片尺寸	512 像素×512 像素
图像数据格式	DICOM

1.2 实验数据与建模

为提取声道形状数据，需要对图像做进一步处理，只保留 MRI 图像中的声道部分，并沿声道中线，在垂直于中线的方向上截出 36 个截面，同时记录垂线与中线的交点坐标以及垂线与水平面的夹角，随后搜索并记录每个截面的边缘点，并对边缘点按顺时针方向排序，边缘点的第一个点定在图像中心正上方的位置。由于每张图的边缘线点数不同，将边缘线上的点归一化，对每一个截面只采集边缘上 36 个点，这样既能很好地显示截面的形状，又能比较好地建立模型。图 1 是汉语普通话单元音[a]的声道截面数据图。图 1a 中粗线是声道中央矢状面边缘线，声道中间灰色的线条是从唇中间到声门的声道中线，垂直于声道中线的细线代表声道截面方向；图 1b 展示了从唇到声门等间距截取声道的 36 个截面，截面 7～9 下方的 2 个黑点是被上下牙、舌头和脸颊封闭起来的牙缝。硬腭是一个圆形穹顶结构，加上舌面中间有条凹槽，所以截面 6～16 的主体呈现扇面结构。整体上，汉语普通话元音[a]的声道口腔区域的截面面积较大，咽腔位置的截面面积较小。

成年人的声道是一个拐角接近直角的复杂腔体。在汉语普通话的 7 个单元音中，牙缝的形状也会随着元音的不同而改变，并且整个声道的形状差别巨大。图 2 是采集 MRI 3-D 声道边缘点重建出的汉语普通话单元音[a]、[i]、[u]立体声道模型，图中每个模型右侧彼此分离的部分是上下唇的表面，向左分别是口腔、咽腔，向下连接着喉管和气管。当发元音[a]时，开口度较大，因此声道前腔大、后腔小且牙缝相对较宽，而且前端与声道主体连接在一起，在声道下端，舌根和会厌软骨距离较近，因此元音[a]的声道模型垂直段腔体粗细较均匀，过渡平稳；发元音[i]时，开口度较小，声道前腔变得

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

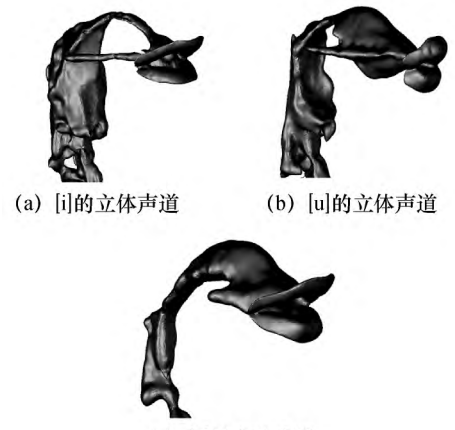


(a) 中央矢状面边缘线、中线和截面方向

(b) 沿截面方向所得的36个截面

图 1　汉语普通话元音[a]的声道截面

狭窄，后腔较大，牙缝变成了细长的条状，只在接近唇的位置与声道主体相连，在声道下端会厌软骨与舌根彼此分离，在舌根处形成一个明显的凹槽，因此元音[i]的声道模型咽腔段比元音[a]粗，且下端向右突出；发元音[u]时，嘴唇突起，声道前腔和后腔都比较大，中间部分狭窄，牙缝细长并且有较长的部分与声道主体相连，会厌软骨与舌根彼此分离，声道模型在舌根处形成一个明显的凹槽，但又与元音[i]的声道模型不同。



(a) [i]的立体声道　(b) [u]的立体声道

(c) [a]的立体声道

图 2　汉语普通话单元音[a]、[i]、[u]的立体声道

图 2 为建立矢量半径驱动的立体声道模型提供参照，也是建立动态声道模型的前提和基础。建立矢量半径驱动的立体声道模型的关键环节是获得声道截面的矢量半径和截面边缘线数据。

表 2 列出了汉语普通话 7 个单元音从唇到声门的 36 个截面的矢量半径的长度。36 个截面足以区分不同的平舌单元音，不会存在 36 个矢量半径相同而元音不同的情况，这就为实现用矢量半径驱动立体声道提供可能，至少可以用这种方法来预测汉语普通话平舌元音的声道形状。

表 2　汉语普通话声道从唇到声门 36 个截面矢量半径

单位：cm

截面序号	[a]	[o]	[ɤ]	[i]	[u]	[y]	[e]
1	0.90	0.70	0.85	0.85	0.25	0.20	0.80
2	1.05	0.60	0.80	0.65	0.25	0.20	0.65
3	0.80	0.80	0.70	1.05	0.45	0.45	1.00
4	1.40	1.45	1.60	0.90	0.90	0.90	0.80
5	1.25	1.65	1.50	0.45	1.35	0.75	0.70
6	1.30	1.90	1.40	0.35	2.05	0.55	0.45
7	1.55	1.80	1.70	0.30	1.75	0.35	0.40
8	1.75	2.15	1.80	0.35	1.65	0.30	0.40
9	1.90	2.25	1.85	0.35	1.65	0.30	0.55
10	1.60	1.65	1.55	0.30	1.60	0.35	0.30
11	1.80	1.50	1.10	0.20	1.10	0.20	0.25
12	1.95	1.50	1.50	0.30	0.85	0.40	0.40
13	1.45	1.50	1.40	0.25	0.80	0.45	0.40
14	1.00	1.90	0.85	0.25	0.85	0.35	0.25
15	0.85	1.10	0.75	0.40	0.45	0.50	0.30
16	0.85	1.00	0.50	0.55	0.40	0.55	0.55
17	0.70	0.80	0.50	0.90	0.30	0.50	0.75
18	0.75	0.55	0.40	1.30	0.30	1.05	1.15
19	0.70	0.50	0.35	1.45	0.20	2.25	1.40
20	0.55	0.50	0.55	2.15	0.55	2.60	1.70
21	0.55	0.40	0.70	2.85	0.85	2.75	2.00
22	0.45	0.30	0.60	2.55	0.75	2.65	1.80
23	0.50	0.75	1.05	2.75	0.95	2.75	1.95
24	0.80	0.80	1.15	2.70	1.40	2.75	1.95
25	1.10	0.65	1.10	2.65	1.10	2.70	1.75
26	1.15	0.80	1.20	2.65	1.25	2.55	1.65
27	1.20	0.90	1.50	2.75	1.45	2.55	1.80
28	1.20	1.15	1.30	2.75	1.90	2.65	1.90
29	1.00	0.90	1.15	2.95	2.35	2.80	1.30
30	1.50	1.25	1.60	2.20	1.55	1.80	1.40
31	0.95	1.45	2.00	2.05	1.60	1.85	0.75
32	0.65	1.05	0.65	0.90	2.00	0.65	0.55
33	0.65	0.50	0.40	0.70	0.35	0.50	0.70
34	1.05	0.60	0.75	1.00	0.45	0.75	1.15
35	1.00	0.55	1.25	1.45	1.20	1.15	1.05
36	0.80	0.90	0.65	1.50	0.70	0.70	0.80

卷舌音[ɚ]、[ar]和舌尖前高元音[ɿ]、[ʅ]在舌尖下出现一个舌下腔，针对卷舌元音的立体声道模型以后再讨论，目前只研究平舌元音。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

将图 1b 中各个截面边缘线等分，就可以得到 36 个截面的边缘点数据，如果在声道边缘线与矢量半径之间建立一组映射关系，那么只要给定一组矢量半径，就可以得到一组声道边缘线，这样就可以组成一个立体网格声道模型。当矢量半径改变时，可以根据已有的 7 个单元音的截面，让截面形状作线性过渡。当边缘线上的点作线性渐变时，截面与水平面的夹角也自然做了调整，于是就得到了用矢量半径驱动的汉语普通话立体声道模型。

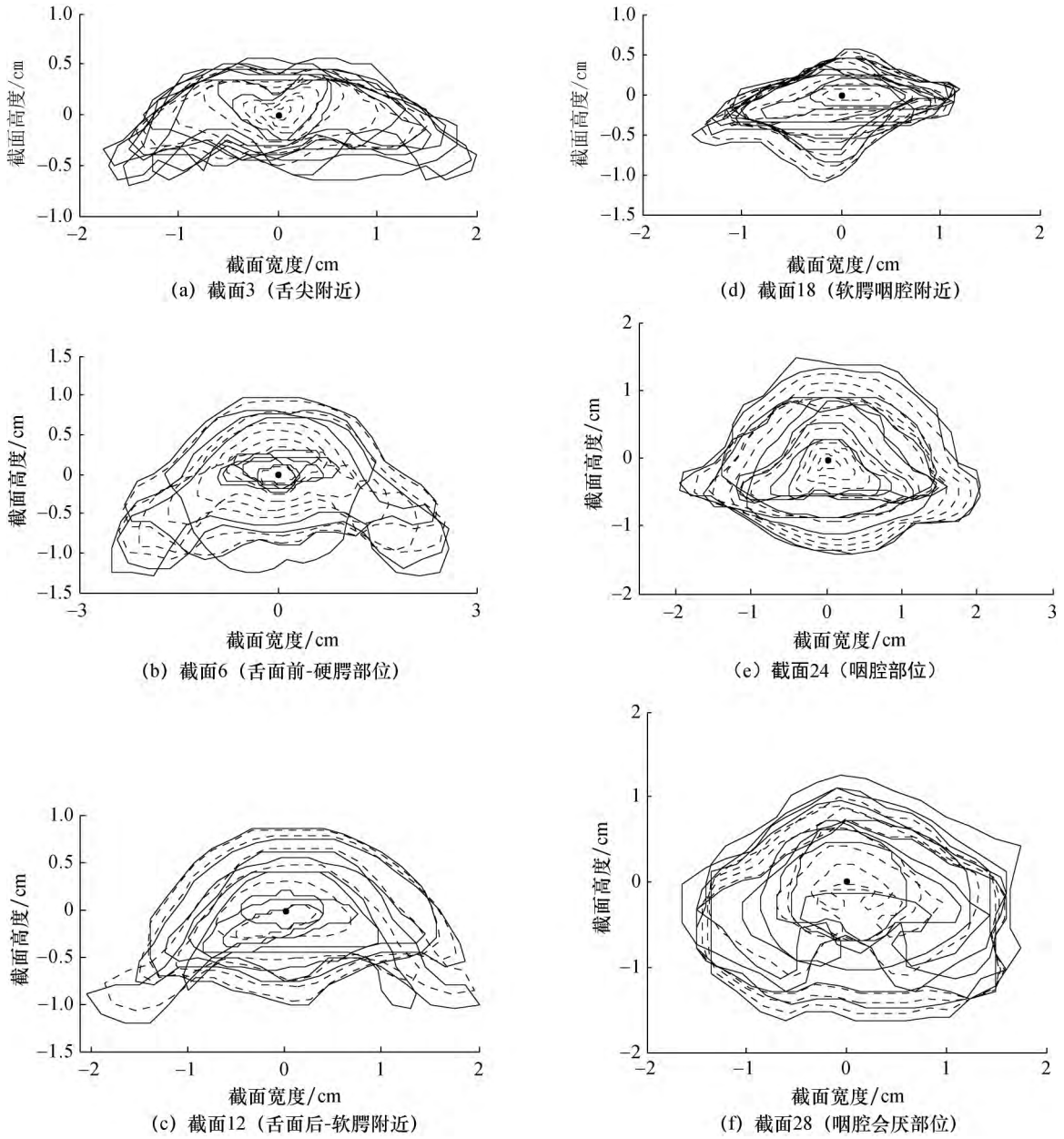


图 3 声道不同位置的截面边缘线

图 3 中的 6 幅图分别是根据矢量半径做线性变化的截面 3、6、12、18、24、28 的边缘线，实线是从 MRI 中获得的 7 个单元音原始截面边缘，虚线是相邻边缘做线性变化得到的边缘线。

截面 3 是在舌尖附近，从图形上看，最大的边缘像一个倒扣的月牙，水平宽度大约有 4 cm，垂直宽度大约 1 cm。

截面 6 在舌面前-硬腭部位，边缘线多半像倒扣的元宝，中间隆起部位是硬腭部位，下边是舌面，其中最大的边缘线水平宽度接近 6 cm，垂直宽度大约 2 cm。

截面 12 在舌面后-软腭附近，最大的边缘像一个贝壳或是一把打开的扇子，线条逐渐向中心收缩为半圆形或椭圆形。其中最大的边缘线水平宽度大

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

约 4 cm，垂直宽度大约 2 cm。

截面 18 在软腭附近，边缘线近似菱形，在这个区域声道的变化比较小，最大的边缘线水平宽度才 1.5 cm 左右，垂直宽度也只有 1 cm 左右。

截面 24 和截面 28 都在咽腔部位，截面 24 较靠上一些，截面 28 较靠下一些。这 2 个截面中，最大的边缘线接近圆形，直径大约 3 cm。

根据声道截面的矢量半径和截面边缘线数据，就可以建立矢量半径驱动的立体声道模型，该声道模型可以简单记作：

$$\mathrm{ShapeVT}=\sum_{i=1}^{n}\mathrm{SectionLine}_i(R_i). \tag{1}$$

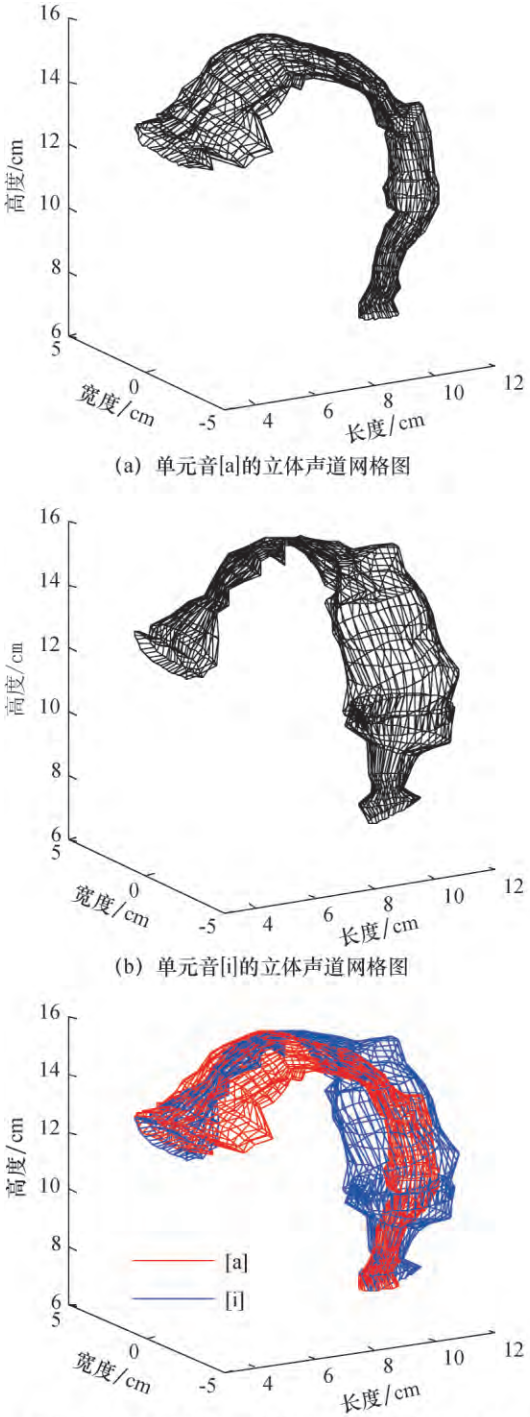
其中：ShapeVT 是声道形状，R 是矢量半径，SectionLine 是截面边缘线，SectionLine 是矢量半径 R 的函数，SectionLine 的形状与[a]、[o]、[ɤ]、[i]、[u]、[y]、[e]这 7 个单元音声道截面边缘的形状有关。该文沿声道中线截取了 36 个截面，所以式(1)中 n 为 36。这里的求和符号并不是数字的累加，而是表示一组边缘线。

图 4 是汉语普通话单元音[a]和[i]的声道三维网格模型及模拟双元音韵母[ai]的起始状态[a]和结束状态[i]的三维立体声道模型。左边开口处是唇的位置，下端开口处是声门。

从图 4c 中可以获得发音时声道变化的细节信息，而且这些信息是很难直接观察到的：从单元音[a]到[i]渐变过程中，口腔体积逐渐收缩，咽腔体积逐渐扩大。而且在从[a]到[i]的发音过程中，硬腭相应地向下运动，因此可以看到[a]和[i]硬腭部位的边缘线并不重合，虚线网格([a])在实线网格([i])上方，而在软腭部位，虚线网格([a])在实线网格([i])下方，这说明在发汉语单元音[i]时，舌面向硬腭方向运动，软腭上抬，因此单元音[i]的咽腔体积比单元音[a]大得多。

改变矢量半径便可生成一组从[a]到[i]过渡状态的立体声道，进而获得相应的声道面积函数和共振峰模式。

图 5a 是汉语普通话单元音[a]到[i]过渡的声道面积函数，从图上看，元音[a]前腔大，后腔小，而元音[i]前腔小，后腔大，这与图 6 一致。图 5b 是汉语普通话单元音[a]到[i]过渡的声道传递函数。



(a) 单元音[a]的立体声道网格图

(b) 单元音[i]的立体声道网格图

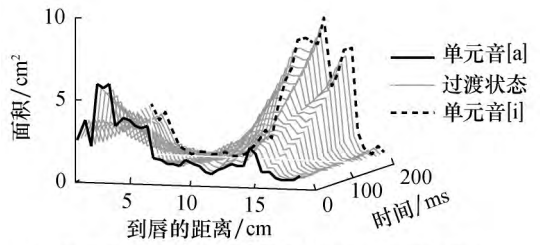
(c) 模拟双元音韵母[ai]的起始状态[a]和结束状态[i]的立体声道模型（网络版彩图）

图 4　汉语普通话立体声道网格模型

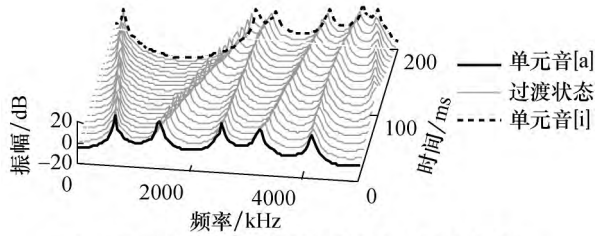
2　模型评估

为了评价该声道模型，需要对比磁共振同步录制的自然语音与用声道面积函数计算的前 4 个共振峰的频率数据(见表 3)，并对合成语音进行听辨。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net



(a) 汉语普通话单元音[a]到[i]过渡的声道面积函数



(b) 汉语普通话单元音[a]到[i]过渡的声道传递函数

图 5 汉语普通话[a]到[i]过渡声道面积函数和声道传递函数

表 3 自然语音与计算模拟的语音共振峰数据

单位：kHz

	*n*F1	*n*F2	*n*F3	*n*F4	*c*F1	*c*F2	*c*F3	*c*F4
[a]	0.87	1.30	3.00	3.70	0.70	1.40	2.55	3.50
[o]	0.68	1.05	2.86	3.47	0.60	1.05	2.55	3.25
[ɤ]	0.68	1.12	2.86	3.58	0.50	1.25	2.90	3.45
[i]	0.35	2.55	3.12	4.06	0.20	2.55	2.95	3.80
[u]	0.35	0.61	2.60	3.70	0.25	0.55	2.9	3.45
[y]	0.32	2.18	2.80	3.08	0.20	1.30	2.50	3.25
[e]	0.47	2.15	2.80	3.55	0.30	2.30	2.75	3.40

表中 *n*F1—4 是 MRI 同步录音的前 4 个共振峰频率，*c*F1—4 是由声道面积推算出来的前 4 个共振峰频率。从表中可以看出元音[y]前 2 个共振峰频率的计算结果与自然语音差别较大。

听辨实验的结果表明：除[y]听起来与 MRI 同步录音存在差别外，其他 6 个单元音都能很清晰地辨认，其中元音[a]、[o]从听感上靠后，因此可以说该声道模型还是比较可靠的。

3 讨论与结论

本文构建了一个用矢量半径驱动的汉语普通话立体声道模型，只要有实时的中央矢状面的声道数据，就可以对人们说话时口腔的形状变化进行实时仿真。根据矢量半径驱动的立体声道模型，不但可以比较清晰地展示声道形状的具体细节，而且可以动态展示声道的变化过程，有利于汉语普通话语音的教学和研究，甚至可以用模型得到精确的声道特性，并用它来合成语音，进行说话人识别，该模型也可以为将来仿真机器人的设计和研究提供参考。

本文实现了从 MRI 3-D 图像数据提取声道边缘，建立立体声道模型，并获得牙缝、梨状窝等声道细节信息，但是在利用声道形状计算传递函数方面，暂时还没有很好的算法将牙缝、梨状窝等腔体的共鸣计算进来，而且声道壁不是一种均匀材质，不同位置的壁阻抗不同，这会影响腔体的共鸣特性，这也是合成的语音与自然语音存在误差的一个重要原因。此外，该模型对舌头的运动处理能力不足，暂时不能模拟卷舌元音。因此研究复杂声道结构的声道特性，添加舌下腔模块等，都是下一步工作的研究内容。

参考文献 (References)

[1] 马大猷. 说话的科学技术[M]. 北京：清华大学出版社，2004.
MA Dayou. Talking Science and Technology [M]. Beijing: Tsinghua University Press, 2004. (in Chinese)

[2] Stevens K N, House A S. Development of a quantitative description of vowel articulation [J]. *Journal of the Acoustical Society of America*, 1955, **27**: 484-493.

[3] Fant G. The Acoustic Theory of Speech Production [M]. Hague: Mouton, 1960.

[4] Heinz J M, Stevens K N. On the derivation of area functions and acoustic spectra from cineradiographic films of speech [J]. *Journal of the Acoustical Society of America*, 1964, **36**: 1037.

[5] Sundberg J. On the problem of obtaining area functions from lateral X-ray pictures of the vocal tract [J]. *Royal Inst Technol STL-QPSR*, 1969, **1**: 43-45.

[6] Chiba T, Kajiyama M. The Vowel: Its Nature and Structure [M]. Tokyo: Kaiseikan Publishing Company, 1942.

[7] Baer T, Gore J C, Gracco L C, et al. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels [J]. *Journal of the Acoustical Society of America*, 1991, **90**(2): 799-828.

[8] Story B H, Hoffman E A, Titze I R. Vocal tract imaging: A comparison of MRI and EBCT [J]. *Medical Imaging Physiology and Function from Multidimensional Images*, *Hoffman*, 1996, **2709**: 209-222.

[9] Narayanan S S, Alwan A A, Haker K. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals [J]. *Journal of the Acoustical Society of America*, 1997, **101**(2): 1064-1077.

[10] Alwan A, Narayanan S, Haker K. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics [J]. *Journal of the Acoustical Society of America*, 1997, **101**(2): 1078-1089.

