语音乐律研究报告

Status Report of Phonetic and Music Research

2014

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

Correlations between Body Heights and Formant Frequencies in Young Male Speakers: a Pilot Study

Honglin Cao [1,2,3], *Yingli Wang* [4], *Jiangping Kong* [3]

[1] Collaborative Innovation Center of Judicial Civilization, China, Beijing 100088
[2] Key Laboratory of Evidence Science (China University of Political Science and Law), Ministry of Education, China, Beijing 100088
[3] Department of Chinese Language and Literature, Peking University, Beijing 100871
[4] Center of Criminal Technology, Public Security Bureau of Guangdong, Guangzhou 510050
caohonglin@pku.edu.cn, wangyingli776@sina.com, jpkong@pku.edu.cn

Abstract

This paper investigates the relationships between body height and formant frequencies in young male adults. A total of 121 speakers were recorded uttering four Chinese vowels. The body height data of the speakers was close to be normally distributed, with a large range and standard deviation. Three subsets of formant parameters: the individual first four formant frequencies, formant interspaces and the sum of different formant frequencies were investigated. Moderate but significant negative relationships were found between most formant parameters of the three subsets and height. Comparatively, the strongest correlations were found between the subset of different formant frequency summations and height. A regression function was also given through multiple regression analysis. These results imply that formant parameters, especially the sum of different formant frequencies, can provide a relatively accurate indication of male speakers' heights.

Index Terms: Standard Chinese, body height, formant frequencies, correlation, regression

1. Introduction

Formant frequencies and structure are always predicted to provide reliable cues to speaker's body size, especially the height. Hypotheses as such are mostly based on the notion that the most important determinant of formant parameters in humans could be the vocal tract length (VTL), and that the human body height could be correlated with VTL. The latter has been corroborated by the authors in [1], who measured the VTLs of 129 normal humans, aged 2-25 years, and revealed a strong positive correlation between VTL and height (r=0.926, p<0.0001). However, few studies, except for [2], were found to investigate the correlation between VTL and height in adults. The authors in [2] investigated 15 males aged from 24-55 years, and found no significant correlation between VTL and height (r=0.08). The study, however, is based on a small sample, therefore the results are questionable.

Compared with the acquisition of the anatomical data of the vocal tract size, formant frequencies and heights are easier to be measured. Thus, there has been a rich repertoire of studies that examined the relationship between formant frequencies and heights. For adult males, however, the previous studies have yielded controversial results. Some authors found no correlation [3-4]. Some studies found correlations but the results did not reach statistical significance. For example, [5] found that men with low-frequency formant and small formant dispersion (Df, defined as the averaged difference between successive formant frequencies (e.g., (F4−F1)/3 or (F3−F1)/2), and was found to be closely tied to both VTL and body size of macaques [6]) tended to be taller (r=−0.36, p=0.06, N=26). In contrast, some studies revealed significant negative correlations between certain vocalic formants and height in male speakers. For example, [7] investigated five Spanish vowels, and found F2 measured in the vowel /e/ correlated with height significantly (r=−0.57, p<0.01, N=27). [8] explored eight German vowels, and found a moderate but significant correlation between male speakers' heights and F3 of [ø:] (r=−0.44, p<0.01, N=43). By averaging two formants, the author also found the maximal correlation was slightly improved in the case of F3 + F4 of [ø:] (r=−0.46). [9] found heights were significantly negatively correlated with Df ((F3−F1)/2 (i.e., first three formants were analyzed) in their study, r=−0.32, p=0.024, N=50). In [10], the authors recorded lists of isolated vowels, words, and sentences spoken in Canadian English by 34 adult males. Not only Df ((F4−F1)/3 (i.e., first four formants were analyzed) in their study, both when all vowels were analyzed together and when tokens of schwa were analyzed separately), but also all four individual formants of schwa were found to perform a significant effect that varied with height. [11] investigated formant frequencies and dispersions of F1-F4 for the vowels /a:/ and /i:/, and only found a weak but significant negative correlation between F2, F3, and F4 of /i:/ and height (r=−0.260, −0.299, and −0.320, respectively, p<0.05, N=60). Recently, A new measure of formant structure, 'formant position' (Pf, which referred to the average standardized formant values, e.g., (F'1+F'2+F'3+F'4)/4) was introduced by [12]. The authors found the Pf was significantly negatively associated with height in two studies(r=−0.24, p<0.01, N=175; r=−0.38, p<0.05, N=32, respectively), and claimed that Pf was more strongly related to height than Df. The findings were supported by [13], which also found a significant negative relation between Pf and males' heights (r=−0.31, p<0.01, N=176).

Owing to the different methodologies used in the previous studies, it is hard to compare their results directly. Undeniably, some studies showed various problems. For example, the sample sizes were too small (15-43) [3-5, 7-8, 10]; the age ranges were too large (18- 68 years in [9] and 18-50 years in [11]); or the height range was too small [3]. Another influencing factor is that different formant parameters, namely, individual formants, average of two or more formants, Df and Pf were used in different studies.

Based on a large and still growing database, the purpose of this paper is to examine the relationship between body heights of young Chinese adult males and various formant frequencies and derived formant measures.

978-1-4799-4219-0/14/$31.00

2. Method

2.1. Speakers and Recording

121 male speakers aged 19-30 years (mean 24.4 years, standard deviation (SD) 2.8 years) were recruited for this study. All speakers' nationalities were self-identified Han. They were students (accounting for a large proportion), teachers, physicians and public servants. None of them had any noticeable voice and speech disorders. All speakers were able to speak Standard Chinese fluently. The speech materials were four monophthongs: [a] (or the character "啊", [a^{55}]), [ə] ("阿(胶)", [ə^{55}]), [i]("衣", [i^{55}]) and [y] ("淤", [y^{55}])in two styles: sustained form and normal-speaking (other three vowels including [u] were also collected, but not shown here, because the numbers of valid speakers were less than 100). Speakers were instructed to produce the sustained vowels at their comfortable levels of pitch and loudness for at least 2s. They also were required to read the vowels (or characters) naturally. Both styles were repeated twice. The SONY ECM-44B microphone was used to record the materials in sound-attenuated rooms at Peking University and the Second Hospital of Dalian Medical University. All recordings were made at a sampling rate of 22 kHz and 16 bit depth.

2.2. Procedures

2.2.1. Body Height Measurement

The body height was measured without shoes from a steel measuring tape (millimeters) stick affixed to a wall. Measurements were made to the nearest 1 millimeter.

2.2.2. Acoustic analysis

Wavesurfer [14] was chosen to extract formant values using the LPC-based algorithm. The steady-state segment of the vowel was chosen by hand for formant tracking. Generally, the center frequencies of first four formants values (F1-F4) were obtained automatically with the default settings. All measurements were compared with automatically extracted values and visual estimates based on the gray-scale spectrogram (bandwidth 250Hz). When values generated were judged to be incorrect, such as lower formants were skipped or F4 and F5 merged, the LPC spectra would be recomputed with altering(increasing) the LPC order and/or number of formants until the LPC peak displayed overlaid on the spectrogram properly. Meanwhile, if the four repeats of one vowel sounded very different and the formants displayed large discrepancies, the vowel would be discarded from analysis. F1-F3 were measured for all valid vowels. If F4 was not clearly enough on the spectrogram, in few cases, F4 was judged to be unmeasurable. Finally, the formant frequencies of four repeats were averaged for calculation.

We then computed the combination of F1-F4 of each vowel in two different ways and formed other two subsets of formant parameters. One was the interspace between two different formants (DFI), including six new variables: F4−F1, F4−F2, F4−F3, F3−F2, F3−F1 and F2−F1. Another was sum of formant frequencies (SFF) , including eleven new variables: F1+F2, F1+F3, F1+F4, F2+F3, F2+F4, F3+F4, F1+F2+F3, F1+F2+F4, F1+F3+F4, F2+F3+F4 and F1+F2+F3+F4. For correlation and regression analysis purposes, the new three variables (F4−F1, F3−F1 and F2−F1) and another three variables (F1+F2, F1+F2+F3 and F1+F2+F3+F4) were equivalent to the widely used parameter Df and recently found parameter Pf, respectively, though they were not averaged or standardized.

2.2.3. Statistical analysis

Pearson's correlation and multiple regression approach were made to estimate the relationship between height and formant parameters using IBM SPSS, version 19.

3. Results

3.1. Descriptive statistics

Figure 1 shows the histogram of all 121 speakers' heights (quantized in terms of 5 cm bins). The number of speakers across each bin are also displayed. The distribution is close to be normal. The height ranges from 155.0 cm to 197.6 cm (mean 176.6 cm, SD 8.9 cm).

Table 1 provides a summary of the mean, minimum, maximum and the SD of individual F1-F4 measures of all speakers. N stands for the number of valid speakers measured. There are three, one and two speakers' data that were disregarded for the vowel [a], [i] and [y], respectively. Meanwhile, F4 of five and three speakers were unmeasurable for the vowel [a] and [ə], respectively.

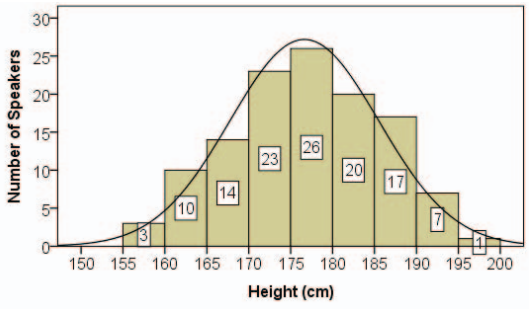


Figure 1: *Histogram showing height distribution*

Table 1. *Mean, min, max and SD for individual F1-F4. N means the number of valid speakers measured.*

V	F(Hz)	Mean	Min	Max	SD	N
[a]	F1	780	646	934	62	118
	F2	1195	1011	1485	92	118
	F3	2644	2117	3013	181	118
	F4	3594	3124	4243	184	113
[ə]	F1	515	408	673	47	121
	F2	1136	932	1468	97	121
	F3	2586	2146	3009	169	121
	F4	3463	2962	3920	217	118
[i]	F1	285	219	359	31	120
	F2	2188	1764	2608	161	120
	F3	3077	2751	3583	169	120
	F4	3658	3130	4269	194	120
[y]	F1	284	225	376	32	119
	F2	1902	1632	2150	105	119
	F3	2265	2003	2628	122	119
	F4	3283	2960	3652	154	119

Table 2. *Pearson's correlation coefficients between formant frequencies of [a], [ə], [i]and [y]and height.*