[11] Espy-Wilson C Y, Boyce S E, Jackson M, et al. Acoustic modeling of American English vertical bar r vertical bar [J]. *Journal of the Acoustical Society of America*, 2000, **108**(1): 343-356.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[12] Story B H, Titze I R. Parameterization of vocal tract area functions by empirical orthogonal modes [J]. *Journal of Phonetics*, 1998, **26**(3): 223-260.

[13] Story B H. A parametric model of the vocal tract area function for vowel and consonant simulation [J]. *Journal of the Acoustical Society of America*, 2005, **117**(5): 3231-3254.

[14] Dang J W, Honda K. Construction and control of a physiological articulatory model [J]. *Journal of the Acoustical Society of America*, 2004, **115**(2): 853-870.

[15] Dang J W, Honda K. Estimation of vocal tract shapes from speech sounds with a physiological articulatory model [J]. *Journal of Phonetics*, 2002, **30**(3): 511-532.

[16] Dang J W, Honda K, Suzuki H. Morphological and acoustical analysis of the nasal and the paranasal cavities [J]. *Journal of the Acoustical Society of America*, 1994, **96**(4): 2088-2100.

[17] Dang J W, Honda K. Acoustic characteristics of the piriform fossa in models and humans [J]. *Journal of the Acoustical Society of America*, 1997, **101**(1): 456-465.

[18] Dang J W, Shadle C H, Kawanishi Y, et al. An experimental study of the open end correction coefficient for side branches within an acoustic tube [J]. *Journal of the Acoustical Society of America*, 1998, **104**(2): 1075-1084.

（**上接第** 944 **页**）

[8] Shepard C, Yu H, Anand N, et al. Argos: Practical many-antenna base stations [C]// Proceedings of the 18th Annual International Conference on Mobile Computing and Networking. Piscataway, NJ, USA: IEEE Press, 2012: 53-64.

[9] Rogalin R, Bursalioglu O Y, Papadopoulos H, et al. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO [J]. *IEEE Transactions on Wireless Communications*, 2014, **13**(4): 1815-1831.

[10] Wei H, Wang D, Wang J, et al. TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent [J]. *Science China Information Sciences*, 2015, **59**(10): 1-10.

[11] Markovsky I, Van Huffel S. Overview of total least-squares methods [J]. *Signal Processing*, 2007, **87**(10): 2283-2302.

[12] Meinilä J, Kyösti P, Jämsä T, et al. WINNER II channel models [M]// Radio Technologies and Concepts for IMT-Advanced. Hoboken: John Wiley & Sons, Ltd., 2008: 39-92.

[13] Gvozdenovic N, Gentner P K, Mecklenbräuker C F. Near-field channel measurements for UWB RFID with on-chip antenna [C]// Antenna Technology (iWAT). Piscataway, NJ, USA: IEEE Press, 2012: 28-31.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

国家哲学社会科学学术期刊数据库
National Social Sciences Database

新寨苗语单字调及双字调声学实验研究*

刘　文　杨正辉　孔江平

[**提要**]本文基于田野实地录音语料，提取基频和时长两项声学参数，并以此分析了新寨苗语单字调及双字调的基频模式。结果表明，该方言有八个单字调：五个平调、两个升平调和一个高降调，五平调是新寨苗语的一大特点。另外，双字组合没有发生变调现象，即双字组连读共有64种组合形式。五平调在双字组合中没有发生变调亦是一个十分特别的现象。在此基础上，本文建立了一套用于描写新寨苗语单音节和双音节声调的区别特征系统。最后，文章结合苗语其它方言讨论了新寨苗语双字组连读不变调的现象。

[**关键词**]新寨苗语　单字调　双字调　基频　时长　区别特征

新寨位于贵州省黔东南苗族侗族自治州凯里市三棵树镇板新村。新寨苗语属于苗语黔东方言北部土语（中国社会科学院、澳大利亚人文科学院 1987:《中国语言地图集 C9》），根据笔者的田野调查，新寨苗语有33个声母、20个韵母和8个声调。

从历史比较的角度看，古苗语有八个声调，四个阴调，四个阳调，阴阳调的分立是根据声母来源的清浊。（张琨 1947；K. Chang 1953、1972；王辅世 1994；王辅世、毛宗武 1995）50年代的民族语言大普查对我国境内的苗瑶语进行了全面的调查，整理了多个苗语方言点的音系，其中最具代表性的苗语黔东方言是养蒿苗语。此外，已发表的苗语黔东方言北部土语的材料还有清江苗语、（Kwan 1971；Kuang 2013）鱼粮苗语。（朱晓农、石德富、韦名应 2012）

黔东苗语在类型学上的一个显著特点是声调系统中平调的个数比较多，例如，养蒿苗语有四个平调，（王辅世 1985）清江苗语和本文讨论的新寨苗语都有五个平调。

本文的重点不在于分析平调个数的多少，而是考察含有多个平调的语言在双音节组合中是否发生连读变调及其背后的机制。基于笔者的田野调查，提取基频和时长两项声学参数来分析新寨苗语单字调和双字调的基频模式。

一　实验说明

本文所用语料来自笔者于2013年1月的田野调查，及2014年5月和2016年6月的补充调查。本项研究选取了四位发音人，两男两女，母语均为新寨苗语，具体见表1。

* 本研究得到国家社科基金重大项目“中国有声语言及口传文化保护与传承的数字化方法及基础理论研究”（10&ZD125）资助。孔江平为本文通讯作者。文章初稿曾在第十二届全国语音学学术会议（内蒙古，2016年7月）上报告，承蒙多位专家指教。田野调查中，得到杨妹勇、杨占国等人的热心帮助。文章投稿后，又蒙《民族语文》匿名审稿专家和编辑部老师惠赐建议，谨一并致谢。文中若有谬误，概由笔者负责。

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

表1　发音人信息表

编号	姓名缩写	年龄	民族	是否母语者	出生地	职业
男1	YZH	25	苗	是	新寨	学生
男2	YJ	28	苗	是	新寨	务农
女1	YZF	24	苗	是	新寨	务农
女2	YYQ	20	苗	是	新寨	学生

录音软件为Cooledit 2.1，采样频率为22050Hz，采用双通道录制，左声道由电容式麦克风（SONY ECM-44B）录制语音信号（Sound Pressure），右声道由喉头仪（Electroglottograph Model 7050A）录制声门阻抗信号（EGG），两路信号通过调音台（XENYX 302 USB）和声卡（SBX）进入电脑。信号分析软件是利用Matlab编写的VoiceLab。

在词表的选取上，考虑到声母对声调的影响，我们要求所选单音节语素例词声母为清不送气塞音声母[p、t、k、q]，韵母为单元音韵母[a、i、u、ə、ɛ、o]。在此标准下，共选取了[pa、ta、qa、pi、ki、to、ko、tɛ、pə、kə、pu、tu、ku]十三组声韵母相同的组合搭配八个不同的声调。由于部分组合在该方言中没有相应词汇形式，所以十三组共有73个例词，每个词读两遍。单字调例词见表2。

表2　新寨苗语单字调录音例词

调类	T1	T2	T3	T4	T5	T6	T7	T8
pa	大腿	把谷子推到一个地方	雄性；父亲	坏机器~	猪	首~歌	——	八
ta	地下	来	长；尾巴	丢	霜；早上	死	——	袋；扔
qa	鸡；主人	——	屎；肠	——	啼；价钱	卡住	——	——
pi	给	掉	满	——	——	木板	——	笔
ki	角；螺	哭；叫	方位词	干水~	结~冰	露出	剪	挖
to	布	冲~水	捣；砍	远	斧头；不	哪儿	梭子	凳子
ko	歌	投	——	喊	男性生殖器	被	——	——
tɛ	儿子	前面	——	有生量词	断	——	——	踏
pə	——	埋	——	刺儿	响	抱	——	——
kə	女婿	——	量词	——	——	——	硬	——
pu	知道	起泡	补	朋友	沸腾	——	打开	见
tu	放置	斗度量单位	书	火	——	爆炸	——	读
ku	——	猜	——	——	钩子	——	鱼篮	撞击

理论上讲，新寨苗语双音节声调有8×8=64种组合形式，每种组合选取两个例词，一共选取了128个词项，每个词读两遍。双字调例词见表3。

表3　新寨苗语双字调录音例词

	T1	T2	T3	T4	T5	T6	T7	T8
T1	$tɕ^hiɛ^1ta^1$	tsa^1na^2	$qɔ^1kiɛ^3$	$qo^1ʑi^4$	tie^1pa^5	qa^1ta^6	$ta^1ɕo^7$	$pie^1paŋ^8$
	扫地	晒谷子	家务	靠椅	杀猪	死鸡	红土地	我们的
	qa^1ta^1	$tiɛ^1tiu^2$	ka^1kia^3	$qaŋ^1ɣiɛ^4$	$ts^ho^1tɕi^5$	$kaŋ^1ɣia^6$	tsa^1to^7	$tsa^1tɕiu^8$
	地下	敲门	炒饭	最小的孩子	吹口哨	蜘蛛	五次来回	五十

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

T2	tiu^{2}pʰiə^{1} 大门 taŋ^{2}ʔɔ^{1} 淹没	ta^{2}ta^{2} 来不来 tɕo^{2}ta^{2} 桌子	nɔ^{2}kia^{3} 吃饭 lo^{2}qa^{3} 厕所	ta^{2}mə^{4} 过去 ta^{2}za^{4} 来吧	tiaŋ^{2}pa^{5} 猪油 ta^{2}to^{5} 来不	ta^{2}na^{6} 量一量 ŋa^{2}ta^{6} 死动物的肉	paŋ^{2}ɕo^{7} 红花 ka^{2}to^{7} 拉织布机的布	ta^{2}tɔ^{8} 来咬 tɕiə^{2}tɕiu^{8} 九十
T3	pa^{3}pʰiə^{1} 大伯 qa^{3}ta^{1} 泥土	kiɛ^{3}ta^{2} 扛来 kʰaŋ^{3}faŋ^{2} 亮处	kʰaŋ^{3}ɬoŋ^{3} 窗子 kʰa^{3}kia^{3} 恰饭	ta^{3}lo^{4} 拿回来 ta^{3}piɛ^{4} 食指	pa^{3}zu^{5} 叔叔 ka^{3}mʰaŋ^{5} 晚餐	po^{3}na^{6} 鼻子 kʰaŋ^{3}to^{6} 何处	tu^{3}ɕo^{7} 红纸 kia^{3}kə^{7} 硬饭	ta^{3}ta^{8} 拿袋子 tiaŋ^{3}to^{8} 用凳子击打
T4	tɕiɛ^{4}ki^{1} 角 to^{4}qa^{1} 鸡（复数）	tɛ^{4}ta^{2} 还来 to^{4}ta^{2} 变远	naŋ^{4}ʔu^{3} 穿衣服 lo^{4}ʦa^{3} 回家	lo^{4}za^{4} 来吧 tu^{4}ta^{4} 散失	naŋ^{4}kʰɔ^{5} 穿裤子 tɛ^{4}zu^{5} 弟弟	na^{4}ta^{6} 死鱼 moŋ^{4}mɛ^{6} 脸庞	na^{4}ɕo^{7} 红鱼 lo^{4}ɕo^{7} 流血	tɛ^{4}tɔ^{8} 还咬 naŋ^{4}za^{8} 老鼠舔
T5	tiə^{5}ha^{1} 穿鞋 to^{5}nʰa^{1} 不动	to^{5}ta^{2} 不来 no^{5}ta^{2} 变多	tɔ^{5}ta^{3} 长树枝 kia^{5}qa^{3} 刨粪	pʰə^{5}ɲi^{4} 盖瓦 to^{5}lo^{4} 不回来	ʔɛ^{5}nia^{5} 做傻子 tiə^{5}ta^{5} 结霜	pa^{5}ta^{6} 死猪 to^{5}ɣiə^{6} 得到力气	to^{5}kiə^{7} 不强 to^{5}ʦo^{7} 别顽皮	ɕi^{5}niə^{8} 摔跤 tiu^{5}tɕiu^{8} 六十
T6	hɔ^{6}ʔɔ^{1} 喝水 tia^{6}qa^{1} 赶鸡	tɔ^{6}ta^{2} 跟来 kʰo^{6}tei^{2} 划着	ha^{6}kia^{3} 盛饭 nɛ^{6}ʦa^{3} 询问家的位置	tɔ^{6}lo^{4} 跟来 ta^{6}na^{4} 死鱼	ŋa^{6}tɔ^{5} 扶树枝 ta^{6}pa^{5} 死猪	tu^{6}tɔ^{6} 跟随 ka^{6}ta^{6} 死鸭子	ka^{6}ɕo^{7} 红鸭子 kʰɔ^{6}ɕo^{7} 划红线	hu^{6}tɔ^{8} 挑选豆子 nia^{6}to^{8} 扶椅子
T7	qei^{7}vo^{1} 挟菜 ki^{7}to^{1} 剪布	ɕo^{7}ta^{2} 变红 kə^{7}ta^{2} 变硬	ɕo^{7}mʰi^{3} 牙齿红 ku^{7}ʦa^{3} 房子封顶	ɕo^{7}mə^{4} 变红 ku^{7}na^{4} 捕鱼	kə^{7}va^{5} 很硬 ɕo^{7}to^{5} 红不红	ɕo^{7}to^{6} 红吗 qei^{7}ko^{6} 挟到	ɕo^{7}ɕo^{7} 红不红 kə^{7}kə^{7} 硬不硬	liu^{7}ta^{8} 摘下来放好 kiɛ^{7}ta^{8} 咬袋子
T8	kaŋ^{8}ɬu^{1} 白盆 na^{8}hi^{1} 担心	zu^{8}na^{2} 收集谷子 ki^{8}ʨaŋ^{2} 挖成	tu^{8}tu^{3} 读书 za^{8}tsa^{3} 八家	ta^{8}mə^{4} 扔下来 za^{8}mə^{4} 强迫去做	na^{8}to^{5} 辣吗 ta^{8}ɣu^{5} 好袋子	niu^{8}tiu^{6} 筷子 ta^{8}nei^{6} 破袋子	ta^{8}ɕo^{7} 红袋子 to^{8}kə^{7} 硬板凳	ta^{8}ta^{8} 每个袋子 tio^{8}vo^{8} 滚动

注：本表采用调类标示法，第一列代表双字组的前字，第一行代表双字组的后字。

二　单字调实验

田野调查中，一个十分有价值的信息是我们找到了三组涵盖八个声调的最小对立对，即声韵母结构相同、声调不同的最小对立对（minimal pair），见表 4。

表 4　新赛苗语单字调最小对立对

调类	T1	T2	T3	T4	T5	T6	T7	T8
[ki]	螺	哭	方位词	干~涸	结~冰	露出	剪	挖
[to]	布	冲	捣	远	不	哪儿	梭子	凳子
[tɕiu]	针	桥	酒	完	竹篾	尽头	刨	十

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

为了给大家一个直观的认识，我们选取了[ki]组的语图，如图 1 所示。

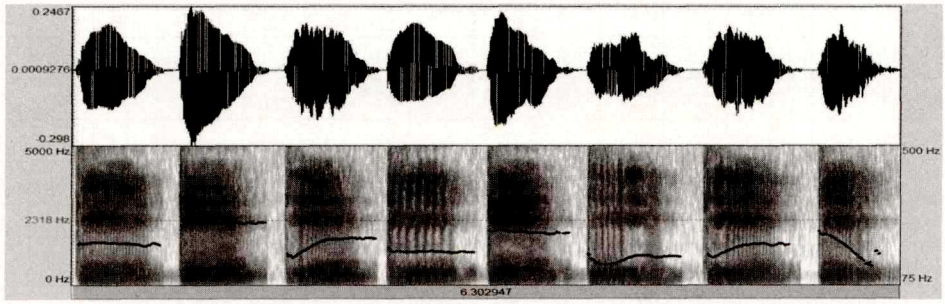


图 1　男 2 [ki] 组语图示例（从左至右：T1→T8）

图 1 显示了新寨苗语八个声调的基频走势。如果忽略弯头降尾，T1、T2、T4、T5、T6 的基频走势为平，并且它们的基频值分别处于五个不同的音高水平上。T2 的基频值最高，其次是 T5，再次是 T1、T4，基频值最低的是 T6。听感上，T6 调的音节存在明显的气嗓音（breathy voice）。T3 和 T7 为升调，但是这种升调不同于汉语普通话的升调模式，汉语普通话的升调是斜线上升，而新寨苗语的 T3 和 T7 调呈现“先升后平”的模式；此外，这两个升调的基频起点大致相同，所不同的是终点，T3 终点的基频值相对要高一些。T8 为降调，由于该方言只有一个降调，所以在田野调查中很容易识别。

基于以上观察，下文我们通过声学分析分别给出新寨苗语男性和女性单字调的基频模式图，目的是确定其声调调值。绘图的具体操作程序如下：

第一步，确定语音波形和声门阻抗信号（EGG 波形）的起点和终点。为了使数据的处理具有内部一致性，在选择样本的起点和终点时，我们利用自编的 MATLAB 程序计算出基频曲线、能量曲线以及共振峰结构等参数。起点选在 EGG 波形图上开始稳定的周期点，终点选在第二共振峰结构稳定、基频曲线平稳、能量恒定的点。

第二步，数据归一化处理。每个例词提取 20 个 F0（基频）值和时长数据，将每个例词的数据进行平均，然后将同一调类下所辖的所有样本进行平均，并对同一声调中不同字的时长进行归一化处理，目的是为了使 F0 和时间相对应，具体参数见表 5。

表 5　新寨苗语单字调基频和时长参数

	T1		T2		T3		T4		T5		T6		T7		T8	
	男	女	男	女	男	女	男	女	男	女	男	女	男	女	男	女
1	195	253	248	303	169	220	176	231	222	278	140	193	163	212	226	266
2	198	252	254	305	173	220	176	229	226	279	138	190	168	213	231	266
3	199	251	257	306	179	223	176	227	228	278	138	188	171	215	234	265
4	200	250	259	306	189	229	176	226	230	278	139	188	175	219	234	261
5	200	249	260	306	198	236	176	226	231	277	141	189	180	225	233	257
6	200	248	261	305	206	242	176	225	231	276	141	190	184	231	231	254
7	200	247	261	305	213	247	176	224	232	276	143	191	189	236	229	249
8	200	246	262	305	219	252	175	223	232	277	145	192	193	240	226	245
9	200	246	262	305	223	255	175	223	232	277	147	192	196	242	222	241

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

10	199	246	262	305	225	259	175	223	233	277	148	193	199	244	218	237
11	199	245	262	305	227	261	175	222	233	277	149	193	200	246	213	232
12	199	245	262	304	228	263	175	222	232	277	149	192	202	247	207	228
13	199	245	262	304	228	264	175	221	232	276	150	192	203	248	200	222
14	198	244	262	303	229	266	174	221	231	276	150	192	203	250	191	216
15	197	243	262	303	229	267	174	221	231	275	149	191	203	252	181	209
16	197	242	262	302	229	267	173	220	230	275	149	191	203	252	172	202
17	197	241	261	301	229	267	173	219	229	274	148	191	203	251	161	195
18	197	239	260	301	229	267	172	217	229	273	147	190	203	249	150	189
19	198	238	260	300	229	264	172	215	229	271	145	189	203	246	140	182
20	197	236	259	300	229	260	172	214	228	268	143	188	202	243	131	176
时长	0.43	0.58	0.38	0.51	0.37	0.53	0.39	0.61	0.39	0.54	0.39	0.58	0.38	0.55	0.21	0.43

注：基频的单位是赫兹（Hz），时长的单位是秒（s）。

第三步，作图。利用上一步得到的数据，将 X 轴定为时间，Y 轴分别定为 F0 值，绘制散点图。需要说明的是，在基频曲线的绘制过程中，为了使得到的数据接近人们的听感，[①]笔者将 F0 转换为半音，见图 2 和图 3。（刘复 1924；孔江平 2015）

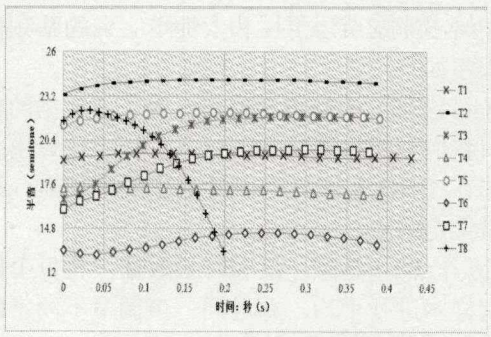


图 2　新寨苗语单字调半音图（男）

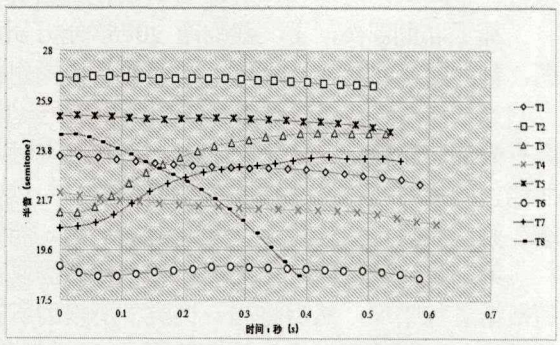


图 3　新寨苗语单字调半音图（女）

从表 5 和图 2、图 3 中可以看出，无论男性还是女性，新寨苗语有八个单字调，含五个平调（T1、T2、T4、T5、T6），一个降调（T8），还有两个升平调（T3、T7）。男性和女性所表现出来的基频曲线的走势完全一致，所不同的是二者的基频值和音域范围。女性发音人的基频值在每一个调类上都大于男性，但转换为半音后，我们可以发现：女性发音人的音域在 18-27 个半音之间，中间相隔 9 个半音，而男性发音人的调域范围在 13-25 个半音之间，中间相隔 12 个半音。尽管男性发音人每一个调类的绝对基频值都小于女性发音人，但是其音高（半音）范围大于女性发音人。

根据半音法所得数据，采用“五度标调法”（Chao 1930），我们一一论述新寨苗语的八个调类。

① 相对于纯物理声学参数的基频而言，半音值比较符合人的听感，因为半音之间的关系反映的是听感上的关系。此外，目前国际上的声调和音调的研究都采用半音法。（孔江平 2015:64-65）

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

需要说明的是，“五度标调”只是一个相对概念，我们采用它只是为了区分既有调类。在调值的具体确定过程中，我们采用刘复（1924）的方案，即，先计算出半音最大值和最小值之间的差值，然后说明一度之间相差几个半音。为了便于理解，图中我们将半音最大值点和最小值点放在一度的中间。由此可知，对男性而言，每两度之间相差 2.8 个半音，而女性相差 2.1 个半音。

T1：无论男性还是女性，T1 都处于音域的中间位置，统一标记为 33。另外，相对于男性而言，女性的 T1 调曲线有些下倾，但是这并不影响母语者将之感知为平调。类似的情况同样出现在女性的 T2、T4 和 T5。这就说明，平调的曲线并非完全平直，当斜率控制在一定范围内，母语者即可感知为平。

T2：亦是一个平调，基频值在整个音域内最大，可以标记为 55，男性在 25 个半音左右，女性在 27 个半音左右。对男性而言，发音的起始位置略升，这是发音人声带由初始状态向高目标值过渡的伴随现象。

T3 和 T7：二者的起点都在第 2 度内，所不同的是上升的幅度有大有小，T3 调的终点落在第 4 度内，而 T7 调的终点落在第 3 度内。另外，就基频而言，T3 和 T7 调在发音的前半程基频呈上升趋势，随后是基频稳定段，这与我们通常所见到的升调模式不同，典型升调基频值的上升贯穿整个音节，例如，汉语普通话的第二声（35 调），而新寨苗语的 T3 和 T7 调则呈现了“升”的另外一种模式，即前升后平，这一点与同属苗语黔东方言的鱼粮苗语的升调模式相同。（刘文、张锐锋 2016）鉴于此，本文将之定义为“升平调”，调值分别标记为 244（T3）和 233（T7）。

T4：无论男性还是女性，T4 是一个中低平调，标记为 22。

T5：无论男性还是女性，T5 是一个中高平调，标记为 44。

T6：男女性 T6 调的音高曲线都有所波动，但其变化范围在同一度内。此外，T6 调在听感上存在明显的气嗓色彩（breathy）。前人研究表明，声调感知不仅涉及到基频，还涉及特殊发声类型和时长等因素。（Yang 2015）T6 调音高曲线的波动应该是由特殊发声类型影响所致。

T8：是一个高降调，并且一降到底，可以标记为 41。

就时长而言，具体到每一个调类，女性单字调的时长都大于男性，并且，降调的时长在所有调类中相对来说最短。

此外，一个十分值得讨论的话题是新寨苗语的五个平调在母语者音域中的分布情况。Maddieson（1978）通过跨语言研究，指出世界上没有超过五个平调对立的语言，也就是说，五个平调是一个语言所能容纳的最大平调数量。由图 2 和图 3 可知，不论男性发音人还是女性发音人，新寨苗语的事实表明五个平调在声学空间中大致呈现均匀分布的状态。关于新寨苗语五平调的感知实验，我们另文讨论。

通过上文分析，我们采用区别特征来表示新寨苗语的单字调系统，如表 6 所示。

表 6　新寨苗语单字调区别特征表

调类	调值		基频
	男	女	
T1	33	33	中平

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

T2	55	55	高平
T3	244	244	高升平
T4	22	22	中低平
T5	44	44	中高平
T6	11	11	低平
T7	233	233	低升平
T8	41	41	高降

由表 5 和表 6 可知，男女性别差异体现在绝对基频值和音域范围上，但是这并不影响每一个调类在音域中的相对分布。

三　双字调实验

根据单字调的声学分析结果可得，新寨苗语的双字调在理论上有 64 种音高组合形式。事实真的如此吗？下文我们以两位男性的数据为例，通过分析双字调的音高半音值对这一问题进行求证。

3.1 双字调基频模式

为了更加清晰地呈现新寨苗语双字调的音高组合模式，本小节从以下两个角度来考察：一是前字是同一个调类，后字分属八个不同的调类；二是后字是同一个调类，前字分属八个不同的调类。双字调半音值每个音节选取 20 个点，暂不考虑时长因素，即，下文双字调音高图中，横轴代表点数，纵轴代表半音值。

图 4（左）为前字 T6（11）调，后字为八个不同声调的情况。11 调作前字时，音高在 13 个半音上下浮动，与单字调相比无显著差异。前字与后字组合形成八种不同的模式，没有产生变调现象。

图 4（右）为后字 11 调的情况，其值和单字调、以及 11 调作前字时相同。在双字组合中，亦没有发生变调现象。需要说明的是，当前字为高平调 55（T2）时，在音节开始处有一个很短的上升段，这是由声带从起始状态到高音高目标值过渡所带来的伴随现象。

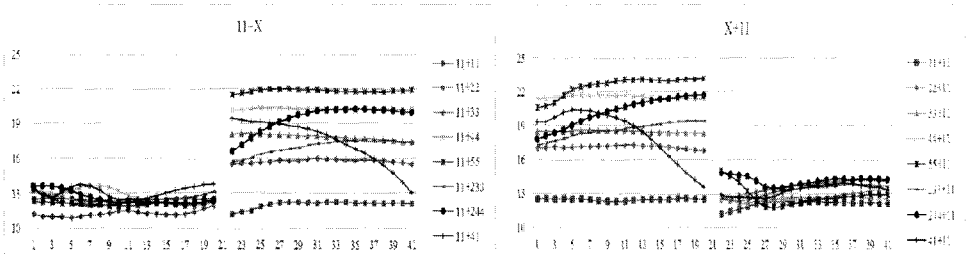


图 4　新寨苗语双字调半音图（左：11+x，右：x+11）

22 调（T4）在作前字和后字时都没有产生变调，基本和单字调一致，音高大约为 17 个半音，如图 5 所示。22 调作前字和后字主要有以下几点表现：第一，22 调作前字时，其后接的 11 调略微有些上升，不过总体仍属于平调范畴；22 调作后字时，11 调的音高值稍低于单字调。值得注意的是，22 调作前字和后字时，其与 11 调之间的半音差值不同，22 调作前字

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

时，11 调和 22 调之间的间距缩小；而 22 调作后字时，11 调和 22 调之间的间距基本与单字调一致，其半音值大于 22 调作前字时的间距。尽管个别调类间的绝对距离有所不同，但整个调域的相对间距（即，11 调和 55 调之间的半音差值）保持不变，例如，22 调作前字时，后接调类的音域在 13-23 个半音间；22 作后字时，前接调类的音域在 12-22 个半音间，也就是说，22 调无论作前字还是后字，与其相连的八个调类的整体音域都是 10 个半音。第二，22 调作前字时，44 调（T5）和 55 调之间的半音差与单字调一致；22 调作后字时，二者有交叉，这与 11 调作后字的情况相同，都是由于声带从休息状态到高音高目标值所带来的伴随现象。第三，22 调无论作前字还是后字，与其组合的高降调 41（T8）的起始点都从中性音高值（3）开始，然后向自己的目标调型进发。第四，22 调作前字时的音高变动范围小于作后字时的值。

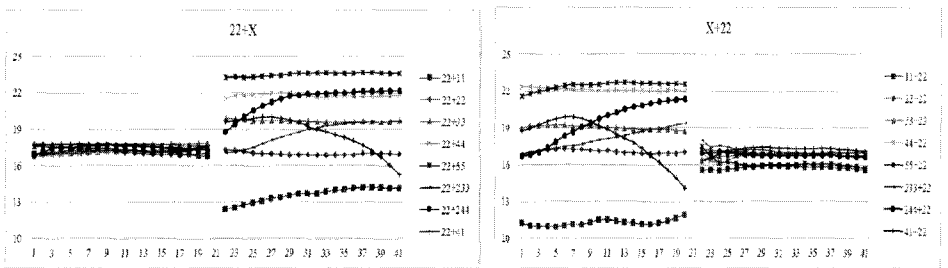


图 5　新寨苗语双字调半音图（左：22+x，右：x+22）

图 6 为 33 调作前字和后字的情况，从图中可见，无论其作前字还是后字都没有变调，基本和单字调一致，大约为 19 个半音。33 调（T1）作前字和后字主要有以下几点表现：第一，33 调作前字时，其后接的 11 调略微有些上升；33 调作后字时，其后接的 11 调略降。无论 11 调略升还是略降，其总体还是平的。第二，33 调作前字时，其后接的 55 调与 44 调在起始处有交叉；33 调作后字时，二者没有交叉。这一现象与 11 调和 22 调的情况不同，它们是在作后字时出现 44 调和 55 调的音高曲线交叉，而 33 调则在作前字时出现两个高平调的音高交叉现象。第三，33 调作前字时，244 调（T3）和 233 调（T7）的调型与单字调一致；33 调作后字时，244 调和 233 调的调型中“升”和“平”的比例关系发生改变，但是这并不足以说明产生了一个新的调类，因为这两种调型在母语者的感知中没有区别意义的功能。第四，33 调作前字时的音高变动范围小于作后字时的值。

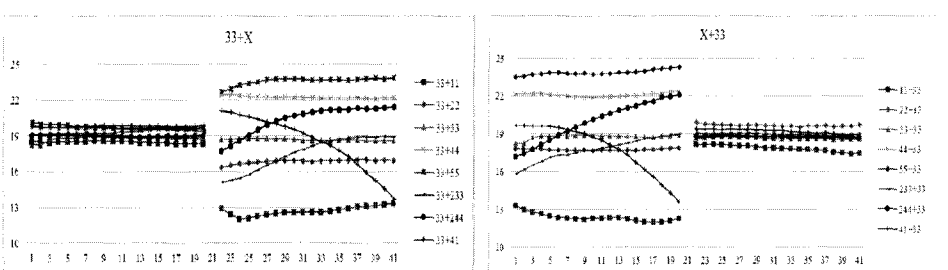


图 6　新寨苗语双字调半音图（左：33+x，右：x+33）

44 调在作前字和后字时没有变调，基本与单字调一致，音高大约为 22 个半音。44 调作前字和后字主要有以下几点表现：第一，44 调作前字时的音高变动范围小于作后字时的值。第二，44 调作前字时，其后接的 41 调的起始音高值与 44 调一致；44 调作后字时，其前接的

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

41 调的起始音高值与 33 调相同，具体见图 7。第三，44 调无论作前字还是后字，44 和 55 两个高平调的音高曲线都没有出现交叉情形。

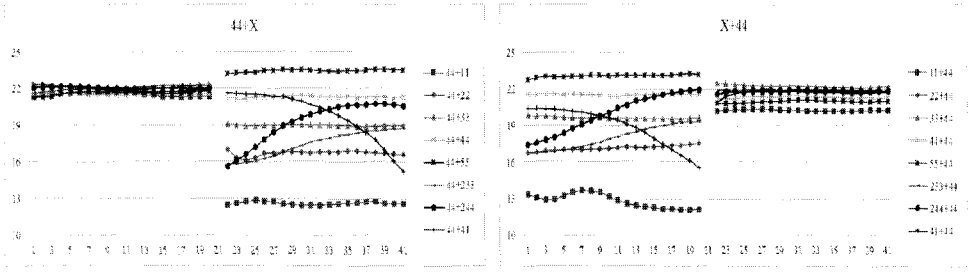


图 7　新寨苗语双字调半音图（左：44+x，右：x+44）

图 8 为 55 调的情况。55 调在作前字和后字时都没有变调，基本和单字调一致，音高在 24 个半音上下浮动。55 调作前字和后字主要有以下几点表现：第一，55 调作前字时，其后接的 11 调先降后平，11 调起始处的下降是由于前接字的音高太高，受协同发音影响所致；55 调作后字时，其前接的 11 调的音高曲线走势为平，不存在先降后平的现象。第二，55 调作前字时，其后接的高降调 41 的起始音高较高，起点与高平调 55 相同；55 调作后字时，前接的 41 调的起始值与 44 调相同。第三，55 调作前字时的音高变动范围小于作后字时的值。

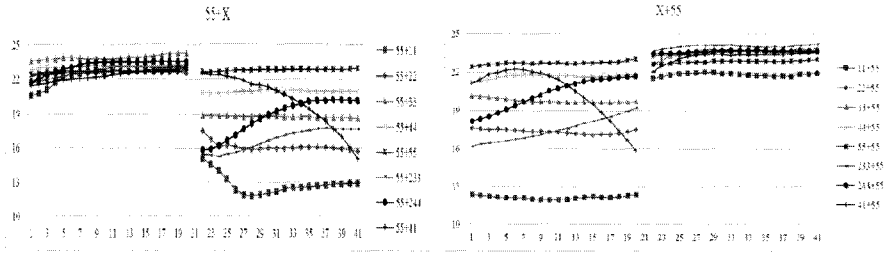


图 8　新寨苗语双字调半音图（左：55+x，右：x+55）

233 调在作前字和后字时都没有变调，基本和单字调一致，音高值大约在 16~20 个半音间。233 调作前字和后字主要有以下几点表现：第一，233 调作前字时，其音高曲线不同于单字调的升平型，由于受到后接字的影响，音高曲线变为升型；233 调作后字时，其走势与单字调相同，仍为升平型。第二，233 调作前字时，其后接的 11 调略微有些上升，不过其总体仍属于平调范畴；233 调作后字时，11 调的音高和单字调基本一致。不过，需要说明的是，11 调和 22 调之间音高差距在 233 调作前字和后字时有所不同，当 233 调作前字时，11 调和 22 调之间的间距缩小；当 233 作后字时，11 调和 22 调之间的间距与单字调一致，但其值大于 233 调作前字时的间距。此外，233 调作前字时，其后接的 44 调和 55 调之间的间距大于 233 调作后字时的情况。综合 11 调和 22 调、44 调和 55 调的情况，可以发现，尽管个别调类间的绝对差值有所改变，但整个调域的相对间距（即 11 调和 55 调之间的距离）保持不变。第三，233 调作前字时，其后接的高降调 41 的起始值较高，与单字调相同；233 调作后字时，前接的 41 调的起始值与 33 调相同。第四，233 调作前字时的音高变动范围小于作后字时的值。见图 9。

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

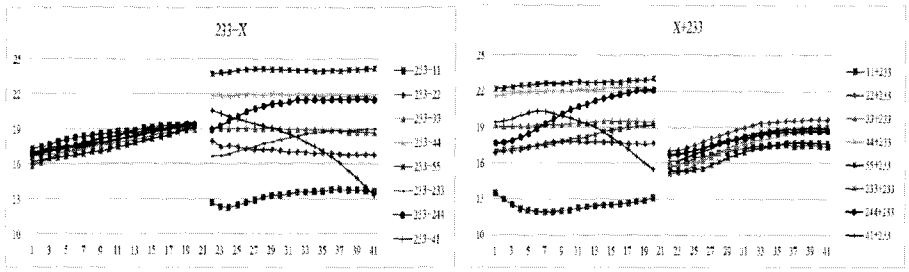


图 9　新寨苗语双字调半音图（左：233+x，右：x+233）

244 调在作前字和后字时都没有变调，基本和单字调一致，音高大约在 17~22 个半音间。244 调作前字和后字主要有以下几点表现：第一，244 调作前字时的音高曲线不如其作后字时与单字调的升平型契合度高。第二，244 调作前字时，其后接的 11 调的音高整体上升；244 调作后字时，11 调的音高走向略微下降。第三，244 调作前字时，其后接的高降调 41 的起始值较高，起点接近 55 调；244 调作后字时，前接的 41 调的起始值与 44 调相同。第四，244 调作前字时，44 调和 55 调之间的距离与单字调一致；244 调作后字时，44 调和 55 调两个高平调在起始处存在交叉现象，这与 11 调和 22 调作后字的情形相同。第五，244 调作前字时的音高变动范围小于作后字时的值。见图 10。

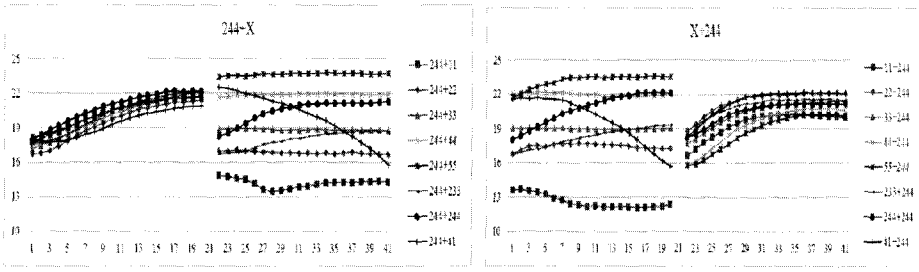


图 10　新寨苗语双字调半音图（左：244+x，右：x+244）

41 调在作前字和后字时都没有变调，基本和单字调一致，音高值大约从 22 个半音降到 13 个半音。41 调作前字和后字主要有以下几点表现：第一，41 调作前字时，44 调和 55 调之间的距离与单字调一致；41 调作后字时，44 调和 55 调两个高平调的音高在起始处有交叉，此外，两个高平调之间的半音差值较小，但是母语者还是能将这两个调类区分开。第二，41 调音节无论作前字还是后字，与其组合的 41 调的起始点都从中性音高值（3）开始，然后向自己的目标调型进发。第三，41 调作前字时的音高变动范围小于作后字时的值。如图 11 所示。

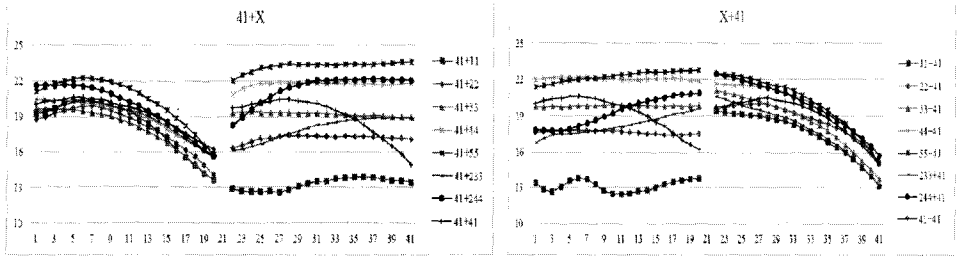


图 11　新寨苗语双字调半音图（左：41+x，右：x+41）

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

从前面的分析中不难发现，无论任何一个声调出现在前字亦或是后字位置，都没有出现调类合流或增生的现象，也就是说，新寨苗语双字组共有 64 种音高组合模式，换句话说，新寨苗语的双字调组合不存在变调现象。表 7 给出了基于区别特征的双字组音高组合模式，从表中可以看出每一种组合的模式都不同于其它组合。

表 7　基于区别特征的双字调音高模式

后字 / 前字	T1（33）	T2（55）	T3（244）	T4（22）	T5（44）	T6（11）	T7（233）	T8（41）
T1（33）	中平+中平	中平+高平	中平+高升平	中平+中低平	中平+中高平	中平+低平	中平+低升平	中平+高降
T2（55）	高平+中平	高平+高平	高平+高升平	高平+中低平	高平+中高平	高平+低平	高平+低升平	高平+高降
T3（244）	高升平+中平	高升平+高平	高升平+高升平	高升平+中低平	高升平+中高平	高升平+低平	高升平+低升平	高升平+高降
T4（22）	中低平+中平	中低平+高平	中低平+高升平	中低平+中低平	中低平+中高平	中低平+低平	中低平+低升平	中低平+高降
T5（44）	中高平+中平	中高平+高平	中高平+高升平	中高平+中低平	中高平+中高平	中高平+低平	中高平+低升平	中高平+高降
T6（11）	低平+中平	低平+高平	低平+高升平	低平+中低平	低平+中高平	低平+低平	低平+低升平	低平+高降
T7（233）	低升平+中平	低升平+高平	低升平+高升平	低升平+中低平	低升平+中高平	低升平+低平	低升平+低升平	低升平+高降
T8（41）	高降+中平	高降+高平	高降+高升平	高降+中低平	高降+中高平	高降+低平	高降+低升平	高降+高降

3.2 双字调基频模式规律

结合上文的分析，从双字组合的基频音高模式中可以总结出以下几条规律：

（1）新寨苗语双字组合没有发生连读变调现象，即，双字组合中既没有产生新的调型，亦没有使得原有调型产生合并。

（2）同一个调类，在其作前字后接不同调类时的音高波动范围小于其作后字时的音高波动范围。也就是说，同一个调类，在其作前字后接不同调类时，其自身变化的方差小于其作后字前接不同调类时的情况。

（3）高降调 41 无论作前字还是后字，其在双字组中的起始音高值都达不到单字调时的高度，音高起始值大致与 33 调一致。

（4）11 调、22 调、244 调和 41 调在作后字时，其前接的高平调 44 和 55 在音高起始处存在交叉情形；但这四个调类作前字时，其后接的高平调 44 和 55 则不存在交叉情形。

（5）升平调 233 和 244 在作前字时，其音高曲线的升平只有在调尾还保持平的特征。也就是说，升平调在作前字时，升和平两段的比例关系相对于单字调发生了变化。

（6）11 调作前字时的音高曲线整体低于作后字时的情况，虽然 11 调和 22 调两个调类间的差值在作前字和后字时有所不同，但整个调类系统的调域基本保持不变。也即，双字调组合中前字的整体调域与后字的整体调域大体保持一致。

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

综上，新寨苗语双字调组合中，相对于前字而言，后字保持单字调的能力更强。

四 讨 论

新寨苗语有八个调类，并且其中五个是平调。从发音和听音的角度看，单单五个平调对非母语者而言就是一大挑战，何况双字组中平调组合竟没有发生合并。

《苗语简志》认为，湘西方言有极简单的连读变调，黔东方言没有连读变调，川黔滇方言连读变调非常复杂。（王辅世 1985:107）从地缘上看，这三大方言的相对位置分处于北部、东部和西部。为什么只有黔东苗语方言没有发生变调现象呢？事实上，《简志》的说法并不全面，例如，川黔滇方言下属的重安江次方言也没有连读变调现象。（王辅世、毛宗武 1995:10）如果撇开前人对苗语方言的分类，只考虑地理位置，重安江次方言的分布区域主要在贵州省凯里和黄平，这与黔东方言的分布区域重合。既然如此，从区域上看，不变调的苗语好像主要集中在东部和南部。但是，孔江平（1992）对川黔滇方言贵阳次方言的紫云苗语研究表明，其五个平调在双音节中都可以相互自由组合，共有 25 个双音节调位，也没有发生连读变调现象。

学界以往对双字调的研究多集中在连读变调的研究上，尽管自主音段音系学给出了具有连读变调的语言或方言中变调出现的条件和规则类型，但是它们并没有对变调与否的深层机制作出解释。本文从发声角度分析了一种双字组未发生变调的语言（新寨苗语），但是，囿于目前的认识，我们还不能解释其不变调背后的机制。关于变调与否的问题，一方面需要考察更多的语言材料和事实，另一方面，还需要从生理、心理、声学和听感等多个角度进行深入地研究。

本文研究表明，新寨苗语有八个单字调：五个平调、两个升平调和一个高降调，其中低平调 11 伴随气嗓音。男女性别差异在单字调上体现为男性发音人在每一个调类上的绝对音高都小于女性发音人，但是男性音高范围大于女性发音人。对双字调的研究表明，新寨苗语在双字调中没有发生连读变调，即双字组共 64 种声调组合模式。另外，在新寨苗语的双字调组合中，相对于前字而言，后字保持单字调调型的稳固性更高。

参考文献

孔江平：《紫云苗语五平调系统的声学及感知研究》，载于马学良主编、中国民族语言学会编《民族语文研究新探》，四川民族出版社，1992 年。

孔江平：《实验语音学基础教程》，北京大学出版社，2015 年。

刘　复：《四声实验录》，群益书社，1924 年。

刘　文、张锐锋：《鱼粮苗语低平调和低降调的声学感知研究》，《语言学论丛》第 54 辑，商务出版社，2016 年。

王辅世：《苗语古音构拟》，国立亚非语言文化研究所，1994 年。

王辅世、毛宗武：《苗瑶语古音构拟》，中国社会科学出版社，1995 年。

王辅世主编：《苗语简志》，民族出版社，1985 年。

张　琨：《苗瑶语声调问题》，《历史语言研究所集刊》第 16 本，1947 年。

中国社会科学院、澳大利亚人文科学院：《中国语言地图集》，香港朗文（远东）有限公司，1987 年。

国家哲学社会科学学术期刊数据库
National Social Sciences Database

国家哲学社会科学学术期刊数据库
National Social Sciences Database

朱晓农、石德富、韦名应：《鱼粮苗语六平调和三域六度标调制》，《民族语文》2012年第4期。

Chang, K. 1953. On the tone system of the Miao-Yao languages. *Language*. 374-378.