Formants	Parameters (Hz)	[a]	[ə]	[i]	[y]
Individual formant frequencies (IFF) F1−F4	F1	−0.409***	−0.025	−0.116	−0.155
	F2	−0.358***	−0.365***	−0.484***	−0.446***
	F3	−0.183*	−0.280**	−0.380***	−0.399***
	F4	−0.196*	−0.361***	−0.460***	−0.475***
Different formant interspaces (DFI)	F4−F1	−0.061	−0.369***	−0.426***	−0.444***
	F4−F2	−0.021	−0.205*	−0.063	−0.196*
	F4−F3	−0.008	−0.168	−0.170	−0.165
	F3−F2	−0.001	−0.068	0.078	−0.020
	F3−F1	−0.043	−0.281**	−0.340***	−0.370***
	F2−F1	−0.108	−0.336***	−0.457***	−0.383***
Sum of formant frequencies (SFF)	F1+F2	−0.414***	−0.322***	−0.494***	−0.470***
	F1+F3	−0.288**	−0.260**	−0.411***	−0.402***
	F1+F4	−0.291**	−0.338***	−0.484***	−0.485***
	F2+F3	−0.322***	−0.392***	−0.509***	−0.457***
	F2+F4	−0.318**	−0.430***	−0.542***	−0.529***
	F3+F4	−0.244**	−0.366***	−0.461***	−0.519***
	F1+F2+F3	−0.378***	−0.369***	−0.525***	−0.465***
	F1+F2+F4	−0.369***	−0.408***	−0.555***	−0.539***
	F1+F3+F4	−0.304**	−0.351***	−0.479***	−0.521***
	F2+F3+F4	−0.330***	−0.420***	−0.528***	−0.531***
	F1+F2+F3+F4	−0.369***	−0.404***	−0.541***	−0.536***

Significant level (2-tailed): *p<0.05. **p<0.01. ***p<0.001.

3.2. Correlation and regression for all speakers

The Pearson's correlation coefficients between 21 formant parameters of each vowel and speakers' body heights are shown in Table 2. The 21 parameters are divided into three subsets, which are individual formant frequencies (IFF), DFI and SFF.

For vowel [a], all IFFs (F1-F4) significantly negatively correlate with height. The predictive power of higher formants (F3 and F4), however, was poorer than of the lower formants (F1 and F2). No significant correlations between the six DFI parameters and height were found. In contrast with DFI, significant negative relationship between all eleven SFF variables and height were observed. The minimum and maximum r value were −0.244 (F3+F4) and −0.414 (F1+F2), respectively. The best predictive variable was F1+F2 (r=−0.414, p<0.001), followed by F1 (r=−0.409, p<0.001). For vowel [ə], F2, F3 and F4 (but not F1) were found to be significantly negatively related to height. F4−F2 and three other variables (F4−F1, F3−F1 and F2−F1), which were equivalent to Df, were significantly negatively associated with height. Similarly, all eleven SFF variables were observed to correlate with height significantly negatively. The best predictive variables of three subsets showing in red were F2(r=−0.365, p<0.001), F4−F1(r=−0.369, p<0.001), and F2+F4(r=−0.430, p<0.001), respectively. Compared with vowel [ə], the situation of vowel [i] was basically the same: F2 to F4, except for F1 were significantly negatively linked to height. Three variables equivalent to Df were significantly negatively associated with height. All eleven SFF variables were shown to associate with height significantly negatively. However, the best predictors of three subsets showing in red were F2(r=−0.484, p<0.001), F2−F1(r=−0.457, p<0.001), and F1+F2+F4(r=−0.555, p<0.001), respectively. Generally, it was evident that the predictive power of vowel [i] was better than of vowel [ə]. For vowel [y], the situation was also comparable to vowel [ə] and [i]. We found significant negative correlations between F2, F3, F4, F4−F2, three variables equivalent to Df and all SFF variables and height. The best predictors of three subsets were F4(r=−0.475, p<0.001), F4−F1(r=−0.444, p<0.001), and F1+F2+F4 (r=−0.539, p<0.001), respectively.

As a whole, all formant variables correlated with height negatively, except for F3−F2 of [i]. More than two IFF variables of each vowel significantly negatively correlated with height. Most DFI variables (including Df) of vowel [ə], [i] and [y], but not of [a], were good predictors for height. The SFF variables (including Pf) showed better predictive ability than IFF and DFI. The maximum correlation coefficient of each vowel was found in the SFF subset, which was displayed in red. An example of one significant correlation was illustrated in scatter plots for F1+F2+F4 of vowel [i] with height (Figure 2).

Subsequently, multiple regression analysis was performed to estimate height from formant parameters. Firstly, for each vowel, IFF variables (F1-F4) were selected as the independent variables. Then all IFF variables of the four vowels (4*4=16 in all) were mixed together as the independent variables. Finally, all DFI and SFF variables ((6+11)*4=68 in all) were combined together as the independent variables. Due to the multi-collinearity between independent variables, a stepwise method was chosen for all regression analyses. The selection criteria for adding and removing variables, based on F-test were set by F<0.05 and F>0.1, respectively. Results of five effective models were shown in Table 3. Better results were obtained

when DFI and SFF variables were included in the model. The regression function with the largest R coefficient was:

Height = 300.575 − 0.008* i(F1+F2+F4)

− 0.008* y(F1+F2+F3+F4) − 0.017* ə(F2−F1)

In addition, from Table 2 and Table 3, we could find that the R value for collection of different formants of each vowel was very similar to (slightly bigger than) the *r* value for the sum of homologous formants, suggesting that the new derived variables (SFF) increasing the predictive ability effectively.

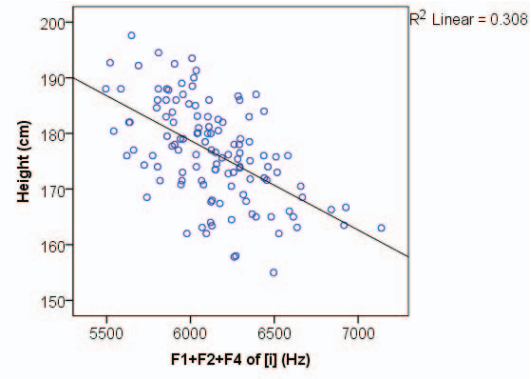


Figure 2: *Scatter plots for F1+F2+F4 of [i](Hz) with height(cm).*

Table 3. *Results of five effective regression models.*

Vowel	Variables	R	adj. R^2	Std.error
[ə]	F2, F4	0.448	0.187	8.006
[y]	F2, F4	0.529	0.268	7.469
[i]	F1, F2, F4	0.566	0.302	7.446
Mixed	i(F4), y(F3), ə (F2)	0.577	0.314	7.319
Mixed	i(F1+F2+F4), y(F1+F2+F3+ F4), ə (F2−F1)	0.619	0.365	7.039