Chang, K. 1972. The reconstruction of proto-Miao-Yao tones. *Bulletin of the Institute of History and Philology*. 44(4): 541-628.

Chao, Y. R. 1930. A system of tone letters. *Le maître phonétique*. 45: 24-27.

Kong, Jiangping. 2007. *Laryngeal Dynamics and Physiological Models: High Speed Imaging and Acoustical Techniques*. Beijing: Peking University Press.

Kuang, Jianjing. 2013. *Phonation in Tonal Contrasts*. UCLA dissertation.

Kwan, Julia Chin. 1971. Ch'ing Chiang Miao phonology. *Tsing Hua Journal of Chinese Studies*. New series 9: 289-305.

Maddieson, I. 1978. Universals of tone, in *Universals of Human Language*. 2: 335-365.

Yang, Ruoxiao. 2015. The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics*. 43(1): 453-472.

[**Abstract**] Xinzhai Miao belongs to Qiandong dialect of the Miao language. Based on our fieldwork, this paper makes use of F0 and Duration to analyze the pitch patterns of tones in monosyllabic and disyllabic words. It turns out that Xinzhai Miao has eight tones for monosyllabic words: five level tones, two rising-level tones and one high-falling tone. The distinctive feature of this dialect is that it has five level tones. In spite of this, it is very interesting to note that there is no tone sandhi in disyllabic words, which is also peculiar for a language with five level tones. Building on the analysis, we establish a set of distinctive features in order to describe the pitch patterns of tones in monosyllabic and disyllabic words in Xinzhai Miao. Finally, in connection with other dialects of Miao, we attempt to interpret the lack of tone sandhi in this speech.

[**Keywords**] Xinzhai Miao monosyllabic words disyllabic words F0 Duration distinctive features

（通信地址：刘　文　100871　北京　北京大学中国语言文学系
杨正辉　100081　北京　中央民族大学少数民族语言文学系
孔江平　100871　北京　北京大学中国语言文学系）

【本文责编　胡鸿雁】

国家哲学社会科学学术期刊数据库
National Social Sciences Database

鱼粮苗语低平调和低降调的声学感知研究*

刘　文　张锐锋

提要　基于朱晓农等（2012）对鱼粮苗语声调的研究，低平调和低降调在同一种语言声调感知中的作用至今还没有得到解决。本文以基频、开商和速度商为参数，对该方言的声调调值和发声进行了系统的声学分析，认为鱼粮苗语中存在四个平调、两个降调、一个升平调。文中进一步通过感知实验着重分析了低平调（T5）和低降调（T6）在声调感知中的作用，即在基频较低的情形下，基频的模式（平、升、降）对母语者的感知有无差异。实验结果表明，在鱼粮苗语中，发声对二者的感知起主要作用，基频也起着一定的作用。

关键词　鱼粮苗语　低平调　低降调　感知实验

1　引言

鱼粮位于贵州省黔东南苗族侗族自治州凯里市湾水镇，地处重安江和清水江交汇处。根据《中国语言地图集》（C9），鱼粮苗语属于苗语黔东方言北部土语。朱晓农等（2012）给出了鱼粮苗语的音系，该方言共有七个声调，其中六个是平调（T1、T2、T3、T4、T5、T6），一个是高降调（T7），并说明这是迄今为止语言学界所知道的唯一一个六平调案例。

收稿日期：2015-09-15；**定稿日期**：2015-10-19；**通讯作者**：刘文。

* 本研究得到国家社科基金重大项目“中国有声语言及口传文化保护与传承的数字化方法”（10ZD&125）和“基于中国语言及方言的语言接触类型和演化建模研究”（14ZBD102）资助。感谢石德富老师和学友杨正辉对本文写作的支持，文章投稿后，又蒙《语言学论丛》匿名审稿专家惠赐建议，在此谨致谢忱。感谢《语言学论丛》编辑部为本文所付出的辛劳。文中若有谬误，概由笔者负责。

世界语言中，存在五个平调的语言并不多见（Maddesion 1978）。据笔者统计，在前人报告的23个苗瑶语方言中（王辅世、毛宗武1995），只有“宗地”（贵州省紫云苗族布依族自治县宗地乡绞坨寨，属苗语川黔滇方言麻山次方言中部土语）有五个平调[①]，其他各苗瑶语方言中的平调最多不超过四个。此外，清江苗语（李方桂调查，Kwan 1971整理）和高坝侗语（Shi *et al.* 1987）也存在五个平调的最小对立。

基于类型学材料和苗瑶语族语言声调的特点，我们对鱼粮苗语六平调这一现象产生了极大的兴趣。因此，本文重新采集了鱼粮苗语的声调材料，首先绘制出鱼粮苗语的基频、开商和速度商曲线，确定了声调的调值和发声类型。其次，通过感知实验着重分析了该方言低平调和低降调在声调感知中的作用。

2 鱼粮苗语的调值和发声

声调的感知涉及多种参数，前人往往将声调与基频等同（孔江平 2015：67）。近年来，随着声学、生理和心理研究的进展，学界逐渐意识到仅仅依靠基频无法准确有效地描写声调，尤其当所描写的语言声调系统中涉及特殊发声类型时。孔江平（2001：288—290）给出了一个描写声调的框架：“声调是由调时发声和调声发声共同作用产生的能区别意义的语音现象。”所以，对声调的描写应该涵盖“调时”（指肌肉对声带振动快慢的调制）和“调声”（肌肉对声带振动方式的调制）两部分。对应到声学参数上，“调时”可以用基频来表现，“调声”可以用开商和速度商来表现。在实验语音学中，基频已经成为声调声学分析的常用参数，但是，开商和速度商的概念并未普及，在此需要做个简单介绍。图1是采用微分法（differential）截取的一段喉头仪信号（Electroglottalgraphy，EGG），其基本原理是根据声带打开或关闭时声门阻抗信号的强弱来定义声带的振动方式，当声带关闭时，电阻变小，EGG信号就会增强。当

声带完全关闭时，电阻最小，EGG信号就会出现峰值；反之，当声带打开时，电阻变大，EGG信号减弱。

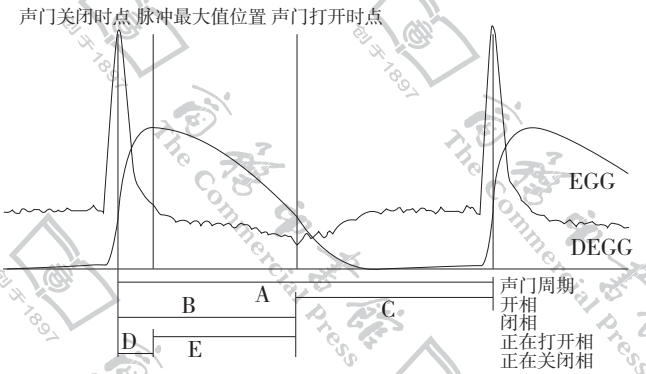


图1 EGG信号微分法示意图

图1EGG波形中，我们把声门阻抗信号微分波形（DEGG）上的最大正峰值在声门阻抗信号（EGG）上的对应位置确定为声门关闭时点，把声门阻抗信号微分波形（DEGG）上的最小负峰值在EGG上的对应位置确定为声门打开时点。A是声带振动的一个完整周期，B是声带关闭相，C是声带打开相，D是声带正在关闭相，E是声带正在打开相。本文基于微分法提取基频、开商和速度商参数。那么，

开商（OQ）=开相（C）/周期（A）×100%

速度商（SQ）=正在打开相（E）/正在关闭相（D）×100%

下面我们使用基频、开商和速度商这三个参数来描写鱼粮苗语的声调系统。

首先，采集录音材料。本文的发音老师为石德富，系中央民族大学少数民族语言文学系老师，母语为鱼粮苗语。录音是在北京大学中文系语音乐律实验室进行的，录音设备有笔记本电脑（ThinkPad E425）、便携式喉头仪（7050A）、驻极体麦克风（SONY ECM-44B）、声卡（SBX）、调音台（XENYX302USB）。录音使用的字词由发音老师提供，具体如下[②]：

表1　录音例词

调类	T1	T2	T3	T4	T5	T6	T7
ta	地	来	长	丢失	弄死	死	翅膀
pa	佩带	排开	雄性	损坏	猪	藤	父亲
ka	炒	拖	饭	矮	挠	鸭	巷
qa	街	斜视	屎	状词	鸣叫	咯痒	锦鸡
ma	妈	拍打	躲藏	马	——	抖动状	砍杀
ɬa	哄骗	漏	狗	——	——	富	乞讨
ɬji	氏族名	疤印	说谎	（碟）浅	撕	脱臼	（鞋）脱落
ki	——	芦笙	路	——	蛋	露出	晴
tu	深	鼓起	冲撞	远	镀	哪	织
pu	栖落	埋	大连襟	刺	坨	抱	——
ɬja	滑	丑	——	突然状	留下	很快状	（撕）乱
maŋ	病	高粱	含	脸/面	——	戴（首饰）	搓揉
tɕaŋ	提把	捞网	果子	弯曲	烤（糍粑）	插秧	禁忌
kaŋ	胸	叠加	——	自从	脆	海碗	剪刀
tɯ	搁	驼	纸	柴火	脚	爆	点火
kɯ	虫	典故	罾	状词	子	——	刺杀
kuŋ	水沟	叶茎	散发	叩击	爪	状词	母鸡护仔
paj	给	掉落	满	拦住	分开	——	猫
pɔ	山	浮萍	宝	状词	报	状词	剥

表1中，每一行的声韵母相同，声调各异；每一列的声调相同，声韵母不同。部分声韵调组合在鱼粮苗语中没有词汇反映形式，表中用“——”标注。考虑到协同发音对声调的影响，本文从上述词表中选取了部分例词作为下文研究的基础。所选例词要求声母为清不送气塞音声母［p、t、k、q］，韵母要求为单元音韵母［a、i、u、ɯ、ɔ］。在此标准下，共选取了［pa、ta、ka、qa、ki、pu、tu、tɯ、kɯ、pɔ］十组例词。由于部分组缺字，所以十组共有66个例词，每个例词读5遍，共计330个样本。

为了给读者一个直观的认识，笔者从录音样本中选取了调类各异但声韵母结构相同的音节［ta］作为例证。

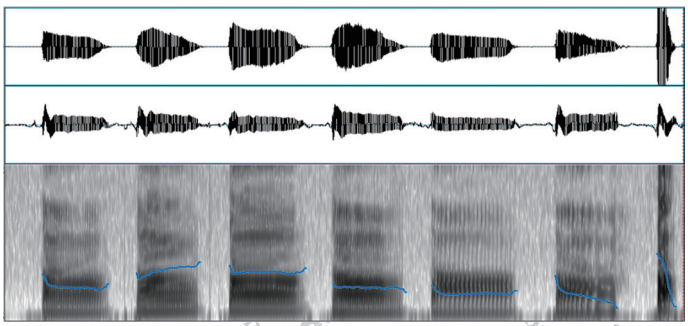


图2 ［tɑ］（声调从左至右依次为T1、T2、T3、T4、T5、T6、T7）

图2由上至下第一栏是语音波形，第二栏是喉头信号，第三栏是三维语图。通过语图中的基频曲线，可以观察到T1、T3、T4和T5的基频曲线都较平，T2略有上升，T6、T7的基频曲线下倾。

其次，我们通过声学分析给出了鱼粮苗语的基频、开商和速度商模式图，目的是确定鱼粮苗语的声调调值和发声类型。绘图的具体操作程序如下：第一步，确定语音波形和声门阻抗信号（EGG波形）的起点和终点。为了使数据的处理具有内部一致性，在选择样本的起点和终点时，我们利用北京大学中文系语音乐律实验室自编的Matlab程序计算出基频曲线、能量曲线以及共振峰结构等参数。起点选在EGG波形图上开始稳定的周期点，终点选在第二共振峰结构稳定、基频曲线平稳、能量恒定的点[③]。第二步，归一化处理。每个例词提取18个F0（基频）、OQ（开商）和SQ（速度商）值，将每个例词读五遍的数据进行平均，然后将同一调类下所辖的所有样本进行平均，并对同一声调中不同字的时长进行归一化处理，目的是为了使F0、OQ和SQ值和时间进行同步关联。第三步，作图。将上一步得到的数据转化成矩阵（F0和时间、OQ和时间、SQ和时间），将X轴定为时间，Y轴分别定为F0、OQ和SQ值，绘制散点图。需要说明的是，在基频曲线的绘制过程中，为了使得到的数据更加接近人们的听感，笔者将F0转换为半音，并将其统筹到“五度制”中。

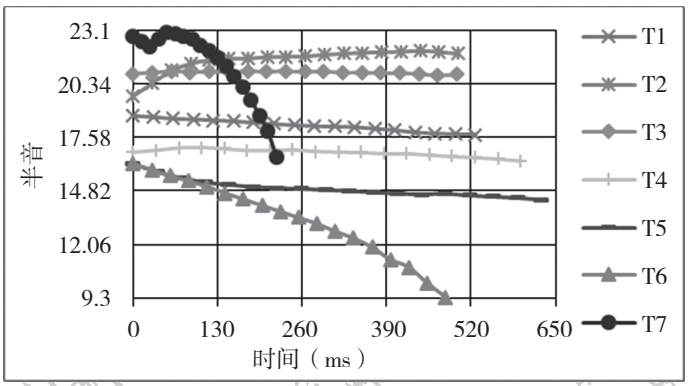


图3　鱼粮苗语基频曲线图

图3进一步证明了我们从图2中观察到的结果，鱼粮苗语的七个声调中包含四个平调（T1、T3、T4、T5），两个降调（T6、T7），还有一个升平调（T2）。根据半音法所得的基频曲线图，如果采用“五度标调法”（Chao 1930），我们可以将它们的调值分别定为53、455、55、44、33、22、31。从基频来看，T2调在发音的前半程（前80ms）基频呈上升趋势，随后是基频稳定段，它与我们通常所见到的升调模式不同。典型升调基频值的上升贯穿整个音节，例如，普通话的第二声（35调），而鱼粮苗语的T2调则展现了另外一种模式，前升后平。至于这两种模式之间的异同，还需另文研究。T7高降调的起点很高，时长较短，在听感上不属于假声，再者它并非一降到底。

低平调（T5）和低降调（T6）的起点相同，所不同的是二者的曲线走势，一个为平，一个为降。朱晓农等（2012）将T6这个低降调也看作平调，并认为在基频较低的情形下，声调曲线的模式对于声调的感知不起作用，即低平、低升、低降、低凹和低凸等都可以看成同一个，统称为“低调”，这也是该文为鱼粮苗语设立六平调的重要前提。低调中，声调曲线的不同模式对于母语者的感知是否有区别意义的作用？为了回答这一问题，下文第三部分设计了一项合成听辨实验来测试低平调和低降调在母语感知中的差异。

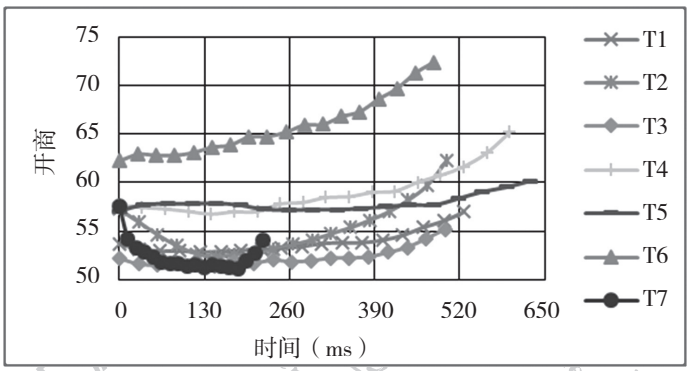


图4 鱼粮苗语开商曲线图

图4给出了鱼粮苗语声调的开商模式，一个比较明显的特点是各个声调的开商在音节的末尾都有上扬趋势，只是程度有所不同。T1调和T3调的的开商走势一致，二者的不同在于T1调的开商值略大于T3调，二者的开商值大多处于52—55之间。T2调的开商和T1调起点相同，其在前200ms的范围内呈下降趋势，数值由57降至53，后半段呈上升趋势，数值由53升至63。T4调的开商从57升至65，其尾部有些气嗓音质。T5调的开商整体较平，数值在57左右，末尾升至60。T6调的开商最大，数值从62上升到73左右，远大于正常嗓音的数值（55左右），这说明母语者在发音时，声带绝大部分都是打开的，由于声带漏气导致声带不能完全闭合。T7调的开商除了在末尾有些上扬外，其他部分都呈下降趋势。根据孔江平（2001：172）对开商的研究，“开商是个比较稳定的值，其总的平均为54.8%，也就是说声带开合的一个周期，通常是开相稍大于闭相。可以这样说，在这个值两端一定的范围内都是正常嗓音。”因此，我们可知，除了T6调属于特殊发声外，其他声调都属于正常嗓音。

图5是鱼粮苗语的速度商模式图。从中可以观察到，速度商的曲线与基频曲线有很大的关联，平调的速度商同基频曲线一样，基本呈平势，例如，T1调、T3调、T5调，T4调的速度商有些下降。T1调的速度商最大，T5调次之，T3调、T4调紧随其后。T2调的基频

呈升平状态，而速度商是下降状态，二者呈反向关系。T7调的基频呈高降趋势，而速度商正好与此相反，呈上升姿态。T6调的速度商与开商正好相反：开商呈上升趋势，而速度商呈下降趋势。T7调的情形与此相反，开商下降，速度商上升。

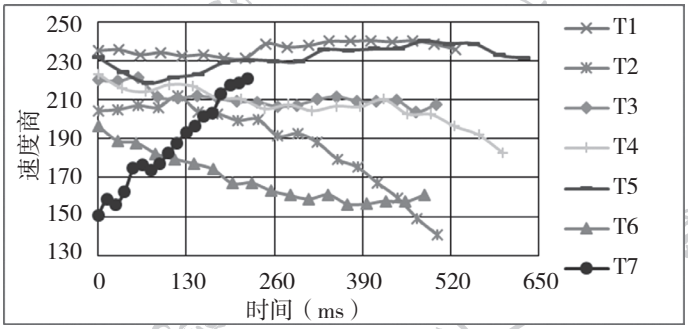


图5　鱼粮苗语速度商曲线图

3　低平调与低降调的合成、感知实验

本小节我们重点讨论低平调（T5）和低降调（T6）在整个声调感知空间中的作用。由于我们的目的是观察基频在低平和低降调中的作用，所以采用基音同步叠加法（PSOLA）来合成实验样本，其调节音高的原理是：基音同步分析，以音高标记为中心，两倍的新基音周期为窗长，加时间窗；截去分析窗以外的部分，保留分析窗以内的部分，按半个窗重叠，重新排列信号；叠加，得到新基音周期。（转引自鲍怀翘、林茂灿 2014：494）

根据上述原理，我们在Praat软件中进行样本的合成。首先从低平调（T5）合成到低降调（T6）。具体做法是：将母语者的原声低平调（T5）均匀分九步通过只改变基频的方式合成到低降调（T6），得到10个样本，其中第一个是原声，其他九个是合成样本。同理，从低降调（T6）合成回低平调（T5）也可以得到10个样本。这样一来一回，我们就可以得到20个样本，合成示意图见图6。本

文我们使用了七组不同的声韵母结构，即［pa ta ka qa ki pu tu］，每组合成20个样本，由此，一共得到140个合成样本。根据基音同步叠加法（PSOLA）的原理可知，虽然我们改变了音节的基频模式，但基本没有影响到声源的发声类型，这就使得我们在保持发声类型不变的前提下进行测试基频对声调的感知贡献成为可能。

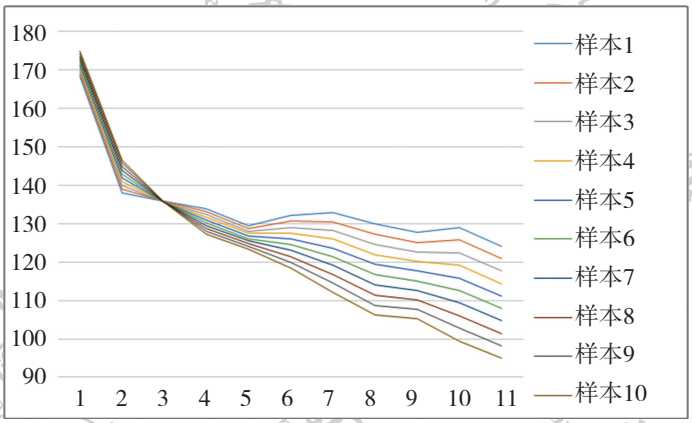


图6 ［ta］音节低平调（T5）和低降调（T6）合成样本示例

得到合成样本后，接下来的工作就是要对母语者进行听觉实验，一般采用“选择适应”（selective adaptation）的实验，实验分为两个部分，一是辨认（identification），即每次随机给出一个样本，让被试指出它是两个目的词中的哪一个，两者必选其一；一是区分（discrimination），采用ABX程序，每次让被试听三个音节，第一个音节A和第二个音节B总是不同的，第三个音节一定是前两个音节中的一个（A或B），被试的任务就是回答第三个音节X到底是A还是B。本文只采取辨认实验，一是因为辨认实验和区分实验的目的都是为了寻找音位界限，二者的功用相同；二是因为在使用“区分测验”时，比较容易受到心理物理学的影响（孔江平 1995）。

听辨时，我们采用DMDX脚本呈现刺激，让它控制每组中的20个合成样本。实验开始后，在电脑上随机播放20个样本，每个样本以1秒的间隔连续播放三次后，电脑屏幕上会立刻出现A、B两

个选项供被试判断，被试有5秒的时间给出答案，并且两者必选其一，不可漏选。参加听辨实验的是鱼粮苗语母语者——中央民族大学少数民族语言文学系教师石德富。

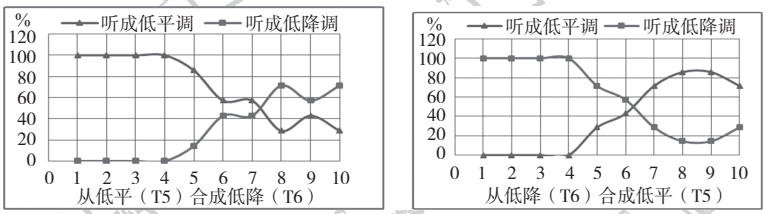


图7　低平调（T5）和低降调（T6）听辨结果

听辨结果如图7所示，鱼粮苗语低平调（T5）和低降调（T6）之间存在着显著的范畴感知边界，即在范畴内部无论如何改变基频值都只能被感知为同一个声调，而一旦跨越边界就会被感知为另外一个声调。从图7可知，由低平调（T5）合成到低降调（T6）时，感知边界在左起第7个和第8个样本之间（偏向第7个样本）。同理，由低降调（T6）合成回低平调（T5）时，感知边界却落在左起第6个样本和第7个样本之间（偏向第6个样本），这两种情形下的感知边界没有落在同一个点上，而是彼此跨越，并且跨越多达两个样本，如图8所示。