4. Discussion and Conclusion

This study found that there were significant negative relationships between adult male speakers' heights and both individual formant frequencies (IFF) and derived formant variables, namely, DFI and SFF. The results disagreed with the findings of [3-5]. The discrepancies could be attributed to several factors: the size or height range of the sample population were too small for an effective test (which would underestimate the underlying correlation), or different formant variables were used in the earlier studies. We found IFF had strong relationships with height. However, formants with different order and of different vowels showed variations. (As one reviewer pointed out that, it was unclear why the higher correlation happened on certain vowels, like [i] and [y]. Certainly, further studies were needed.) These results were supported by [7-8, 10-11], who found that certain order formant frequency of certain (not all) vowel did have a moderate but significant correlation with height. In the present study, by contrast, most IFFs of all investigated four vowels strongly associated with height. One possible explanation is that the sample size was larger, and the height distribution clearly trended to be normal with a large range and SD in our study. Different from other three vowels, unexpectedly, F1 of [a] showed better estimation accuracy than the higher formants, which needed further studies. In line with studies [12] and [13] based on large database (175 and 176 speakers), which reported that Pf strongly related to height, we found that SFFs as new variables, some of which were equivalent to Pf, showed the strongest relationship with height. Although DFIs (including Df) were also found to correlate with height, the correlations were slightly weaker. One possible explanation was given in [12]: Adding one formant and subtracting another, as in Df calculation, captures information about formant spacing but partly cancels information about formant positions. The results of multiple regression analysis confirmed the ability of SFFs, and also implied the effectiveness of combining the formant variables of different vowels. Although the correlations were significant, anyhow, as the correlation coefficients were not very high ($|r|\leq 0.555$, adjusted $R^2 \leq 0.365$), from a practical point of view, the application of the estimation of adult males' body heights from formant frequencies should still be with caution.

This pilot study carries three limitations. Firstly, only four vowels, other than long materials were investigated. Secondly, one critical hypothesis of the database is that the height of whole adult male population distributes normally, and the current data does not satisfy the real condition of Chinese young adults very well. Meanwhile, as one recent authoritative investigation [15] showed that the mean and SD heights of Chinese school students aged 19 to 22 years were 172.07 cm and 6.19 cm (N=23770), respectively. The corresponding values in our study were both larger comparatively, which would maybe slightly overestimate the underlying correlation. Thirdly, the results were based on high-quality recordings, other than telephone calls with restricted frequency.

In conclusion, the present study revealed that there were moderate but significant negative relationships between adult male speakers' body heights and formant parameters, which included individual formant frequencies, formant interspace and the sum of different formant frequencies. However, as the correlation coefficients are not very high, it is suggested that for forensic purpose the estimation of adult males' heights only based on formant frequencies should still be done with caution. In future studies, more speakers should be recruited to make the height distribution more practical. Other variables, such as standard deviation of fundamental frequency [16] and Long-Term Formant Distribution [17], should also be integrated. The exploration of adult female speakers will be interesting as well. Such investigations are currently undertaken by the authors.

5. Acknowledgements

The authors are grateful to Dr. Michael Jessen for his long-standing suggestions and encouragement in doing this project. They are also like to thank Dr. Youjing Lin, Dr. Yinghao Lee and three anonymous reviewers for their helpful and valuable comments and suggestions received on an earlier version of this paper. This research was funded by the Ministry of Education of China (No. 10JJD740007).

6. References

[1] Fitch, W. T. and Giedd, J., "Morphology and development of the human vocal tract: A study using magnetic resonance imaging", *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1511-1522, 1999.

[2] Hatano, H., Kitamura, T., Takemoto, H., et al., "Correlation Between Vocal Tract Length, Body Height, Formant Frequencies, and Pitch Frequency for the Five Japanese Vowels

Uttered by Fifteen Male Speakers," in *Proc. of INTERSPEECH-2012*, pp. 402-405, 2012.
[3] van Dommelen, W. A. and Moxness, B. H., "Acoustic parameters in speaker height and weight identification: sex-specific behaviour," *Lang Speech*, vol. 38, no. 3, pp. 267-287, 1995.
[4] Collins, S. A., "Men's voices and women's choices," *Anim Behav*, vol. 60, no. 6, pp. 773-780, 2000.
[5] Bruckert, L. Lienard, J. S., Lacroix, A., M., et al., "Women use voice parameters to assess men's characteristics," *Proc Roy Soc B-Biol Sci*, vol. 273, pp. 83-89, 2006.
[6] Fitch, W. T., "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," *J. Acoust. Soc. Am.*, vol. 102, no. 2, pp. 1213-1222, 1997.
[7] Gonzalez, J., "Formant frequencies and body size of speaker: a weak relationship in adult humans," *J Phonetics*, vol. 32, no. 2, pp. 277-287, 2004.
[8] Greisbach, R., "Estimation of speaker height from formant frequencies," *Forensic Linguistics*, vol. 6, no. 2, pp. 265-277, 1999.
[9] Evans, S., Neave, N. and Wakelin, D., "Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice," *Biol Psychol*, vol. 72, no. 2, pp. 160-163, 2006.
[10] Rendall, D., Kollias, S., Ney, C., et al., "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," *J. Acoust. Soc. Am*., vol. 117, no. 2, pp. 944-955, 2005.
[11] Hamdan, A.-L. H., Barazi, R. A., Khneizer, G., et al., "Formant Frequency in Relation to Body Mass Composition," *J Voice*, vol. 27, no. 5, pp. 567-571, 2013.
[12] Puts, D. A., Apicella, C. L. and Cardenas, R. A., "Masculine voices signal men's threat potential in forager and industrial societies," *Proc Roy Soc B-Biol Sci*, vol. 279, pp. 601-609, 2012.
[13] Kempe, V., Puts, D. A. and Cárdenas, R. A., "Masculine Men Articulate Less Clearly," *Hum Nat*, vol. 24, no. 4, pp. 461-475, 2013.
[14] http://www.speech.kth.se/wavesurfer/
[15] Ministry of Education of P. R. China, et al., Reports on the Physical Fitness and Health Research of Chinese School Students, Higher Education Press, 2012. (in Chinese)
[16] Cao, H., Kong, J. and Wang, Y., "Relationship between fundamental frequency and speaker's physiological parameters," *J Tsinghua Univ (Sci & Tech)*, vol. 53, no. 6, pp. 848-851, 2013. (in Chinese)
[17] Nolan, F. and Grigoras, C., "A case for formant analysis in forensic speaker identification," *Int J Speech Lang Law*, vol. 12, no. 2, pp. 143-173, 2005.

Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2013;65:294–302
DOI: 10.1159/000360760

Published online: June 5, 2014

Formant and Voice Source Properties in Two Male Kunqu Opera Roles: A Pilot Study

Li Dong[a] Johan Sundberg[b] Jiangping Kong[a]

[a]Department of Chinese Language and Literature, Peking University, Beijing, China; [b]Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden

Key Words
Kunqu Opera · Formant · Flow glottogram · Electroglottogram

Abstract
Objective: This investigation analyzes flow glottogram and electroglottogram (EGG) parameters as well as the relationship between formant frequencies and partials in two male Kunqu Opera roles, Colorful face (CF) and Old man (OM). ***Participants and Methods:*** Four male professional Kunqu Opera singers volunteered as participants, 2 singers for each role. Using inverse filtering of the audio signal flow glottogram parameters and formant frequencies were measured in each note of scales. Two EGG parameters, contact quotient (CoQ) and speed quotient, were measured. ***Results:*** Formant tuning was observed only in 1 of the OM singers and appeared in a pitch range lower than the passaggio range of Western male opera singers. Both the CF and the OM role singers showed high CoQ values and low values of the normalized amplitude quotient in singing. For 3 of the 4 singers CoQ and the level difference between the first and second partials showed a positive and a negative correlation with fundamental frequency (F0), respectively. ***Conclusions:*** Formant tuning may be applied by a singer of the OM role, and both CF and OM role singers may use a rather pressed type of phonation, CF singers more than OM singers in the lower part of the pitch range. Most singers increased glottal adduction with rising F0.

Introduction

Colorful face (CF) and the Old man (OM) are two important male Kunqu Opera roles. The voice of the CF role is typically described as 'resonant' and 'vigorous' and uses a number of vocal effects. The OM singers, by contrast, use deep and hoarse voice as they play the roles of middle-aged or elderly gentlemen. In the tradition of the Kunqu Opera, both use modal register to sing and recite on stage. However, the details of the phonation and articulation method have not been investigated. The voice quality depends on the muscular, aerodynamic, and acoustical conditions in the glottis and the vocal tract. Therefore, it is relevant to analyze subglottal pressure, voice source and resonance characteristics associated with the voice timbres of these two roles.

In our previous investigations, both similarities and differences were found between the OM role and the CF role with regard to equivalent sound level, fundamental frequency (F0) and long-term-average spectrum (LTAS)

KARGER

E-Mail karger@karger.com
www.karger.com/fpl

1021–7762/14/0656–0294$39.50/0

Jiangping Kong
Department of Chinese Language and Literature
Peking University
Beijing 100871 (China)
E-Mail jpkong@pku.edu.cn

[1, 2]. As compared with the CF role the OM role showed significantly higher equivalent sound level and lower mean F0 in singing. The main LTAS peak of the CF role appeared at higher frequency than that of the OM role and in both roles this peak was higher in frequency and wider in bandwidth than in nonsingers' standard conversational speech. In addition, the CF role singers demonstrated a speaker's formant peak in their LTAS curves. These observations raise the question to what extent these differences emanate from the voice source and the vocal tract. Therefore, in the present article, flow glottogram and electroglottogram (EGG) as well as the formant frequency characteristics will be analyzed.

A flow glottogram shows transglottal airflow versus time. It reflects glottal opening and closure from the perspective of time and airflow amplitude and is commonly obtained by inverse filtering. This technique eliminates the contributions of the vocal tract to the output sound. In cases of high F0 it is unreliable or impossible to use. Hence it is applicable to the voice of CF and OM roles as their F0 ranges are considerably lower than those of other Kunqu Opera roles.

Several parameters are typically used to characterize the voice source, such as the peak-to-peak pulse amplitude and the maximum flow declination rate (MFDR). These parameters seem to be useful to reveal glottal features. The peak-to-peak pulse amplitude has been found to be strongly correlated with the amplitude of the fundamental [3, 4]; MFDR has been found to be closely related to vocal intensity [5], sound pressure level [3], and to subglottal pressure [6]. The ratio between peak-to-peak pulse amplitude and MFDR, i.e. the amplitude quotient (AQ), has also been analyzed in the parameterization of the glottal source [7–9] and has been found to systematically reflect changes in phonation type [9, 10]. However, it typically differs between sexes due to the F0 differences [9]. Therefore, the normalized version of AQ (NAQ), which is the product of AQ and F0, was introduced as a complement to AQ [11]. NAQ has been used both in speech [12, 13] and singing [10, 14] analyses. High AQ or NAQ values have been shown to indicate a less adducted phonation type. Closed quotient (ClQ), defined as the ratio between the closed phase of glottal flow and the period, is a frequently used time-based parameter. It has been found to be associated with the level difference between the first and second voice source partials (H1-H2) [6, 15], and low values of H1-H2 are typically associated with pressed voice and strong high-frequency partials [16–18].