图8　低平调（T5）和低降调（T6）感知边界跨越示意图

是什么导致了这种感知跨界现象？我们从音强、音长、音质、音高和发声这些参数入手进行分析。首先，用于合成的每组样本都是同一个发音人在相对稳定的状态下发出的，并且根据基音同步叠加法（PSOLA）合成的原理，在合成过程中并不改变声音样本的振幅，因此可以排除音强这一因素。其次，从基频曲线图可知，归

一化后的低平调（T5）的时长要长于低降调（T6），是否是时长导致了这种感知的错位呢？答案同样是否定的，因为在低平调（T5）和低降调（T6）的合成过程中，已经将时长这一因素考虑在内。合成采用目标调的时长，即，从低平调（T5）合成低降调（T6）时，采用低降调（T6）的时长；从低降调（T6）合成低平调（T5）时，采用低平调（T5）的时长。感知测试的结果也证明了时长在二者的感知中不起作用。再者，我们也可以排除音质这一因素，因为音质主要由共振峰结构决定，由于每组样本都采用相同的声韵母结构，所以音质也不是造成感知错位的原因。排除完音强、音长和音质三个要素后，我们接下来考察音高这一要素。根据上文所知，基音同步叠加法（PSOLA）合成是在保持其他条件不变的情形下只改变基频以期达到合成语音的目的。低平调（T5）和低降调（T6）的合成正是由基频的改变而来，所以感知错位现象的出现一定与基频有关联，但其又不是唯一决定性因素。因为如果在听辨中只有基频起作用，那么听辨的结果就不会出现边界错位现象，而应该是边界重合（张锐锋、孔江平 2014）。笔者在与被试交流中得知，听辨实验中，被试感觉每组20个样本中总有五六个无法判断是低平调（T5）还是低降调（T6），这也说明基频在二者的感知中是起作用的，因为合成实验设计只改变了基频这一参数。

考虑完上述几个因素后，接下来我们分别从听感、EGG波形对比以及开商、速度商三个方面分析发声这一参数。

听感上：笔者在录音过程中，明显感知到母语者在发低降调（T6）时，喉头带有强烈的摩擦，也即我们所说的气嗓音（breathy voice）。

EGG波形对比：为了更加清晰地呈现低平调（T5）和低降调（T6）具有不同的喉头波形，我们引入了属于正常发声模式的中平调（T1）。图9显示低平调（T5）和中平调（T1）属于同一种发声模式，即开相和闭相的时长几乎相等。而低降调（T6）则明显不同，其开相远大于闭相，这说明在发低降调（T6）时，声带存在着

严重的漏气现象。

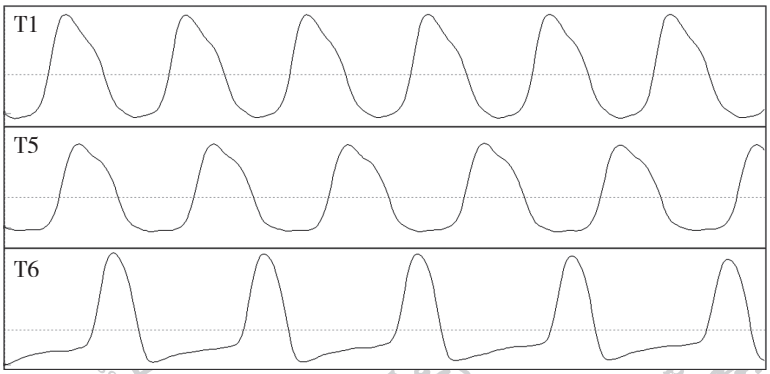


图9　T1、T5、T6调EGG波形对照

开商、速度商参数：开商决定着声源第一谐波的相对强度，可以通过改变开商来改变第一谐波的振幅。开商高，则h1-h2的值也高，也就是说，相比于第二谐波振幅及嗓音频谱上更高次成分的振幅，嗓音基频的幅度h1就更见突出（Fant and Kruckenberg 2004）。因此，h1-h2的值标示着声门气流开相的相对时长，对应于声门气流波形的开商。发气嗓音的时候，声门在其大部分长度上都保持一定程度的打开状，两条声带从来都不在声门中线处相遇，因此有大量的漏气，这就导致气嗓音的第一谐波很强。也可以说开商越大，气嗓音音质越明显。结合图10鱼粮苗语的情况，低降调（T6）的开商明显大于低平调（T5）。从图11的速度商来看，情况正好相反，低降调（T6）则明显小于低平调（T5）。这与Kong（2007：95）利用高速数字成像技术得出的结论一致：气嗓音具有很大的开商和相对较小的速度商。声门气流波形的下降段越是陡峭，声门从开始关闭到完全闭合所用的时间就越短，关闭的速度就越快，速度商也就越大，则此时声门脉冲频谱上的高频谐波也就越强。相反，声门气流波形越是对称，则低频谐波的振幅就显得越强（方特1988）。因此，减小速度商对嗓音频谱的影响就是导致频谱在高频处向下倾斜（Fant and Kruckenberg 2004）。在一个正常的声门脉冲

中，速度商一般为200%。

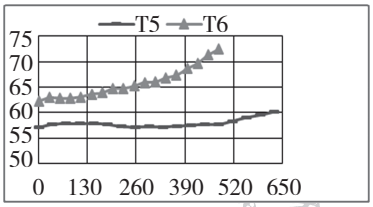


图10 T5、T6的开商模式

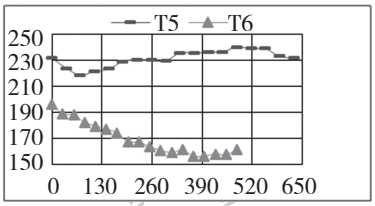


图11 T5、T6的速度商模式

孔江平（2001：188）给出的嗓音区别性特征矩阵，低降调（T6）的音调和速度商低于正常嗓音，而开商大于正常嗓音。由此，我们可以比较肯定地说低降调（T6）属于气嗓音这一发声范畴。所以，特殊发声在低平调（T5）和低降调（T6）的感知中也起到了一定的作用，但必须说明的是，特殊发声方式也不是导致感知边界错位的唯一因素。如果发声是决定性因素，那么，我们在合成时所采取调整基频的策略应该是没有任何作用的，因此感知的结果会呈现两条平行的直线，而事实并非如此。

由此我们可知，低平调（T5）和低降调（T6）之所以会出现感知错位的现象，一方面是二者具有不同的基频，另一方面，这与二者具有不同的发声方式也有一定的关联。至于发声和基频在感知中的权重，还需要进一步的研究。

4 结论

本文首先通过声学分析给出了鱼粮苗语七个声调的调值：四个平调（55、44、33、22）、两个降调（54、31）、一个升平调（455）。这与朱晓农等（2012）所分析的六个平调和一个降调的结果不同，二者的区别在于对升平调（T2）和低降调（T6）的处理上。文章后半部分重点讨论了低平调（T5）和低降调（T6）在同一种语言声调感知中的作用。鱼粮苗语的合成感知实验表明，低平调（T5）和低降调（T6）的区分是发声和基频二者共同作用的结

果，也即，在声调基频较低的情形下，基频模式的走向（平、升、降）对母语者的感知是有影响的。这也表明，我们不能笼统地将含有不同基频模式的低调统一处理为某种形式。因此，就目前的认识来看，世界范围内尚不存在六平调的语言。

另外，观察四个平调在声调感知空间中的分布，可以发现一个很有意思的现象：T5（22调）和T4（33调）分别处于“五度制”中第二度和第三度的顶端，T1（44调）和T3（55调）分别处于第四度和第五度的底部。对于平调因基频大小在声调空间中分布差异的一种推测性解释是母语者为了平衡发音的经济性和感知的效能性所做出的适应性调整。当然，这一问题还需进一步深入研究。

附　注

①　紫云县界牌村也有五个平调，除汉语借词外，该方言只有五个调值，并且都是平调。（孔江平 1992；王辅世、刘援朝 1993）

②　调类是发音老师参照王辅世（1994）《苗语古音构拟》标注。

③　孔江平（2015：61）给出了另外两种方案：一是根据振幅量曲线，将振幅曲线做微分，微分后的振幅信号会出现两个尖的峰值，这两个点之间的基频就是要保留的基频数据。二是根据振幅曲线，设定一个尺度来确定基频。

④　半音的计算公式为：半音$S=12\times\log_2 F/F_0$，参考频率F0为64.66Hz，转自沈炯（1985：86）《北京话声调的音域和语调》，载于林焘、王理嘉等著（1985）《北京语音实验录》。

⑤　T5在五度制中介于第二度的顶部和第三度的底部，其基频的起点值和31相同，因为其大部分处于第二度中，所以本文将其处理为22。

参考文献

鲍怀翘、林茂灿　（2014）《实验语音学概要》（增订版），北京大学出版社，北京。

方　特　（1988）嗓音声源研究，《声学学报》第13卷第2期，135—146页。

孔江平　（1992）　紫云苗语五平调系统的声学及感知研究，马学良主编、中国民族语言学会编《民族语文研究新探》，152—163页，四川民族出版社，成都。

孔江平　（1995）　藏语（拉萨话）声调感知研究，《民族语文》第3期，56—64页。

孔江平　（2001）《论语言发声》，中央民族大学出版社，北京。

孔江平 （2015）《实验语音学基础教程》，北京大学出版社，北京。
林焘、王理嘉等著 （1985）《北京语音实验录》，北京大学出版社，北京。
沈 炯 （1985） 北京话声调的音域和语调，林焘、王理嘉等著《北京语音实验录》，北京大学出版社，北京。
王辅世 （1994）《苗语古音构拟》，国立亚非语言文化研究所，东京。
王辅世、刘援朝 （1993） 贵州紫云界牌苗语的语音特点和方言归属，《语言研究》第1期，165—182页。
王辅世、毛宗武 （1995）《苗瑶语古音构拟》，中国社会科学出版社，北京。
张锐锋、孔江平 （2014） 河南禹州方言声调的声学及感知研究，《方言》第3期，206—214页。
中国社会科学院、澳大利亚人文科学院 （1987）《中国语言地图集》，香港朗文（远东）有限公司，香港。
朱晓农、石德富、韦名应 （2012） 鱼粮苗语六平调和三域六度标调制，《民族语文》第4期，3—12页。
Boersma, Paul, and David Weenink （2013） Praat: Doing Phonetics by Computer（version 5.3.52）. www.praat.org.
Chao, Y. R. （1930） A System of Tone Letters. *Le maître phonétique* 45: 24-27.
Fant, G., and A. Kruckenberg （2004） An Integrated View of Swedish Prosody, Voice Production, Perception and Synthesis. In Gunnar Fant, *Speech Acoustics and Phonetics*, Kluwer Academic Publishers, 249-300.
Kong, Jiangping （2007） *Laryngeal Dynamics and Physiological Models—High Speed Imaging and Acoustical Techniques*. Peking University Press.
Kwan, Julia Chin （1971） Ch'ing Chiang Miao Phonology. *Tsing Hua Journal of Chinese Studies*. New series 9: 289-305.
Maddieson, Ian （1978） Universals of Tone. *Universals of Human Language*, Vol. 2: 335-365.
Shi Feng, Shi Lin and Liao Rongrong （1987） An Experimental Analysis of the Five Level Tones of the Gaoba Dong Language. *Journal of Chinese Linguistics* 15（2）: 335-361.

An Acoustic and Perceptual Study on the Low Level Tone and Low Falling Tone of Yuliang Miao

LIU Wen, ZHANG Ruifeng

Abstract: According to Zhu et al.（2012）, the mechanisms involved

in the perception of low-level and low-falling tones remain understudied. We therefore undertake an acoustic study on tonal categories and phonation of Yuliang Miao, investigating three parameters, i.e., pitch, open quotient, and speech quotient. The result shows that this language has seven tones: four level tones, two falling tones and one rising-level tone. In order to investigate whether it is true that native speakers cannot differentiate pitch patterns（level, rising or falling）at very low pitch, we undertake a perceptual experiment on the low-level and low-falling tones. The results suggest that while phonation types serve as the primary cues for the perception of these tones, pitch still plays a not so significant but indispensable role in it.

Keywords: Yuliang Miao, low level tone, low falling tone, perceptual experiment

（刘文：北京，北京大学中国语言文学系　liuwen3214@163.com；
张锐锋：新乡，河南师范大学外国语学院　ruifeng_72@163.com）

ISSN 1000-0054 清华大学学报（自然科学版） 2017 年 第 57 卷 第 9 期 12/19
CN 11-2223/N J Tsinghua Univ (Sci & Technol), 2017, Vol. 57, No. 9 963-969

语速对普通话音段产生的影响

李英浩[1]， 孔江平[2]

（1. 延边大学 外国语学院，延吉 133001；2. 北京大学 中国语言文学系，北京 100871）

摘 要：语速影响音段的语音实现以及相邻音段发音动作的时间关系，然而对普通话音段产出的影响却并不清楚。该文使用动态电子腭位方法分析语速对普通话跨音节辅音丛/ŋ#C2/中音段发音动作的时间关系以及 V1#C2V2 中音段的生理和声学特征的影响。研究发现：辅音丛音段发音动作的绝对时差和动作重叠度随语速的加快而增大；音段的生理/声学时长与语速线性相关，语速对辅音动作幅度的影响与辅音发音部位和方式有关；元音动作幅度减弱，在声学上表现出央化的趋势。研究结果表明：普通话语速控制与音段发音动作的时间关系以及音段动作的时空特征有关，音段发音属性和发音人发音策略也影响不同语速条件下的音段产出。

关键词：语速；音段产生；动态电子腭位；普通话

中图分类号：H017 **文献标志码**：A

文章编号：1000-0054(2017)09-0963-07

DOI：10.16511/j.cnki.qhdxxb.2017.26.048

Effects of speech rate on segment production in Standard Chinese

LI Yinghao[1], KONG Jiangping[2]

(1. Foreign Languages College, Yanbian University, Yanji 133001, China;
2. Department of Chinese Language and Literature, Peking University, Beijing 100871)

Abstract: The phonetic realization of speech segments and the timing relation of the articulations of contiguous segments vary for different speech rates. How the speech rate affects the segmental articulations in Standard Chinese is still not very understood. This paper studies the effect of the speech rate on the segment production in Standard Chinese using electropalatography to focus on the articulatory timing relation in heterorganic consonant clusters /ŋ#C2/ and segment articulations in V1#C2V2. The results show that the absolute latency and the gestural overlap of the consonant clusters increases with faster speech rates. Significant linear relations were found between the articulatory or acoustic length for the segments. Effect of the speech rate on the consonantal displacement depended on the place and manner of the articulation. Gestural weakening of vowels was found accompanied by acoustic centralization. The results indicate that the articulatory mechanisms for the speech rate control are associated with the timing relation of contiguous gestures and the temporal-spatial properties of the segments. The articulatory attributes of consonants and individual strategies also affect the segment production for different speech rates.

Key words: speech rate; segment production; electropalatography; Standard Chinese

语速影响音段的语音实现以及相邻音段发音动作的时间关系。通过考察不同语速条件下发音器官的运动特征，不仅可以建立正常发音的语音产出模型，同时也可以研究病理语音产出的根源[1]。

语速一般用单位时间内的音节数来衡量，有 2 种计算方法：一种把静音停顿纳入到计算中；另外一种不包括静音停顿，又称为音速。研究结果发现，语速控制在声学上主要表现为音段/音节时长随语速的加快而缩短，音段时长缩短的幅度与音段属性有关，辅音时长缩短的程度小于元音[2]，送气音的浊音起始时间的缩短幅度要大于同部位的不送气塞音[3]。语速加快使得元音发音不到位或者元音声学空间缩小[4-5]，但也有研究表明，语速不影响元音的频谱特征[6]。

语速对发音器官运动特征影响的研究集中在 3 个方面[7]：1）发音器官运动最大速度以及速度变化曲线；2）运动幅度（或位移）；3）不同发音器官运动的时间关系。在慢速条件下，发音器官的速度曲线表现为非对称、多峰值的变化，而在快速条件下表现为对称的、单峰值的变化[8]。动态电子腭位（electropalatography，EPG）语音研究发现[9]，语速加快使得音段的动作时长缩短，但并不一定伴随着

收稿日期：2016-05-27

基金项目：教育部人文社会科学研究项目（11YJAZH055）；
吉林省教育厅“十三五”社会科学研究项目
（吉教科文字 2016 第 3 号）

作者简介：李英浩（1973—），男，教授。

通信作者：孔江平，教授，E-mail：jpkong@pku.edu.cn

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

动作幅度的减弱。语速对不同发音器官动作的时间关系的影响没有一致的结论[7]，然而，生理研究发现语速对不同发音器官动作的绝对时差(absolute latency)有影响，相邻音段的发音动作重叠程度则取决于音段的属性和语音环境[9]。

汉语普通话是声调语言。语速影响音节或音段的声学和生理特征，在语速较快的条件下，普通话音节/音段的声学时长缩短，缩短程度与其在语流中的韵律位置有关[2,10]；元音和辅音声学时长缩短的程度存在差异[2]。随着语速的加快，辅音发音动作的生理时长缩短，动作幅度减弱，相邻音段发音动作的绝对时差缩短，动作重叠度增大[11]。本文使用62个电极的EPG考察语速对普通话音段产出的影响，从语音生理和声学角度分析语速对普通话音段产出的影响。

1 方法

1.1 语料设计

EPG数据来自北京大学语言学实验室建立的2位发音人(男性XS，女性JZ)的汉语普通话动态电子腭位数据库。语料分为2类：1) 跨音节C1#C2辅音丛(#为音节边界)，设计了8个V1C1#C2V2双音节，C1为软腭鼻音/ŋ/，C2为齿龈塞音/t/和舌尖前擦音/s/，V1和V2为/i/或/a/；2) V1#C2V2双音节，把目标元音/i，ɿ，ɻ/分别置于双音节前字和后字韵母位置上，设计了11个双音节/ʂɿ#pV2/、/pV1#ʂɻ/、/sɿ#pV2/、/pV1#sɿ/、/pi#pa/、/pi#pi/和/pa#pi/(V1为/i/或/a/)。把辅音音段/t，s，ʂ/置于后字声母位置，设计了4个双音节/pa#ta/、/pa#ti/、/pa#sa/和/pa#ʂa/。JZ缺少部分辅音的语料，双音节语料置于负载句“我说___这个词”。发音人语速控制参考文[8]，设计了慢速、正常和快速3种语速条件。每种语速条件下每个句子连续读3或4遍。使用负载句的头2个音节韵母时长的平均值作为语速参数(speech rate，SR)，SR越大，说明语速越慢。

1.2 生理和声学参数

对辅音的EPG关键帧进行定义。把假腭前3行电极定义为齿龈区域，该区域的舌腭接触与/t，s/的舌尖发音动作有关。舌尖动作起点定义为该区域电极开始接触的时刻，动作目标(幅度)定义为辅音持阻段在齿龈区域的最大舌腭接触比率，首帧目标腭位的定义参照文[12]，动作结束点定义为齿龈区域舌腭接触消失前最后一帧的时刻。舌尖辅音后接高元音/i，ɿ，ɻ/的时候，在元音时段齿龈区域也有接触，此时舌尖动作结束点定义为元音声学起点前的腭位帧时刻。舌尖后擦音/ʂ/的成阻部位位于齿龈后到硬腭中的部位，此时选择整个假腭区域并定义关键帧时刻。

假腭后2行中部2列电极定义为软腭区域，这个区域与软腭鼻音/ŋ/的舌体动作有关。如果软腭区域存在舌腭接触，则可以定义舌体动作的EPG关键帧时刻。如果软腭区域没有接触，则采用声学标准，以语图中元音和鼻音的边界点作为舌体动作的起点时刻，声学结束点作为舌体动作结束时刻，此时无法获得舌体动作的运动幅度。

C1#C2辅音丛的参数有2类：1) C1舌体动作和C2舌尖动作各自的参数，包括动作幅度(displacement，DIS)和生理时长(length，LEN)。DIS为音段最大舌腭接触比率，LEN为动作起点和结束点的间距。C1和C2的2个参数分别用$C1_{DIS}$和$C1_{LEN}$以及$C2_{DIS}$和$C2_{LEN}$表示。2) C1和C2发音动作时间关系参数。时差参数包括C1和C2发音动作起点的间距△OSS和发音动作目标的间距△PKS。发音动作重叠参数为C2发音动作在C1动作时域内与C1的重叠度$C1_{OL}$，即C1和C2动作重叠时段与C1动作时段的比率。

在V1#C2V2语料中，C2的参数包括动作幅度$C2_{DIS}$、生理时长$C2_{LEN}$、成阻段时长$C2_{APP}$、接触重心(center of gravity，CoG)、趋中指数(contact centrality，CC)以及擦音的谱重心和谱偏度。$C2_{APP}$为舌尖动作起点和首帧腭位目标时刻的间距。CoG和CC的计算方法分别见文[13]和[14]，擦音声学参数计算方法见文[15]。元音的参数包括DIS、CoG、CC、元音声学时长vLEN和元音声学中点的共振峰F1和F2。元音的DIS定义为元音声学时段中点附近连续3帧腭位中舌腭接触最大帧的接触比率，CoG和CC也基于该帧腭位进行计算。

1.3 统计分析

本文考察SR与音段生理和声学参数是否存在线性关系，使用线性回归的分析方法，计算回归方程的回归系数和确定系数R^2。如回归方程显著，则说明参数与语速之间存在线性关系；确定系数越大，则说明语速和参数的线性关系越明显。

2 结果与分析

2.1 辅音丛发音动作时间关系

使用C1#C2语料，以SR为自变量，7个参数

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

为因变量，对 2 类辅音丛分别进行线性回归分析。表 1 是具有显著性的回归方程的 2 个系数，括号外的数值表示确定系数，括号内的数值表示回归系数，“n. s.”表示回归方程不显著。

表 1 辅音丛回归方程的确定系数 R^2 和回归系数

参数	发音人	辅音丛	
		/ŋ#t/	/ŋ#s/
ΔOSS	XS	0.76[a](0.4)	0.79[a](0.6)
	JZ	0.43[a](0.5)	0.44[a](0.6)
ΔPKS	XS	—	—
	JZ	0.43[b](0.4)	0.56[a](0.7)
$C1_{OL}$	XS	0.44[a](−1.9)	0.31[a](−1.4)
	JZ	0.50[a](−2.1)	0.42[a](−1.9)
$C1_{DIS}$	XS	—	—
	JZ	n. s.	n. s.
$C1_{LEN}$	XS	0.80[a](0.4)	0.85[a](0.7)
	JZ	0.20[b](0.4)	0.40[a](0.6)
$C2_{DIS}$	XS	0.10[c](51.4)	n. s.
	JZ	0.25[a](81.9)	n. s.
$C2_{LEN}$	XS	0.41[a](0.4)	0.26[a](0.3)
	JZ	0.21[b](0.3)	n. s.