The EGG reflects vocal fold contact [19] and is only indirectly related to glottal airflow [20]. Some EGG parameters appear to be related to phonation type, such as F0, open quotient, contact quotient (CoQ) and speed quotient (SQ) [21]. The open quotient and CoQ are reciprocal parameters; the former is the ratio between the de-contact phase of the EGG signal and the fundamental period while the latter is the ratio between the contact phase and the period. A high CoQ is typically associated with a pressed voice while a low CoQ is mostly observed for breathy voice. The CoQ and the ClQ are not necessarily equal, since transglottal airflow may occur during incomplete glottal closure [22]. The definition of SQ is the ratio between de-contacting phase and contacting phase in the EGG. In other words, high SQ indicates that the glottal closing is quicker and the voice has more high-frequency energy.

Formant tuning or vocal tract tuning is a topic that has been discussed in recent years [23–27]. It increases the sound level of the vocal output, and it is assumed to help the singer avoid register breaks [27–31]. In some male singing voices, formant tuning has been observed in and above the passaggio, between the pitches of E4 and G4 [27, 31], while formant detuning has been found in professional male opera singers [32]. Two types of formant tuning have been reported: (1) tuning F1 and/or F2 to a harmonic partial, and (2) tuning F1 or F2 to a frequency just above its nearest partial. On the other hand some singers have been found to keep F1 and F2 constant and independent of F0.

Method

Four male professional Kunqu Opera singers volunteered as participants, 2 singers for the CF role (CF1 and CF2) and 2 singers for the OM role (OM1 and OM2). CF1, CF2 and OM1, who are performers of the Kunqu Opera Theater of Jiangsu Province, were recorded in a quiet living room, about 4 × 5 × 3 m. OM1 works at the Northern Kunqu Opera Theater and was recorded in an anechoic room, about 3.6 × 2.6 × 2.2 m. Each singer was asked to sing as on stage four songs from their respective role repertoires. A Sony Electret Condenser Microphone, placed off axis on the chest at 21 cm for all singers, was used to record the audio signals. The EGG signal was obtained by an EGG system (Electroglottograph Model 6103; Kay, USA). Those signals were simultaneously recorded and digitized with 16 bits resolution at a sampling frequency of 20 kHz and recorded on dual-channel wav files into the ML880 PowerLab system. A 1-kHz calibration tone was recorded and its sound pressure level, measured by means of a TES-52 Sound Level Meter (TES Electrical Electronic Corp., Taiwan, ROC), was announced at the end of each song. The analyses were conducted on all /a/ vowels extracted from the recordings.

For formant frequency analysis, the files were converted into smp format and calibrated by means of the Soundswell Core Signal Workstation (Hitech Development, Stockholm, Sweden) using the

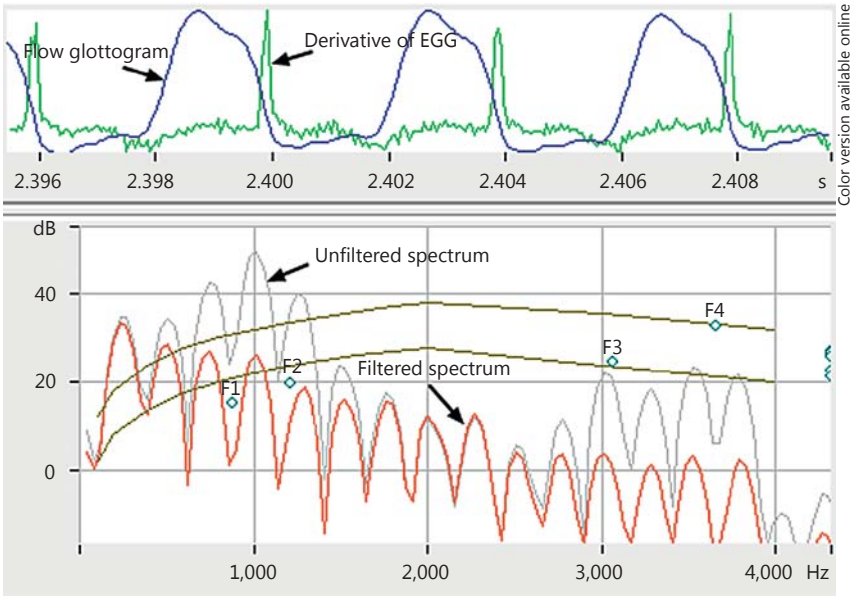


Fig. 1. Example of Decap display showing the flow glottogram and dEGG signals in the upper window and the spectrum before and after inverse filtering in the lower window. The small circles in the lower window represent the frequencies and bandwidths of the inverse filters.

sound level calibration tone. Most of the syllables contain more than one tone. Sections from the middle part of the tone were selected for inverse filtering. The filtering was performed by means of the custom-made DECAP program (Svante Granqvist, KTH). It displays waveform and spectrum in separate windows (fig. 1). The frequencies and bandwidths of the inverse filters are set manually (in the lower window). Then the acoustical signal is filtered with the inverse of the transfer function associated with the given formant frequencies and bandwidths and is displayed as a flow signal (in the upper window). The derivative of the EGG (dEGG) signal was displayed along the same time axis of the flow glottogram and delayed by about 1 ms, corresponding to the travel time of sound from the glottis to the microphone. Four criteria were used to help adjusting the formant frequencies and bandwidths: (1) a ripple-free closed phase; (2) smoothly declining source spectrum envelope; (3) synchronicity between the MFDR and the positive dEGG peak, and (4) the open phase starts no later than the negative dEGG peak. The reliability of these measures was tested by comparing them for 32 samples with those independently obtained by the second author. The results showed no significant difference (paired t test, $p > 0.05$) from the original formant frequency results; parameters of a linear correlation between the original and second measures are listed in table 1.

Analyses of the flow glottograms obtained were carried out using the Snaq module in the Soundswell Core Signal Workstation. After manual marking of the period and of the closed phase, the program calculates the following parameters: (1) F0, (2) MFDR, (3) AQ, (4) NAQ, (5) H1-H2, and (6) ClQ.

The EGG signals were analyzed using Matlab-based VoiceLab (Linguistic Lab of Peking University). The low-frequency component of EGG signals, which was caused by the up and down laryngeal movements, and the high-frequency noise were reduced by means of the wavelet transform. The moments of glottal contact and of loss of glottal contact were approximated using the commonly used 35% of the EGG amplitude criterion (fig. 2). Three parameters were calculated: (1) F0; (2) CoQ and SQ. F0 was transformed into semitones and the mean and standard deviation of CoQ and SQ were calculated for each pitch.

Table 1. Slopes, intercepts, correlations, and standard deviations for the relationships between the formant results determined by two authors

Roles	Formant	Slope	Intercept	R^2	Standard deviation, Hz
CF1	F1	1.54	–441	0.61	28
	F2	1.44	–558	0.43	24
CF2	F1	0.88	111	0.89	10
	F2	0.76	270	0.83	20
OM1	F1	0.87	112	0.84	14
	F2	0.78	227	0.72	11
OM2	F1	1.34	–314	0.95	9
	F2	1.18	–224	0.98	12

Results

Formant Tuning

Figure 3 shows F1, F2 and harmonic partials for 2 CF and 2 OM singers. The relationships between formants and F0 were analyzed by means of linear regression, showing the R^2 and slope values listed in table 2. F1 and

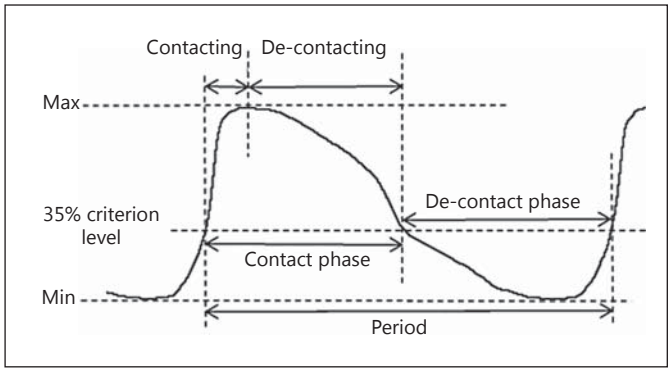


Fig. 2. Example of an EGG waveform. The top and bottom horizontal dashed lines represent maximum and minimum amplitude and the middle one 35% of the EGG amplitude used for defining the contact phase. SQ is defined as the ratio between the de-contacting and contacting phases.

Table 2. Correlations and slope constants for the relationships between F0 and F1 and between F0 and F2 for the four participants

Singer	F1		F2	
	R^2	slope	R^2	slope
CF1	0.29**	0.58	0.20**	0.36
CF2	0.33**	0.23	0.86**	0.85
OM1	0.00	0.05	0.22*	0.50
OM2	0.41**	0.50	0.56**	1.12

Correlations significant at the * $p < 0.05$ and ** $p < 0.01$ levels (two-tailed).