显著性水平：$p<0.001^a$，$p<0.01^b$，$p<0.05^c$。

由表 1 可知，SR 与 C1 舌体和 C2 舌尖动作的 ΔOSS 线性相关，ΔOSS 随语速的加快(SR 减小)而缩短。ΔPKS 数据来自 JZ 的前字韵母为/iŋ/条件下的样本，与 SR 也线性相关，ΔPKS 随语速的加快而缩短。$C1_{OL}$ 与 SR 线性相关，随着语速的加快，$C1_{OL}$ 增大，说明辅音丛的协同发音程度增大。

图 1 是 SR 与 ΔOSS 以及 $C1_{OL}$ 的线性关系示意图，图中的回归方程是基于 2 位发音人数据得出的结果。由图 1a 可知，在相同条件下，/ŋ#s/的 ΔOSS 大于/ŋ#t/的情况。此外，前者的 ΔPKS 也大于后者的。由图 1b 可知，在相同条件下，/ŋ#s/的 $C1_{OL}$ 小于/ŋ#t/的情况。

下面分析 C1/ŋ/和 C2/t，s/的动作参数。$C1_{DIS}$ 来自 JZ 的前字韵母为/iŋ/条件下的样本。由表 1 可知，SR 与 $C1_{DIS}$ 无关，但是与 $C1_{LEN}$ 线性相关。

图 2 是 SR 与 $C2_{DIS}$ 以及 $C2_{LEN}$ 的线性关系示意图。可以看出，语速对 $C2_{DIS}$ 的影响与 C2 的发音方式有关，语速对塞音 $C2_{DIS}$ 影响较大，但是擦音受影响的程度较低。由图 2a 可知，在快速条件下，/t/的 $C2_{DIS}$ 降低，说明出现了动作减弱的趋势；/s/的

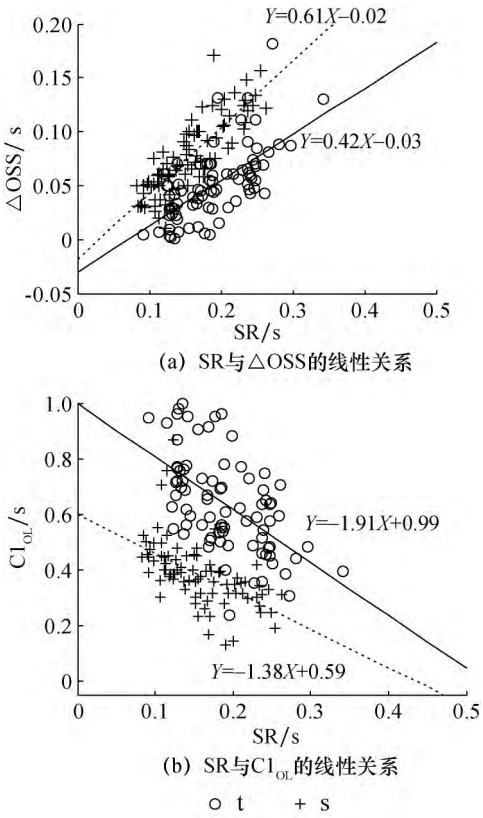


图 1 SR 与时间关系参数的线性关系

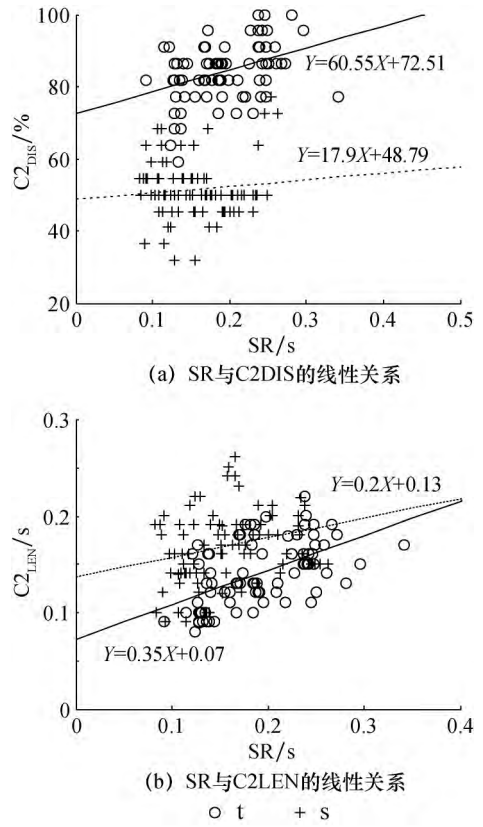


图 2 SR 与 C2 动作参数的线性关系

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$C2_{DIS}$基本不变。SR 与 $C2_{LEN}$ 线性相关，然而 JZ 的/s/的 LEN 不受语速影响。由图 2b 可知，在相同条件下，/s/比/t/的 LEN 更长；在快速条件下，/t/的 LEN 下降较快，/s/的变化则相对较慢。

上述结果表明：1）语速对辅音丛的生理控制机制表现为音段发音动作时间关系的变化，SR 与 ΔOSS 以及 ΔPKS 呈现正相关，与 $C1_{OL}$ 负相关；这种关系受到 C2 发音方式的影响，擦音/s/条件下的时差参数较大，重叠度相对较小；2）SR 和 $C2_{LEN}$ 呈现正相关；快速条件下 C2 动作幅度是否减弱取决于辅音的发音方式，擦音的动作幅度受影响较少，塞音的动作幅度易于减弱；3）C1 舌体动作幅度不受语速影响，但是 SR 与 LEN 呈现正相关。

2.2 语速对辅音发音动作和声学特征的影响

使用 V1＃C2V2 语料，以 SR 为自变量，C2 的 7 个参数为因变量进行线性回归分析。表 2 是具有显著性的回归方程的确定系数和回归系数，图 3 和 4 分别为 $C2_{DIS}$ 以及 $C2_{LEN}$ 与 SR 的线性关系。

表 2 C2 参数回归方程的确定系数 R^2 和回归系数

参数	发音人	辅音		
		/t/	/s/	/ʂ/
$C2_{DIS}$	XS	0.72^{a}(291.5)	0.09^{d}(59.0)	0.31^{b}(32.0)
	JZ	—	n. s.	0.44^{c}(−86.9)
$C2_{CoG}$	XS	n. s.	n. s.	n. s.
	JZ	—	0.78^{a}(3.9)	0.27^{d}(2.9)
$C2_{CC}$	XS	0.60^{a}(0.7)	n. s.	n. s.
	JZ	—	n. s.	n. s.
$C2_{LEN}$	XS	0.45^{a}(0.4)	0.12^{d}(0.2)	0.60^{a}(0.4)
	JZ	—	0.95^{a}(0.9)	0.82^{a}(0.8)
$C2_{APP}$	XS	n. s.	0.32^{b}(0.3)	0.23^{b}(0.2)
	JZ	—	n. s	n. s.
谱重心	XS	—	n. s.	0.23^{b}(3 206)
	JZ	—	n. s	n. s.
谱偏度	XS	—	n. s.	0.28^{b}(−2.3)
	JZ	—	n. s	n. s.

显著性水平：$p<0.001^{a}$，$p<0.01^{b}$，$p<0.05^{c}$，$p<0.1^{d}$。

由表 2 可知，3 个辅音的动作幅度受语速影响的结果不一致。/t/的 $C2_{DIS}$ 与 SR 正相关；/s/的 $C2_{DIS}$ 受语速影响的程度相对较低，XS 的 SR 与 $C2_{DIS}$ 的关系只表现为趋势效应（$p=0.09$），JZ 的 $C2_{DIS}$ 与语速无关。SR 与/ʂ/的 $C2_{DIS}$ 线性相关，但是存在个体差异：XS 表现为正相关，在快速条件下/ʂ/的舌面与假腭的电极接触数目下降；JZ 则表现为负相关，进一步的分析发现，在快速条件下 $C2_{DIS}$ 上升主要与假腭后 4 行接触电极数目增多有关，前 4 行接触电极数目不受语速影响。这说明元音对擦音的协同发音幅度增大。

从图 3 可知，XS 产出的塞音和擦音的 SR 与 $C2_{DIS}$ 的关系存在差异：塞音/t/的 $C2_{DIS}$ 变化幅度较大，在快速条件下，$C2_{DIS}$ 快速降低；/s，ʂ/的 $C2_{DIS}$ 变化范围相对较小。

CoG 和 CC 反映音段舌腭接触模式的细节特征。由表 2 可知，JZ 产出的 2 个擦音的 CoG 与 SR 存在正相关关系，SR 与/ʂ/的 CoG 的关系呈现出趋势效应（$p=0.07$）。在快速条件下，CoG 的降低主要与假腭后部接触电极数目增多有关。

上述结果说明：不同语速条件下舌尖擦音的主发音器官动作相对稳定；在快速条件下，由于元音协同发音的增强，舌尖后擦音/ʂ/不参与成阻的舌面上抬。

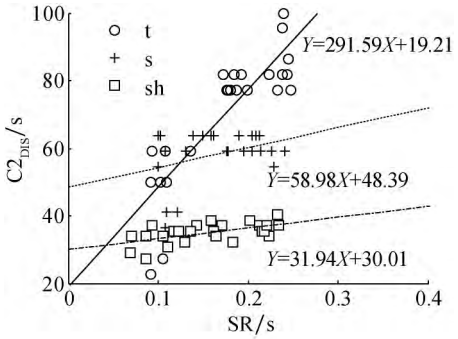


图 3 发音人 XS 的 SR 与 $C2_{DIS}$ 的线性关系(图中 sh 表示/ʂ/)

由表 2 可知，XS 产出的/t/的 CC 与 SR 线性相关。在快速条件下，由于塞音出现了舌尖齿龈不完全闭合的情况，因此 CC 明显降低。

由 XS 擦音的结果可知，SR 与 $C1_{APP}$ 线性相关，这说明在较快语速条件下，擦音的舌尖动作能够在较短的时间内到达指定位置。进入到持阻段后，舌尖动作相对稳定，表现为舌腭接触面积基本不变；在慢速条件下，持阻段的舌尖动作存在调整过程。这说明在快、慢速条件下擦音的动作策略存在差别，这与文[8]的结果一致。舌尖塞音的 $C1_{APP}$ 不受语速影响，原因在于 $C1_{APP}$ 非常短（0.01～0.04 s 之间）。SR 和 $C2_{LEN}$ 线性相关。由图 4 可知，在相同条件下，/t/比/s，ʂ/的 LEN 要低一些。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

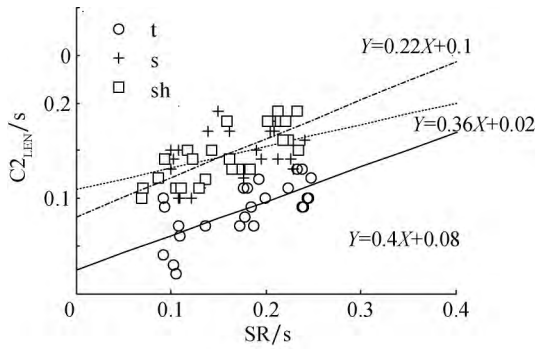


图 4 发音人 XS 的 SR 与 $C2_{LEN}$ 的线性关系
（图中的 Sh 表示|ʂ|）

语速影响 XS 产出的/ʂ/的声学特征。在快速条件下，/ʂ/的谱重心下降，谱偏度增大，这与发音动作幅度减弱有关。

上述结果表明：1）辅音动作时长与语速正相关，语速越快，辅音的动作时长就越短；2）语速只影响擦音的 $C1_{APP}$，在快、慢速条件下擦音的动作过程存在差异；3）在快速条件下，塞音/t/的舌尖动作出现弱化趋势；擦音发音动作受语速的影响小于塞音，擦音/ʂ/动作幅度的降低伴随频谱的变化；在快速条件下，元音对擦音的协同发音幅度增大。

2.3 语速对元音发音动作和声学特征的影响

元音动作和频谱特征受多个因素的影响，除语速外，发音人、音节位置以及元音前后的辅音都会影响元音音段的产出。因此，本文对 2 位发音人的 11 个双音节分别进行回归分析，考察语速对元音的 DIS、CoG、CC、vLEN 以及元音的 F1 和 F2 共振峰的影响。表 3 是元音参数回归方程的确定系数和回归系数。SR 与 vLEN 高度线性相关（R^2 = 0.82～0.99，p<0.000 1），限于篇幅，不放入表中。

表 3 元音参数回归方程的确定系数 R^2 和回归系数

双音节	发音人	元音参数				
		DIS	CoG	CC	F1	F2
/sɿ#pa/前	XS	0.44^{c}(51.9)	0.46^{c}(3.5)	0.47^{c}(1.1)	0.77^{a}(−462)	0.35^{c}(−408)
	JZ	0.79^{a}(76.0)	n. s.	0.86^{a}(2.9)	0.74^{b}(−520)	n. s.
/sɿ#pi/前	XS	0.72^{a}(56.9)	0.71^{a}(3.7)	0.45^{c}(1.1)	0.67^{a}(−531)	0.69^{a}(−1 871)
	JZ	n. s.	n. s.	n. s.	0.60^{c}(−913)	n. s.
/pa#sɿ/后	XS	0.83^{a}(169.2)	0.78^{a}(12.4)	0.83^{a}(3.2)	0.74^{a}(−793)	0.88^{a}(−3 170)
	JZ	n. s.	n. s.	n. s.	0.38^{c}(−631)	0.82^{a}(−4 873)
/pi#sɿ/后	XS	0.78^{a}(109.1)	0.77^{a}(7.1)	0.81^{a}(2.3)	0.83^{a}(−568)	0.91^{a}(−2 458)
	JZ	n. s.	n. s.	n. s.	n. s.	0.91^{a}(−3 683)
/ʂɿ#pa/前	XS	0.31^{c}(27.9)	n. s.	n. s.	0.58^{a}(−311)	n. s.
	JZ	0.44^{c}(66.3)	0.38^{c}(−3.7)	n. s.	0.44^{c}(−569)	0.70^{b}(2 114)
/ʂɿ#pi/前	XS	n. s.	n. s.	n. s.	n. s.	n. s.
	JZ	0.33^{d}(34.1)	0.50^{c}(−3.3)	n. s.	0.56^{c}(−708)	0.72^{b}(1 905)
/pa#ʂɿ/后	XS	0.56^{b}(75.1)	0.43^{c}(3.2)	n. s.	0.35^{c}(−433)	0.53^{b}(−785)
	JZ	0.48^{c}(−47.5)	n. s.	n. s.	n. s.	0.54^{c}(−1 015)
/pi#ʂɿ/后	XS	n. s.	0.26^{d}(3.2)	n. s.	n. s.	n. s.
	JZ	n. s.	0.28^{d}(−4.3)	0.40^{c}(−1.5)	n. s.	n. s.
/pi#pa/前	XS	n. s.	n. s.	0.39^{c}(0.7)	0.74^{a}(−493)	n. s.
	JZ	n. s.	n. s.	0.44^{c}(1.3)	0.62^{b}(−1 282)	0.34^{d}(910)
/pa#pi/后	XS	0.72^{a}(84.9)	0.26^{d}(2.1)	0.79^{a}(1.0)	0.68^{a}(−389)	n. s.
	JZ	n. s.	n. s.	n. s.	n. s.	0.56^{c}(1 453)
/pi#pi/前	XS	0.60^{b}(91.8)	0.54^{b}(2.8)	0.52^{b}(0.8)	0.64^{a}(−796)	n. s.
	JZ	0.41^{d}(−56.4)	0.48^{d}(−2.8)	0.38^{d}(−0.7)	0.48^{c}(−1 021)	0.88^{a}(2 094)
/pi#pi/后	XS	0.52^{c}(85.1)	n. s.	0.76^{a}(1.3)	0.42^{c}(−410)	n. s.
	JZ	n. s.	0.34^{d}(−3.8)	n. s.	0.48^{c}(−794)	0.50^{c}(1 641)

显著性水平：p<0.001^{a}，p<0.01^{b}，p<0.05^{c}，p<0.1^{d}。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

由表 3 可知，XS 的前字元音/ɿ/的 DIS 和 F1 受语速影响，在快速条件下，DIS 下降，F1 上升。JZ 的前字元音/ɿ/的 DIS、CoG、F1 和 F2 受语速的影响，语速越快，DIS 下降，CoG 上升，F1 上升，F2 下降。首字元音/ɿ/的动作在快速条件下趋于弱化，而声学上表现为央化的趋势。XS 的后字元音/ɿ/在快速条件下发音动作减弱，DIS 和 CoG 下降，F1 和 F2 上升。JZ 的后字元音/ɿ/随语速的变化方式与 XS 不同，语速加快，DIS、CC 和 CoG 上升，F1 不变，F2 上升。

XS 的元音/i/的生理参数受语速的影响，语速越快，DIS、CoG 和 CC 下降，说明在快速条件下元音动作幅度和舌腭收紧程度均降低。/i/的频谱特征也发生了变化，语速越快，F1 上升。JZ 前字元音/i/的 CC 和频谱特征受到语速的影响，在快速条件下，CC 下降，F1 上升，F2 下降。上述结果表明：快速条件下元音收紧程度的降低使得元音频谱特征发生相应变化。JZ 的后字元音/i/的结果与 XS 的不同，DIS、CoG 和 CC 随语速加快上升，F1 上升，F2 下降，这有可能与快速条件下后续音节声母/tʂ/的影响有关。

语速对 XS 的舌尖前元音/ɿ/的发音动作影响显著，语速加快，DIS、CoG 和 CC 降低，F1 和 F2 上升。表 4 为 3 种语速条件下 XS 产出的元音/ɿ/的腭位，样本取自双音节/sɿ ♯ pa/中的前字元音，图中黑点表示电子假腭的电极，灰色区域表示舌腭接触的区域。可以看出，随着语速的加快，前腭区域的收紧程度明显降低，收紧点截面积增大，舌尖动作出现了明显的弱化；F1 和 F2 的升高表明元音有央化的趋势(F1 变化范围为 359～532 Hz，F2 变化范围为 925～1 420 Hz)。语速对 JZ 的前字元音/ɿ/的 DIS 和 CC 有影响，表现为语速加快，DIS 下降，前腭区域的收紧程度降低；F2 不受语速影响，但是 F1 随语速的加快而上升。JZ 的后字元音/ɿ/的生理参数不受语速影响，但是 F1 和 F2 却随语速的加快而上升。上述结果表明：在快速条件下，元音/ɿ/的舌尖动作减弱，在声学上表现出元音央化的趋势。需要注意的是，舌尖动作弱化还有可能与音段间协同发音的增强有关，前字元音的腭位有可能受到跨辅音元音/a/的影响，后字元音的腭位有可能受到后续音节声母/tʂ/的影响。

表 4 3 种语速条件下 XS 产出的元音/ɿ/的腭位

项目	快速	正常	慢速
腭位			

元音分析结果表明：1) SR 与 vLEN 具有高度线性关系；2) 在快速条件下，元音的发音动作易于弱化，元音频谱特征表现出央化的趋势；3) 在快速条件下，元音发音动作易受周边音段的协同发音影响。

3 讨论和结论

本文使用 EPG 分析了语速对普通话音段产生的影响。普通话语速调整与发音动作指令间距密切相关，快速条件下，相邻音段发音动作指令间距缩短。不同发音器官的动作在时域上重叠，因而协同发音程度增大。然而，发音动作的时间关系受辅音发音方式的制约。研究发现，擦音影响发音动作时间关系，使得动作间绝对时差变大，动作重叠度变小，这是因为鼻音和擦音对软腭的动作要求相互冲突，后者需要关闭鼻咽通道以保持口腔内保持较高的气压条件。

语速对辅音和元音音段的影响主要表现为生理(声学)时长的缩短和动作弱化。在快速条件下，辅音动作时长缩短，发音动作幅度一般降低，但是降低程度取决于辅音的发音方式。塞音动作易于弱化，擦音的动作受语速影响较小，后者受擦音动作控制、空气动力条件和音段感知条件的制约。擦音的成阻动作时长随语速的变化而变化，且在快、慢速条件下产出动作存在差异。元音时长随语速的加快而线性缩短，并伴随着动作幅度的减弱。在快速条件下，元音动作易于弱化，在声学上表现为元音央化的趋势；同时，元音受语音环境影响的程度变大。

分析结果表明：普通话语速控制与音段发音动作的时间关系以及音段动作的时空特征有关，音段发音属性和发音人发音策略也影响不同语速条件下音段产出的结果。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

参考文献 (References)

[1] Tsao Y, Weismer G, Iqbal K. Interspeaker variation in habitual speaking rate: Additional evidence [J]. *Journal of Speech, Language, and Hearing Research*, 2006, **49**(5): 1156-1164.

[2] **张家騄，朱维彬**，Eva Gårding. **汉语句子的韵律结构** [J]. **声学学报**，2002，**27**(5)：389-392.
ZHANG Jialu, ZHU Weibin, Gårding E. Temporal structures of spoken Chinese sentence [J]. *Acta Acustica*, 2002, **27**(5): 389-392. (in Chinese)

[3] Kessinger R H, Blumstein S E. Effects of speaking rate on voice-onset time in Thai, French, and English [J]. *Journal of Phonetics*, 1997, **25**(2): 143-168.

[4] Agwuele A, Sussman H M, Lindblom B. The effect of speaking rate on consonant vowel coarticulation [J]. *Phonetica*, 2008, **65**(4): 194-209.

[5] Fourakis M. Tempo, stress, and vowel reduction in American English [J]. *Journal of the Acoustical Society of America*, 1991, **90**(4): 1816-1827.

[6] Gay T. Effect of speaking rate on vowel formant movement [J]. *Journal of the Acoustical Society of America*, 1978, **63**(1): 223-230.

[7] Berry J, Weismer G. Speaking rate effects on locus equation slope [J]. *Journal of Phonetics*, 2013, **41**(6): 468-478.

[8] Adams S G, Weismer G, Kent R R. Speaking rate and speech movement velocity profiles [J]. *Journal of Speech, Language, and Hearing Research*, 1993, **36**(1): 41-54.

[9] Byrd D, Tan C. Saying consonant clusters quickly [J]. *Journal of Phonetics*, 1996, **24**(2): 263-282.

[10] **曹剑芬. 语速特征及其变化**[C]// **第六届全国现代语音学学术会议. 天津，中国**，2003：143-148.
CAO Jianfen. Speaking rate and its variations [C]// Proc National Conference on Modern Phonetics. Tianjin, China, 2003: 143-148. (in Chinese)

[11] LI Yinghao, KONG Jiangping. Effects of speech rate on inter-segmental coarticulation in Standard Chinese [C]// Proc ISCSLP. Piscataway, NJ, USA: IEEE Press, 2010: 44-49.

[12] Liker M, Gibbon F E. Difference in EPG contact dynamics between voiced and voiceless lingual fricatives [J]. *Journal of International Phonetic Association*, 2013, **43**(1): 49-64.

[13] Hardcastle W J, Gibbon F, Nicolaidis K. EPG data reduction methods and their implications for studies of lingual coarticulation [J]. *Journal of Phonetics*, 1991, **19**: 251-266.

[14] Fontdevila J, Pallares M D, Recasens D. The contact index method of electropalatographic data reduction [J]. *Journal of Phonetics*, 1994, **22**(2): 141-154.

[15] Forrest K, Weismer G, Milenkovic P, et al. Statistical analysis of word-initial voiceless obstruents: Preliminary data [J]. *Journal of the Acoustical Society of America*, 1988, **84**(1): 115-123.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

藏语声调起源的多维视角

孔江平*

北京大学中文系，中国语言学研究中心

语言与人类复杂系统联合研究中心

提要

本文从不同的视角对藏语声调产生和起源进行了研究，研究主要得到了以下结论：1）同音词的增加是导致藏语产生声调的主要内在动力。从结构的演化来看，声调的产生又受到藏语音位系统结构的制约；2）同音词的分布在藏语各方言中都是遵循指数曲线，这是藏语音位的基本内在性质；3）从音位负担的演化来看，在藏语信息量为常数的情况下，声、韵、调三者的音位负担可以相互转换，藏语声母负担的增加直接导致了声调的产生，因此，藏语的音调转换为声调；4）提出了“准音位”的感念；5）从语音结构在词汇中的扩散和语音演化看，藏语声、韵母的减少形成了藏语语音结构在词汇中的扩散；6）声调产生过程中，语音的演化主要受制于发音生理机制和感知机制,声母清化是声调产生的感知基础。

主题词

声调起源　**同**音词率　**音**位负担量　**词**汇扩散　**言**语生理感知

1. 引言

藏语主要分布在中国的西藏自治区、青海省、甘肃省、云南省和四川省西部等高原地区。藏语是研究汉藏语系声调起源的一种重要语言，因为：1）藏语使用者主要居住在海拔相对较高的地区，由于交通工具不发达，居住地理环境相对隔绝，形成了藏语的单一性；2）藏族是藏语的主要使用者，大部分藏区民族成份相对单一，这种相对单纯的人文环境使得藏语在演化过程中和其他语言的接触较少；3）藏语从

*孔江平，北京大学中文系，中国语言学研究中心，语言与人类复杂系统联合研究中心。地址：中国北京市海淀区颐和园路 5 号，北京大学中文系。邮编：100871。jpkong@pku.edu.cn。

7 世纪开始，逐渐形成了自己的传统文字，在历史的长河中，积累了大量的藏文文献，为研究提供了丰富的书面材料；4）从共时的平面看，藏语形成了许多无声调和有声调的方言，它们反映了声调演化的过程，也反映了藏语形成前底层语言的差异，是研究声调起源珍贵的活的语料，也是研究汉藏语系语言声调起源的活化石。

藏文是拼音文字，创立于 7 世纪，之后有过三次厘定，现在标准的藏文基本代表了 9 世纪的语音。从藏语音韵学的角度看，藏语在九世纪前后有 220 个声母和 100 个韵母，没有声调，这是藏语古音韵的基本结构（格桑居冕等，1991，2002,2004；孔江平，2010；孔江平于洪志等，2011）。在 220 个声母中，有 30 个是单辅音声母，其它为复辅音声母。在 100 个韵母中，有 5 个单元音韵母和少量复元音韵母，大部分是带韵尾的韵母。另外，藏语有大约 5190 个单音节语素（孔江平，2010，2011）。从古藏文音韵系统和现代方言看，整个藏语音韵的变化趋势是复辅音声母和韵母都逐渐减少，元音和韵母变得越来越复杂，声韵组合结构的数量越来越少。根据藏文和方言的音韵结构，我们可以看出藏语声调的产生主要和声母的清浊和韵尾的性质有密切关系（马学良主编，2003；黄布凡，1994）。

藏语有较大的方言差别，国内一般将藏语分为“安多”、“康巴”、“卫藏”三个方言（瞿霭堂，1963，1987；Shafer R.，1955；）。从藏文和藏语方言的音位系统上看，藏语的声母（复辅音和单辅音）按藏文、安多、康巴和卫藏的顺序递减，藏语在藏文和安多方言中还没有声调，但有些已经形成了单音节和双音节的音调模式。卫藏已经产生了声调，通常一个方言有 4 至 6 个单音节声调。康方言通常认为有声调，但根据我们的实地调查，这一说法很不确定，有的康方言有声调，如德格藏语，有的康方言并没有形成完整的声调系统，如玉树藏语。根据目前的调查资料，将安多藏语定为无声调（non-tonal）及将卫藏藏语定为有声调（tonal）是比较符合实际的。至于康方言，定义为正在产生声调（tonal developing）比较符合实际。

以往对藏语声调起源和产生的研究主要是在历史语言学的框架下（Light T.，1978；Sun，1997,2003）。在方法论上也主要是采取语音对应规律的原则，这方面的研究取得了一定的成果。如藏语声调的形成和声母的清浊及韵母的类型有关，即清声母形成高调，浊声母形成低调，塞韵尾能使声调变高，喉塞尾能使调尾降低等。但藏语语音的演化并不是那么简单（Sun，2003），为了进一步解释藏语声调产生的内在动因，本文根据藏语方言的实际情况：1）从藏语语音演变过程的角度，采用分析藏语语音结构演化过程的方法、2）从藏语音位负担量演化的角度，采用分析藏语信息量的方法、3）从藏语语音结构在词汇上扩散和语音生理演化的角度，采用声学分析和语音结构在词汇上扩散的方法，研究和探讨了藏语声调起源和产生的内在动

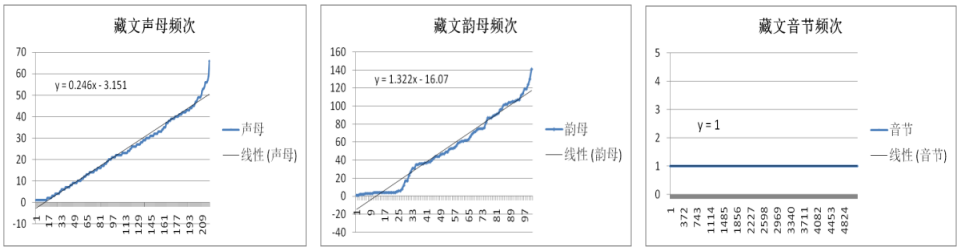
力以及生理机制，科学地分析和解释了藏语声调起源和产生的内在本质。同时本文还讨论了研究藏语声调起源的心理感知范畴和脑科学研究的可能性与前景。

2. 藏语声调起源的音位结构演化

在以往对藏语声调产生和起源的研究中，主要是采用历史语言学语音对应规律的研究方法。然而，从历史语言学的角度，还可以采用其他新的方法来对藏语声调起源进行更深入地研究。传统历史语言学的方法通常只注重语音对应规律的结果，对藏语声调产生的原因、动力和过程解释甚少。这一节是从历史语言学的角度，采用研究藏语各方言音位结构演化的方法，探讨藏语音位结构演化和声调产生及起源的关系。本项研究的数据仅限于藏语 5190 个单音节语素（孔江平，2011），除了藏文外，方言包括泽库、化隆、德格、循化、夏河、巴塘、玉树、日喀则、拉萨 9 个点。每个数据库中包含藏文、拉丁转写、汉义、声母、韵母、音节等项。

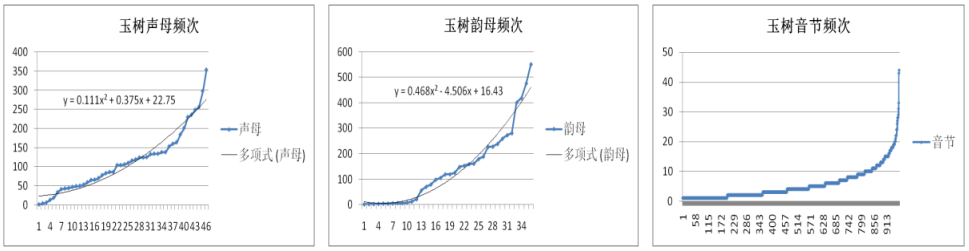
2.1 声母、韵母和音节的频次分布

历史语言学的初步研究表明，藏语声调的产生和声母的清浊及韵尾有关，同时还表明藏语在声调形成的过程中，声母主要表现为复辅音的消失。因此本文从藏语各方言声母、韵母和音节的分布来研究藏语音位结构的演化。研究对藏文和方言的声母、韵母及音节的出现频次进行了统计，并将结果以图形表示出来，见图一。



图一、藏文声母（左）、韵母（中）和音节（右）在 5190 个语素中出现的频次曲线

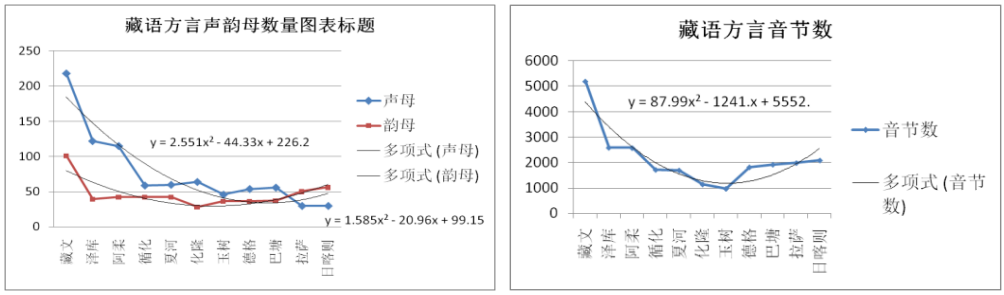
从图一藏文声母（左）的曲线可以看出，藏文声母的分布基本上可以用线性方程来描写。这说明有的声母出现的频次很小，有的声母出现的频次很大，它们之间是一个线性关系。同样，从图一藏文韵母（中）的曲线可以看出，藏文韵母的分布也可以用线性方程来描写，虽然数量不同，但分布的性质相同。从图一藏文音节（右）曲线可以看出，藏文的音节分布是一条直线，这是因为在 5000 多个藏语单音节语素中是没有考虑同音词的出现频次，这就形成了藏语语音结构共时断面的一个基准。



图二、玉树声母（左）、韵母（中）和音节（右）在 5190 个语素中出现的频次曲线

从图二玉树声母（左）曲线可以看出，玉树藏语声母的分布基本上可以用二次方程来描写，即有些声母出现频次越来越高。从图二玉树韵母（中）曲线可以看出，玉树藏语韵母的分布也应该用二次方程来描写，即有些韵母出现频次越来越高，和声母有同样的分布特性。从图二玉树音节（右）的分布曲线可以看出，玉树藏语音节可以用大致用指数函数（a>1）来描写，即在曲线的后部数值突然增大，这说明有些同音词迅速大量增加。

分析藏文和玉树藏语声母频次分布、韵母频次分布和音节频次分布曲线不难发现，两者有很大的差别，藏文声母、韵母频次为线性分布，而玉树藏语声母、韵母频次为二次方程分布，由于这种差别，导致两者在组合产生音节后，形成了玉树藏语音节的接近指数的分布。分析其它藏语方言点的数据不难发现，它们和玉树藏语在声母、韵母和音节的频次分布上都具有相同的性质，它们和藏文声母、韵母和音节的分布有很大差别。这表明藏语方言音节数量在迅速减少，同时产生了大量同音词，给语言的交际造成了困难。



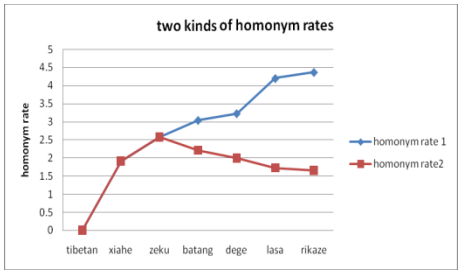
图三、藏语方言声、韵母（左）和音节（右）的数目曲线

从图三（左）中可以看出，藏语声母从藏文开始呈下降分布，但韵母的变化不如声母那样大，这说明声母有剧烈的变化，而韵母相对在一个较小的范围内变化。

从图三（右）中可以看出，玉树左边是无声调方言，右边是有声调方言，玉树藏语的音节数最小，也就是说随着声母的减少音节也减少，同音词增加，而在产生了声调后，音节又逐渐增加，同音词减少。

2.2 同音词对声调产生的贡献

本文作者研究了藏语同音词和藏语声调产生的关系（孔，2012），在此做一些简单的介绍，并以此来讨论下一节的内容。为了讨论的方便我们将“同音词”定义为两种，一种是声韵母相同的音节，但在有声调的方言中并不是真正的同音词，在无声调的方言中是真正的同音词。另一种是声韵调相同的音节。我们把按第一种定义计算出的结果称为同音词率一，把按第二种定义计算的结果称为同音词率二。



图四、同音词率一和同音词率二的曲线

从图四的曲线可以清楚地看出，没有声调的前三个藏语方言中，同音词率一和同音词率二是相同的，第四至第七方言是有声调的，同音词率一和同音词率二开始有不同的走向。第三个方言点的同音词率反映了藏语声调产生的起点。同音词率二的左边为无声调的方言，而右边为有声调的方言，这个范围在 2 到 3 之间。也就是说在这个同音词率范围内，藏语的语音系统会产生声调，这可以来对藏语方言进行是否有声调的预测。这个结论说明，同音词的产生和声调的产生有极大的相关性，这种相关性可以解释为藏语声调产生的一种内在动力。

2.3 同音词分布与语音结构

从语言交际的角度看，音位起着承载信息的作用，但当音位负担加重时就会寻求系统的改变。根据本文的研究，举例说明藏语方言在产生声调时音位结构起到的作用。见表一。

表一、藏语拉萨话词例

藏文　　汉义　　音标　　藏文　　汉义　音标　　藏文　　汉义　音标

ཚགས་པ་ཡིན་ 宣布　tʂak131　　ཐག་ཡིན་　过于　tʂak131　　ཐིག་　　干肉　tʂak131

ཚགས་པ་　宣布　tʂak131　　ཐག་　　污垢　tʂak131　　ཟིག་　　害怕　tʂak51

ཐག་ཡིན་　傲慢　tʂak131　　བཅག་　　痊愈　tʂak131　　ཐིག་　　煎　　tʂak51

ཐག་　　噎住　tʂak131　　བཅག་　　割　　tʂak131　　བཅིག་　　空隙　tʂak51

ཐག་　　痊愈　tʂak131　　བཅག་ཡིན་ 烧焦　tʂak131　　ཐེག་　　煎　　tʂak51

ཐེག་　　岩石　tʂak131　　བཅེག་　　油罐　tʂak131　　ཚིག་ཡིན་ 姓氏　tʂak51

ཕྱག་　　已割　tʂak131　　བཅྱག་　　剪　　tʂak131　　ཚིག་　　光泽　tʂak51

ཕྱག་ཡིན་　剃　　tʂak131　　ཕྱོག་　　邮递　tʂak131　　བཅགས་ཡིན་　大叫　tʂak131

ཕྱོག་ཡིན་　已随同　tʂak131

表一列出了 25 个词，这是拉萨藏语中声韵母组合频次最高的音节，分析藏文的声母可以发现，藏文的基字不同，不仅有清音 p,k，也有浊音 g,d,b，但每个声母都有下加字 r，在拉萨藏语中为/tʂ/。元音有 a 和 e，在拉萨藏语中为/a/。藏文的韵尾 g 和 gs 在拉萨藏语中为/k/。同时，分析表一不难发现，声调有两个，一个是/131/，另一个是/51/。从藏文和拉萨藏语语音的对应规律来看，声调的产生和声母的清浊及韵尾有关。我们在此要说明的是，虽然在复辅音声母减少的过程中，拉萨藏语中形成了数目较多的同音词，但音节结构对声调产生的制约性还是很强，并没有产生较多的声调。

表二、藏语拉萨话词例

藏文	汉义	音标	藏文	汉义	音标
ཁྲིམས་	权力	ŋa53	བཀྲམས་	排列	ŋa55
རྔ་	鼓	ŋa53	ཁྲིམས་	甜的	ŋa55
ཨ་	辅音字母	ŋa53	ཧྲིལ་	往昔	ŋa55
ཧྲིལ་	早	ŋa53	སྐྱེལ་	力量	ŋa15
ཁྲིལ་	收割	ŋa53	སྐྱེ་	我	ŋa13

表二是另一个例子，从藏文的角度看，共有 10 个单音节词。在拉萨藏语中，这

10 个词的声韵母都相同，但有 4 个不同的声调，这样就形成了 4 个音节。虽然藏文中仅有 10 音节，数量比表一的音节少，但却产生了 4 个声调。仔细分析我们可以发现，并不是同音词多就会产生较多的声调。声调的产生不仅需要有较多的同音词，即同音词率平均在 2.5 左右，同时又要遵循音位系统的整个结构系统的制约。研究表明，同音词对声调产生有贡献，贡献在于有需要区别词义，但具体声调的产生在结构上取决于声母和韵母的结构制约，我们将这种制约称为藏语声调产生过程中音位系统的“结构制约”。

3. 藏语声调起源的音位负担量演化

上一节讨论了藏语语音结构的演化和由结构演化导致的同音词演化，这显然是声调起源和产生的两个方面。某些音节大量同音词的出现，引起了声调的产生，是一种内在的动力，然而声母和韵母结构的演化是声调产生的形式。在这一节我们将进一步从音位负担量的角度，讨论语言信息量在藏语声调产生中的作用。

3.1 音位负担定义

在音位负担的定义上，我们将一个单音节语素定义为一个语言的信息单位，一个信息单位由几个不同的音位来负担。如果由一个音位来区别这两个语素，这个信息单位就负载在这个音位上，如果是由多个音位来区别，词信息量就负载在多个音位上。根据藏语的情况，本文采用“声母”、“韵母”和“声调”为音位单位。

“音位负担”定义为：一种语言音位系统内部音位结构、音位分布和音位功能的负担，称为“音位结构功能负担（phoneme structural functional load)”，简称“音位负担”(phoneme load)。语言音位系统内部音位结构、音位分布和音位功能负担研究称为“音位结构功能负担研究”，简称“音位负担研究”。“音位负担研究”主要是以一种语言基本语素的内部音位结构、音位分布和音位功能负担为研究对象。而“语言功能负担（functional load)” （王士元，1967）是以一种语言的实际文本为基础的音位功能负担研究，两者有很大不同。

在汉藏语系语言中，基本语素大多是单音节，每个音节均由声母和韵母或声母、韵母和声调组成。以藏语有声调的方言为例，音节和音节之间的对立类型总共包括：三项对立、两项对立、单项对立和无对立四种，其中三项对立有一种类型，二项对立有三种类型，单项对立有三种类型，无对立有一种类型，共八种类型，形成了藏语音位系统结构和分布的基本形式和框架，见表三。

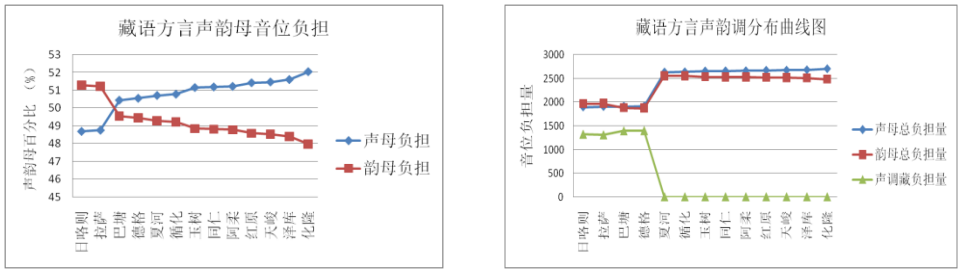
表三、藏语音位对立类型表

对立类型	对立方式		
三项对立	声/韵/调对立		
两项对立	声/韵对立	声/调对立	韵/调对立
单项对立	声母对立	韵母对立	声调对立
无对立	声/韵/调相同		

三项对立是指两个单音节语素之间声韵调都不同；两项对立是指两个单音节语素之间只有两个音位单位不同，即声韵不同或声调不同或韵调不同；单项对立是指两个单音节语素之间只有一个音位单位不同，即声母不同或韵母不同或声调不同；无对立是指两个单音节语素之间没有音位单位不同，即同音词[1]。对于藏语无声调的方言来说，对立类型是不同的，由于没有声调，其对立类型要少。首先是无三项对立，因为没有声调，只有声韵和韵母。两项对立只有一种，及声韵母对立。单项对立有两种，即声母对立和韵母对立。最后是一个无对立类型。这样一共只有 3 中对立类型，但在计算上，为了计算程序的通用性我们按八种类型来设计。这样在无声调藏语方言的计算上，声调的参数统一确定为 0，因此最终计算的结果并不影响结论。为了比较有声调方言的数据和无声调方言的数据，我们将有声调的方言按两种方式计算，一种是按声、韵来计算，另一种是按声、韵、调来计算。

3.2 藏语方言音位负担量的性质

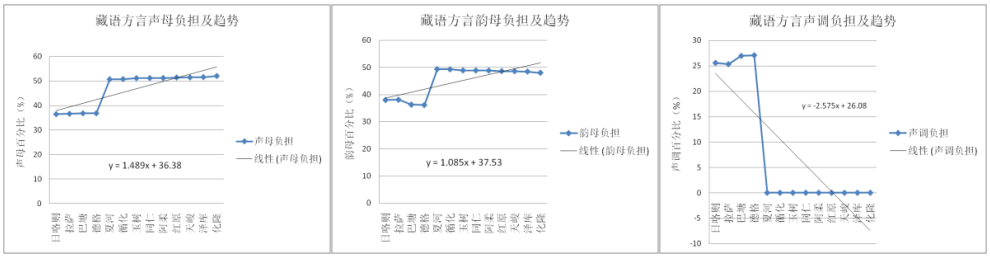
由于藏语方言有的有声调，有的无声调，在计算它们的音位负担时采用了两种计算方法，这样可以看出藏语方言声韵的演化总过程和趋势，见图五。



图五、藏语方言音位负担百分比曲线

从图五（左）可以看出，声母的音位负担是按从小到大来排序的，其中日喀则和拉萨的声母负担最小，排在左边。紧接着的巴塘和德格藏语，声母的音位负担较

小。而化隆和泽库的声母负担最大，排在右边。但韵母的排列正好和声母相反，形成镜像分布。目前对藏语的研究认为，日喀则和拉萨有声调属卫藏方言，巴塘和德格也有声调，属康方言。其他的无声调，属安多方言。分析藏语方言声韵母的音位负担可以发现：1）声母和韵母的音位负担在藏语方言中并不相同，由于承载的信息量是一个常数，声母负担大，韵母负担就小；2）日喀则和拉萨的声韵母负担和其他方言的负担之间有一个跳跃，这说明日喀则和拉萨在音位结构负担上比较接近。虽然通常认为巴塘和德格是有声调的方言，但它们在音位结构负担上更接近安多方言。3）日喀则和拉萨的韵母负担要大于声母的音位负担，而其它方言都是声母音位负担大于韵母音位负担。由此可以看出，声母音位负担小于韵母音位负担是声调系统稳定的一个重要标志。图六（右）是按声、韵、调三项计算的结果，其中日喀则、拉萨、巴塘和德格声调有实际的调值，其它声调按“零”来计算。同样也是按照声母从小到大排序，分析图中的曲线可以发现：1）有声调的方言其声母和韵母的负担大大降低，分散给了声调；2）韵母的负担和声调的负担成反比关系，与声调成正比关系。从古藏语的音位系统看，藏语的声母是逐渐减少的，因而，整个音位系统的负担会越来越重。在实际的藏语方言交际中，声调的产生并不是突变的，无声调的藏语方言首先通过音调模式来减轻负担。在现代音位学的理论中没有“半音位”或“准音位”这样的单位和概念，但实际上在对藏语方言的调查过程中发现，许多无声调方言中有些词之间已经形成了声调的对立，但由于数量较少，不能建立声调的整体系统，这为藏语声调产生过程的精确描写和研究带来了困难。在此我们提出“准音位”的概念，以便引起学界在音位演化研究中对这种现象的重视。



图六、藏语方言声母（左）、韵母（中）和声调（右）负担及趋势拟合曲线

图六是藏语方言声母、韵母和声调负担及趋势拟合曲线，这三张图中藏语方言都是按声母从小到大排序的。如果将最右边的方言假设为完全没有声调功能的语言，从图六声母（左）可以看出，有声调的前 4 个方言声母的音位负担大约在 36%左右，无声调方言声母的音位负担量大约在 50%以上，差别很大。从图六韵母（中）的曲线

看也基本上如此。图六（右）显示了声调的数据和拟合趋势曲线，无声调的藏语方言数据为零。众所周知，康方言和部分声母数量较少的安多方言已经有了所谓的“习惯调”。从语音学的角度，本文将这种现象称为音调模式（pitch pattern）以区别于声调模式(tone pattern)。由于音位学无法定义和描写这种音位系统的渐变过程，本文用拟合曲线来简单表示其变化的趋势，因此这三条曲线体现了藏语从无声调到有声调的变化过程，为我们认识藏语语音演化和声调产生的过程以及藏语方言的田野调查提供了依据。从图六（右）的拟合曲线可以看出，“准音位声调”在无声调的方言中的功能是不同的，夏河、循化等较强，泽库、化隆等较弱。

4. 语音结构在词汇中的扩散和语音演化

根据我们调查的资料，藏语声调在产生过程中，无论是词汇还是语音都存在着逐渐演化的过程，这一过程很复杂，涉及词汇和语音两个方面。下面将从语音结构在词汇中的扩散和音调及声调感知两个角度简要介绍一下演化过程的两个方面。

4.1 声调产生中的词汇扩散

从藏语声母在各方言的情况看，由于复辅音声母的减少，会出现大量的同音词，同音词的增加会给语言的交际带来困难。根据第一节的讨论可知，虽然声调的产生和同音词相关，但声调产生却必须遵循整个音位结构安排，不是同音词出现多就会分化出更多的声调。在这一节主要讨论声调产生的过程。众所周知，语音的演化往往会遵循词汇扩散的原则，下面本文从词汇扩散的角度来分析一下藏语方言复辅音声母减少和声调产生的过程，看是否有词汇扩散的现象。

本文以拉萨话声母和韵母组合最多的/tʂak/为例加以说明。由于有藏语方言有的有声调，有的无声调，音节/tʂak/在有声调的方言中不一定都是同音词，但在无声调的方言中，/tʂak/都是同音词。根据所有举例方言出现该音节的情况，我们将藏文最大公倍数作为该音节的总数，这样藏文中一共有38个单音节词。其中拉萨话、日喀则和德格分别有25个词为该音节，巴塘有15个词为该音节，循化和夏河分别有12词为该音节，化隆有8个词为该音节，泽库有7个词为该音节。根据具体的情况，本文将其分为5个部分加以说明，见表四至八。

表四、藏语例词表

藏文	汉义	转写	拉萨	喀则	德格	巴塘	循化	夏河	化隆	泽库
賱繳	煎	krag	tʂak51	tʂak51	tʂaˀ51	tʂaˀ51	tʂəx	tʂəx	tʂak	tʂak

藏文	汉义	转写	拉萨	日喀则	德格	巴塘	循化	夏河	化隆	泽库
光泽	bkrag	tʂak51	tʂak51	tʂaʔ51	tʂaʔ51	tʂəx	tʂəx	tʂak		
姓氏	bkrag	tʂak51	tʂak51	tʂaʔ51	tʂaʔ51	tʂəx	tʂəx	tʂak		
煎	sprag	tʂak51	tʂak51	tʂaʔ51	tʂaʔ51					
害怕	skrag	tʂak51	tʂak51	tʂaʔ51	tʂaʔ51					
空隙	dprag	tʂak51	tʂak51	tʂaʔ51						

表四共有 6 个词，基字为 k 或 p，下加字都是 r，元音都是 a，韵尾是 g 或 gs。从拉萨、日喀则、德格和巴塘 4 个有声调的方言看，都是高降调。其中拉萨、日喀则和巴塘 6 个词都出现该音节。德格只有 5 个词出现该音节。循化、夏河和化隆有 3 个词出现该音节。泽库只有一个词出现该音节。因此不难看出，各方言词汇在演化过程中的情况是不同的，有的快有的慢，并不是一致的。

表五、藏语例词表

藏文	汉义	转写	德格	藏文	汉义	转写	德格
木澡盆	krab	tʂaʔ51		筛净	dkrab	tʂaʔ51	
选择	bkrab	tʂaʔ51		跳舞(古)	skrab	tʂaʔ51	
选择	bkrabs	tʂaʔ51		空隙	dprag	tʂaʔ51	

表五有 6 个词，基字都是 k，下加字都是 r，元音都是 a，韵尾是 g、b 或 bs。表五中只有德格一个方言有出现该音节，都是高降调。不难看出，德格在韵尾上出现了 b 和 bs 韵尾，这是其它方言没有的，比较特别。

表六、藏语例词表

藏文	汉义	转写	拉萨	日喀则	德格	巴塘	循化	夏河	化隆	泽库
痊愈	dag	tʂak121	tʂak121	tʂaʔ121	tʂaʔ121	tʂəx	tʂəx	tʂak	tʂak	
过于	dags	tʂak121	tʂak121	tʂaʔ121	tʂaʔ121	tʂəx	tʂəx	tʂak	tʂak	
已割	breg	tʂak121	tʂak121	tʂaʔ121	tʂaʔ121	tʂəx	tʂəx	tʂak		
剃	bregs	tʂak121	tʂak121	tʂaʔ121	tʂaʔ121	tʂəx	tʂəx			
岩石	brag	tʂak121	tʂak121	tʂaʔ121	tʂaʔ121	tʂəx	tʂəx	tʂak		

軿繳	污垢	dreg	tʂ̥ak121	tʂ̥ak121	tʂ̥aˀ121	tʂ̥aˀ121	tʂ̥əx	tʂ̥əx	tʂ̥ak	tʂ̥ak
軿繳霻	傲慢	dregs	tʂ̥ak121	tʂ̥ak121	tʂ̥aˀ121	tʂ̥aˀ121	tʂ̥əx	tʂ̥əx	tʂ̥ak	
羒繳	据说	grag	tʂ̥aˀ121	tʂ̥aˀ121			tʂ̥əx	tʂ̥əx	tʂ̥ak	
羒繳霻	传遍	grags	tʂ̥aˀ121	tʂ̥aˀ121			tʂ̥əx	tʂ̥əx	tʂ̥ak	

表六共有 9 个词，基字为 d、b 或 g，下加字都是 r，元音都是 a 或者 e，韵尾是 g 或 gs。从拉萨、日喀则、德格和巴塘 4 个有声调的方言看，都是低升降调。其中拉萨、日喀则和巴塘 7 个词都出现该音节。德格、巴塘、循化和夏河 9 个词都出现该音节。化隆有 5 个词出现该音节，泽库有 6 个词出现该音节。因此也不难看出，各方言词汇在演化过程中的情况是不同的，有的快有的慢，并不是一致的。

表七、藏语例词表

藏文	汉义	转写	拉萨	日喀则	藏文	汉义	转写	拉萨	日喀则
繼毵繳	宣布	bsgrag	tʂ̥ak121	tʂ̥ak121	繍鱠繳	剪	vbreg	tʂ̥ak121	tʂ̥ak121
繼毵繳霻	宣布	bsgrags	tʂ̥ak121	tʂ̥ak121	繍輫繳	油罐	vbrag	tʂ̥ak121	tʂ̥ak121
軨繳	噎住	sgreg	tʂ̥ak121	tʂ̥ak121	繍軿繳	割	vdreg	tʂ̥ak121	tʂ̥ak121
腵繳	干肉	sbreg	tʂ̥ak121	tʂ̥ak121	繍軿繳霻	烧焦	vdregs	tʂ̥ak121	tʂ̥ak121
繍羒繳霻	大叫	vgrags	tʂ̥ak121	tʂ̥ak121	腫繳	邮递	sbrag	tʂ̥ak121	tʂ̥ak121
繍輅繳	痊愈	vdag	tʂ̥ak121	tʂ̥ak121	腫繳霻	已随同	sbrags	tʂ̥ak121	tʂ̥ak121

表七有 12 词，基字为 b 或 g，下加字都是 r，元音都是 a 或者 e，韵尾是 g 或 gs。表七只有拉萨和日喀则两个有声调的方言有该音节出现，都是低升降调，其它各个方言在这 12 个词中并没有该音节出现。

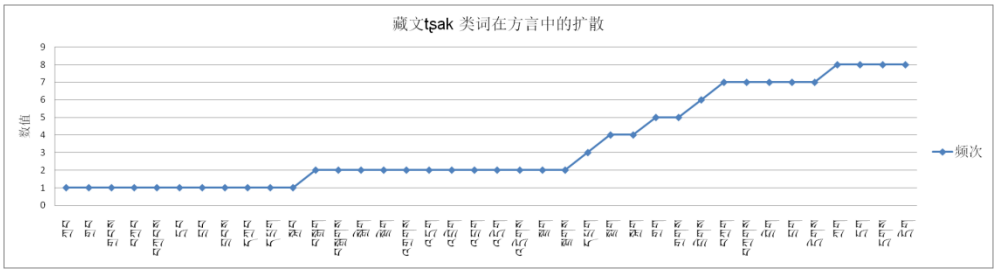
表八、藏语例词

藏文	汉义	转写	德格	藏文	汉义	转写	德格
羒繿	长木条	grab	tʂ̥aˀ121	輫繿	鞭打	brab	tʂ̥aˀ121
羒繿霻	临近	grabs	tʂ̥aˀ121	輫繿霻	已鞭打	brabs	tʂ̥aˀ121
輅繿	椽子	drab	tʂ̥aˀ121				

表八有 5 个词，基字分别是 g、d 和 b，下加字都是 r，元音都是 a，韵尾是 b 或 bs。表八中只有德格一个方言有出现该音节，都是低升降调。不难看出，德格在韵尾上出现了 b 和 bs 韵尾，这是其它方言没有的，比较特别。

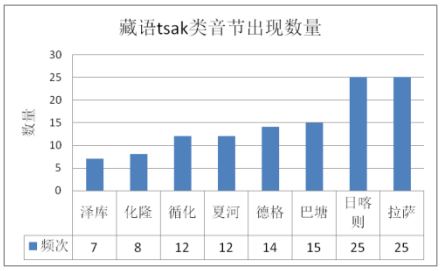
从以上 5 个表中的例子可以看出，藏文中的 38 个单音节词，复辅音声母脱落的过程和速度是不同的，各个方言之间也有差别。但从藏文看，其主要的规律是下加字 r 形成了/tʂ/，基字可以是 g、d、b、k 和 p。韵尾可以是 g、gs、b 和 bs。清声产生高降调，浊声产生低升降调。

图七、藏语声调产生词汇扩散分布



图七是 38 个词在八个方言中扩散的情况，从图中可以看出，横轴是 38 个藏文词汇，纵轴是方言数量，从左边起，词 1 至 11 共有 11 个词出现在一个方言中，词 12 至 23 共有 12 个词出现在两个方言中，词 24 只有 1 个词出现在三个方言中，词 25 至 26 共有 2 个词出现在四个方言中，词 27 至 28 共 2 有个词出现在五个方言中，词 29 只有 1 个词出现在六个方言中，词 30 至 34 共有 5 个词出现在七个方言中，词 34 至 38 共有 4 个词出现在 8 个方言中。

图八、藏语声调产生词汇在方言中的扩散



图八是 38 个词在个方言中出现的频次，横轴是方言，纵轴是词汇数量。从图中

可以看出这 38 个词在各方言中语音出现的频次。从小到大，泽库出现了 7 个词，化隆出现了 8 个词，循化和夏河各出现了 12 个词，德格出现了 14 个词，巴塘出现了 15 个词，日喀则和拉萨出现了 25 个词，高举榜首。

很显然，无论是从词汇的角度，还是方言的角度，我们都可以看出，在复辅音声母减少、同音词增加和声调产生的过程中，存在词汇扩散的现象，这种扩散是结构性的。实际上，在对藏语方言的调查中，我们也发现藏语方言，特别是康方言，其音位系统很难整理，不仅是辅音和元音音位难以确定，单音节声调也很难确定，而双音节声调就更不稳定了。因此可以确定，藏语在声调产生的过程中，是以语音结构在词汇中扩散的方式演化的。而语音结构在词汇中的扩散还需进一步研究。

4.2 声调产生的语音演化

在语音学的研究中，我们知道，如果调型不同，声调的感知属于范畴感知，而在调型相同或相近音高不同的情况下，感知的范畴会减弱或消失。另一方面，声母根据发音方法的不同，后接元音的音高有显著的差别。清声母后接元音的音调高，而浊声母后接元音的音调要低，这种现象称为辅音的内在音高。根据最近语音学的研究发现，辅音的内在音高和“嗓音起始-接触时间（VAT）”有关（Baken R.J.，Orlikoff R.F.,1998），嗓音起始-接触时间可以反映发声的状态，如，硬起音和软起音，因此可以解释为什么清声母可以导致高基频起点，而浊音可以导致低基频起点。表九是拉萨藏语声调和夏河、玉树音调的举例。

表九、拉萨、夏河和玉树声调或音调举例

声韵特征	夏河	玉树	拉萨
清长	51/51	51/51	55
清短	51/51	51/51	53
清塞	51/51	55/55	51
浊长	2251/51, 251/51, 251/25	5351/51, 351/351, 353/353	15
浊短	2251/51, 251/51	251/51, 351/51, 353/353	13
浊塞	2251/51, 251/51	3151/51, 55/55, 2125/15	132

表九第一列是藏语音节类型，分别由声母清浊和韵母长、短及塞尾表征。在“/”前的数值是实际基频的五度转写，在‘/’后的数值表示感知数值的五度转写。可以清楚地看出，清声母基频和感知的数值比较明确，前后完全相同。但浊

声母的数值比较复杂，实际基频的数值和感知的数值有很大差别，很难用一个数值来描写不同音节的音调和声调。其原因主要有两个，一是藏语在无声调方言中有浊的复辅音声母，其前浊段很长，因而形成了两条基频曲线，会被感知为两个音调，因此无法用一个数值来描写。二是由于浊辅音内在音高的影响，韵母段的基频值为升降曲线。语音学的研究表明，基频为升降曲线时，有可能被感知为三种音调或声调，即升、降和升降。被感知为哪一种取决于升或降段的时长和振幅。

分析表九的数据可以发现声调产生的过程。从语音学的角度分析，浊音声母基频变短、振幅减弱和消失都会对该音节的感知产生影响。这几个因素的变化构成了藏语从无声调音位系统到有声调音位系统演化的语音生理要素或者说生理制约，这种制约是必然的。这也为我们解释藏语声调的产生提供了科学的依据。具体来看，浊长在夏河藏语中可以被感知为 51 和 25 两种音调，在玉树藏语中可以被感知为 51、351 和 353 三种音调，在拉萨被感知为 15 一种声调。浊短类型在夏河被感知为 51 一种音调，在玉树藏语中被感知为 51 和 353 两种音调，在拉萨被感知为 13 一种声调。浊塞类型在夏河被感知为 51 一种音调，在玉树藏语中被感知为 51、55 和 15 三种音调，在拉萨被感知为 132 一种声调。从这些变化不难看出音调到声调的演化过程，这种演化也是在词汇上慢慢扩散的。

5. 结论

根据以上的研究，可以得到以下几点主要的结论：1）从结构的演化来看，同音词的增加是导致藏语产生声调的主要内在动力，但在语音的结构上，声调的产生又受到藏语音位系统结构的制约；2）同音词的分布在藏语各方言中都是遵循指数曲线，这是藏语（汉语也基本如此）音位的基本内在性质；3）从音位负担的演化来看，在藏语信息量为常数的情况下，声、韵、调三者的音位负担可以相互转换，藏语声母负担的增加直接导致了声调的产生，因此，藏语的音调转换为声调；4）提出了“准音位”的感念；5）从语音结构在词汇中的扩散和语音演化看，藏语声、韵母的减少形成了藏语语音结构在词汇中的扩散；6）声调产生过程中，语音的演化主要受制于发音生理机制和感知机制，声母清化是声调产生的感知基础。

根据多年的研究，本文主要讨论了藏语声调产生和起源研究的不同视角及其研究方法。可以看出，藏语声调产生和起源的研究才刚刚起步，还很不系统。因此，藏语声调起源的多维视角将会为藏语声调起源研究奠定一个科学基础，推动藏语乃至汉藏语声调起源的研究，最终从科学的角度解释汉藏语声调演化的本质。

注释

1. 具体的算法见“音位结构功能负担计算方法研究”一文，发表过程中。

致谢

本项研究由中国社会科学基金重大项目支持，批准号：10&ZD125。感谢李永宏博士在计算音位负担量上所做的工作。同时感谢藏语方言所有的发音人和北京大学及西北民族大学参与调查的所有研究生们。

Multi Viewpoints on the Origin of Tibetan Tones

Kong Jiangping[♣]

Department of Chinese Language and Literature, Research Center for Chinese Linguistics, Joint Center for Language and Human Complexity, Peking University

ABSTRACT

The production and origin of Tibetan tones were studied from the multi viewpoints. The following results and conclusion were obtained: 1) homophone is the internal impetus of the production and the origin of Tibetan tones but the origin of Tibetan tones was restricted by the phoneme structure; 2) the distributions of homophone in all the Tibetan dialects are close to exponential curve, which is the basic nature of Tibetan; 3) from the viewpoint of phoneme evolution, the phoneme load could transfer among the initials, finals and tones. The increase of initial loads lead to that the pitch patterns transferred to the tone patterns; 4) according this study, the concept of ‘quasi-phoneme’ was proposed; 5) from the viewpoint of structure diffusion in the lexicon, the structure diffusion were caused by the decrease of initials and finals; 6) in the production and origin of Tibetan

[♣] Kong Jiangping, Department of Chinese Language and Literature, Research Center for Chinese Linguistics, Joint Center for Language and Human Complexity, Peking University. Address: Dept. of Chinese Language and Literature, No. 5, Yiheyuan road, Haidian district, Beijing, China. Post code:100871. jpkong@pku.edu.cn。

tones, the formation of tones were restricted by the phonations and perception physiologically and psychologically. The devocalization of the initial is the basis of the origin of tones.

SUBJECT KEYWORDS

Origin of tones **H**omophone rate **P**honeme load **L**exical diffusion **S**peech perception

参考文献

GESANG Jumian 格桑居冕.1991.Zangwen zixingfa yu guzangyu yinxi 藏文字性法与古藏语音系.*Minzu yuwen* 民族语文 6:12-23.

GESANG Jumian 格桑居冕.GESANG Yangjing 格桑央京.2002.Zangyu fangyan gailun 藏语方言概论.*Minzu chubanshe* 民族出版社.

GESANG Jumian 格桑居冕.GESANG Yangjing 格桑央京.2004.Shiyong zangwen wenfa jiaocheng(xiudingban)实用藏文文法教程.*Sichuan minzu chubanshe* 四川民族出版社.

HUANG Bufan 黄布凡.1994.Zangyu Fangyan shengdiao de fasheng he fenhua tiaojian 藏语方言声调的发生和分化条件.*Minzu yuwen* 民族语文 3:1-9.

KONG Jiangping 孔江平.2010.Guzangyu yinwei xitong de jiegou he fenbu 古藏语音位系统的结构和分布.Yanjiu zhi Le:qingzhu wang shiyuan xiansheng qishiwu shouchen xueshulunwen ji 研究之乐：庆祝王士元先生七十五寿辰学术论文集.Pan Wuyun,Shen Zhongwei Zhubian 潘悟云,沈钟伟主编.*Shanghai jiaoyu chubanshe* 上海教育出版社

KONG Jiangping,YU Hongzhi Deng 孔江平,于洪志等.2011.Zangyu fangyan diaocha zibiao 藏语方言调查表.*Shangwu yinshuguan* 商务印书馆。

MA Xueliang Zhubian 马学良主编.2003.Hanzangyu gailun 汉藏语概论(An Introduction to Sino-Tibetan Languages).*Minzu chubanshe* 民族出版社.

QU Aitang 瞿靄堂.1963.Zangyu gaikuang 藏语概况(An Introduction to Tibetan Language) .*Zhongguo yuwen* 中国语文 6:

QU Aitang 瞿靄堂.1987.Zangyu bufen 藏语部分.zhongguo shehui kexueyuan he aodaliya renwen kexueyuan hebian 中国社会科学院和澳大利亚人文科学院合编.Zhongguo yuyan dituji 中国语言地图集 (Atlas of Chinese Languages),

Longman, Hong Kong.

WANG Shiyuan 王士元.2009.Wang Shiyuan Yuyinxue Lunwen Ji 王士元语音学论文集.*Shijie tushu chuban gongsi beijing gongsi* 世界图书出版公司北京公司.

LIGHT T., 1978, Tonogenesis: Analysis and Implications, *Lingua 46*, (1978) 115-131.

SHAFER R., 1955, Classification of the Sino-Tibetan Languages, *Word*, 11:1.

SUN, Jackson T.-S. 1997. The Typology of Tone in Tibetan. Chinese Languages and Linguistics IV: Typological studies of languages in China, Taipei, Taiwan: Symposium Series of the Institute of History and Philology-Academia Sinica.

SUN, Jackson T.-S. 2003. "Variegated tonal developments in Tibetan." *Language variation: papers on variation and change in the Sinosphere and the Indosphere in honour of James A. Matisoff* ed. by D. Bradley, R. LaPolla, B. Michailovsky and G. Thurgood, 35-51. Canberra: Pacific Linguistics.

KONG Jiangping，2012，A Study on the Origin of Tibetan Tones by Homonym Rate，Yuyanxue Luncong 语言学论丛.第四十五辑.*Shangwu Yinshuguan* 商务印书馆.

W.S-Y. WANG.,1967, *The Measurement of Functional Load*. Phonetica

BAKEN RJ, ORLIKOFF RF. 1998，Estimating vocal fold adduction time from EGG and acoustic records. In: Schutte HK, Dejonckere P, Leezenberg H, Mondelaers B, Peters HF, eds. *Programme and abstract book*: *24th IALP congress, Amsterdam*;:15.