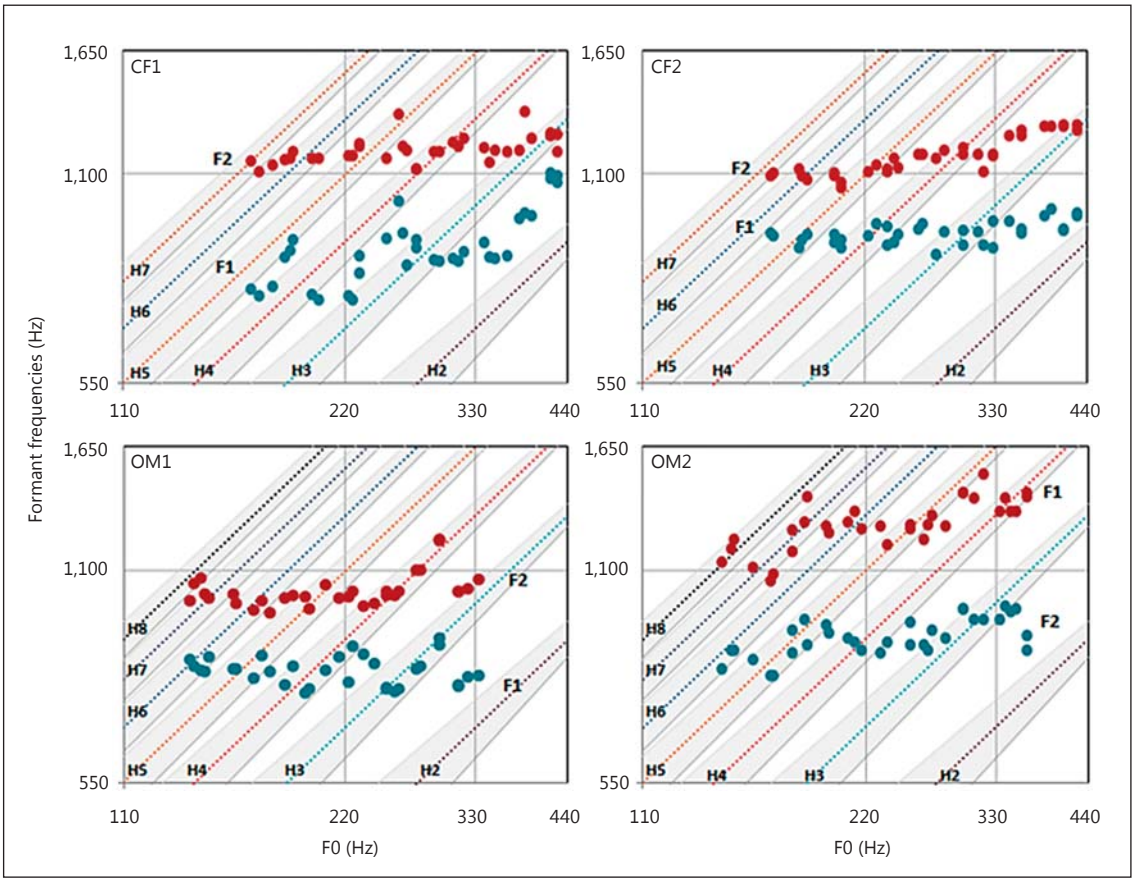


Fig. 3. F1, F2 (big dots) and the harmonics (dotted lines marked with H2–H8). The gray areas represent the frequency ranges that meet the 50-Hz criterion of formant tuning.

Table 3. Scale tones where F1 and F2 fell within 50 Hz of a harmonic in all measurements of that pitch for the indicated singers

Singers	Pitch range	Pitch where F1 is close to a harmonic	Pitch where F2 is close to a harmonic
CF1	E3–G#4		C#4–D#4, G#4
CF2	E3–G#4	F3, F#3, A3, D4	F3, F#3, G#3, A3, D4, G#4
OM1	C#3–E4	D3, D#3, G3, G#3, B3–D4	C#3, D3, E3, G3, G#3, A#3–D4
OM2	C#3–F#4	C#3–F3, G#3, A3, D#4	C#3, D#3, A3, B3, E4

The second column shows the singers' pitch ranges.

Table 4. Pearson correlation between indicated flow glottogram parameters

Singers	Parameters	F0	MFDR	AQ	NAQ	H1-H2	ClQ
CF1	F0	1.00	0.66*	–0.83*	–0.15	–0.75*	0.35*
	MFDR	0.66*	1.00	–0.78*	–0.55*	–0.72*	0.42*
	AQ	–0.83*	–0.78*	1.00	0.65*	0.73*	–0.33
	NAQ	–0.15	–0.55*	0.65*	1.00	0.29	–0.22
	H1-H2	–0.75*	–0.72*	0.73*	0.29	1.00	–0.59*
CF2	F0	1.00	0.54*	–0.91*	0.27	–0.67*	0.38*
	MFDR	0.54*	1.00	–0.50*	0.19	–0.54*	0.29
	AQ	–0.91*	–0.50*	1.00	0.12	0.74*	–0.52*
	NAQ	0.27	0.19	0.12	1.00	0.05	–0.25
	H1-H2	–0.67*	–0.54*	0.74*	0.05	1.00	–0.78*
OM1	F0	1.00	0.57*	–0.84*	0.23	–0.45*	0.18
	MFDR	0.57*	1.00	–0.64*	–0.28	–0.42*	0.14
	AQ	–0.84*	–0.64*	1.00	0.31	0.37*	–0.31
	NAQ	0.23	–0.28	0.31	1.00	–0.10	–0.19
	H1-H2	–0.45*	–0.42*	0.37*	–0.10	1.00	–0.46*
OM2	F0	1.00	0.57*	–0.86*	0.02	–0.24	–0.20
	MFDR	0.57*	1.00	–0.65*	–0.36*	–0.49*	0.06
	AQ	–0.86*	–0.65*	1.00	0.47*	0.23	0.09
	NAQ	0.02	–0.36*	0.47*	1.00	–0.02	–0.08
	H1-H2	–0.24	–0.49*	0.23	–0.02	1.00	–0.70*

F0 in semitones. Correlations significant at the * $p < 0.05$ level (two-tailed).

F2 tended to increase significantly with F0 ($p < 0.01$) for CF1, CF2 and OM2, while for OM1 significant correlation was found only for F2 ($p < 0.05$). The slope values are smaller than 1 except for F2 of CF2.

F1 and F2 are close to a partial at several pitches (fig. 3). A frequency difference less than 40 or 50 Hz between F1 or F2 and its nearest partial has been applied as a criterion of formant tuning [33, 34]. Table 3 lists the tones where the 50-Hz criterion was met in our material. F1 or F2 was close to a harmonic in many cases, especially for OM1, although for nonadjacent scale tones. For most cases in the table the formant was close to but not identical with the frequency of the partial.

Flow Glottogram Data

Many flow glottogram parameters analyzed show a correlation with other such parameters, as can be seen in table 4. For all singers: (1) F0 is significantly and positively correlated with MFDR and negatively correlated with AQ; (2) MFDR is significantly negatively correlated with AQ and H1-H2. H1-H2 is positively correlated with AQ and negatively correlated with F0, but only significantly

Folia Phoniatr Logop 2013;65:294–302
DOI: 10.1159/000360760

Table 5. Slopes of the linear regression between the indicated pairs of parameters

Singer	Parameter pair (independent, dependent)						
	F0, MFDR	F0, AQ	F0, H1-H2	MADR, H1-H2	AQ, MFDR	AQ, H1-H2	H1-H2, ClQ
CF1	66	–0.024	–0.6	–0.006	–2,764	20	–0.010
CF2	74	–0.017	–0.5	–0.003	–3,649	27	–0.015
OM1	132	–0.020	–0.2	–0.001	–6,248	6	–0.016
OM2	121	–0.027	–0.1	–0.001	–4,464	4	–0.019

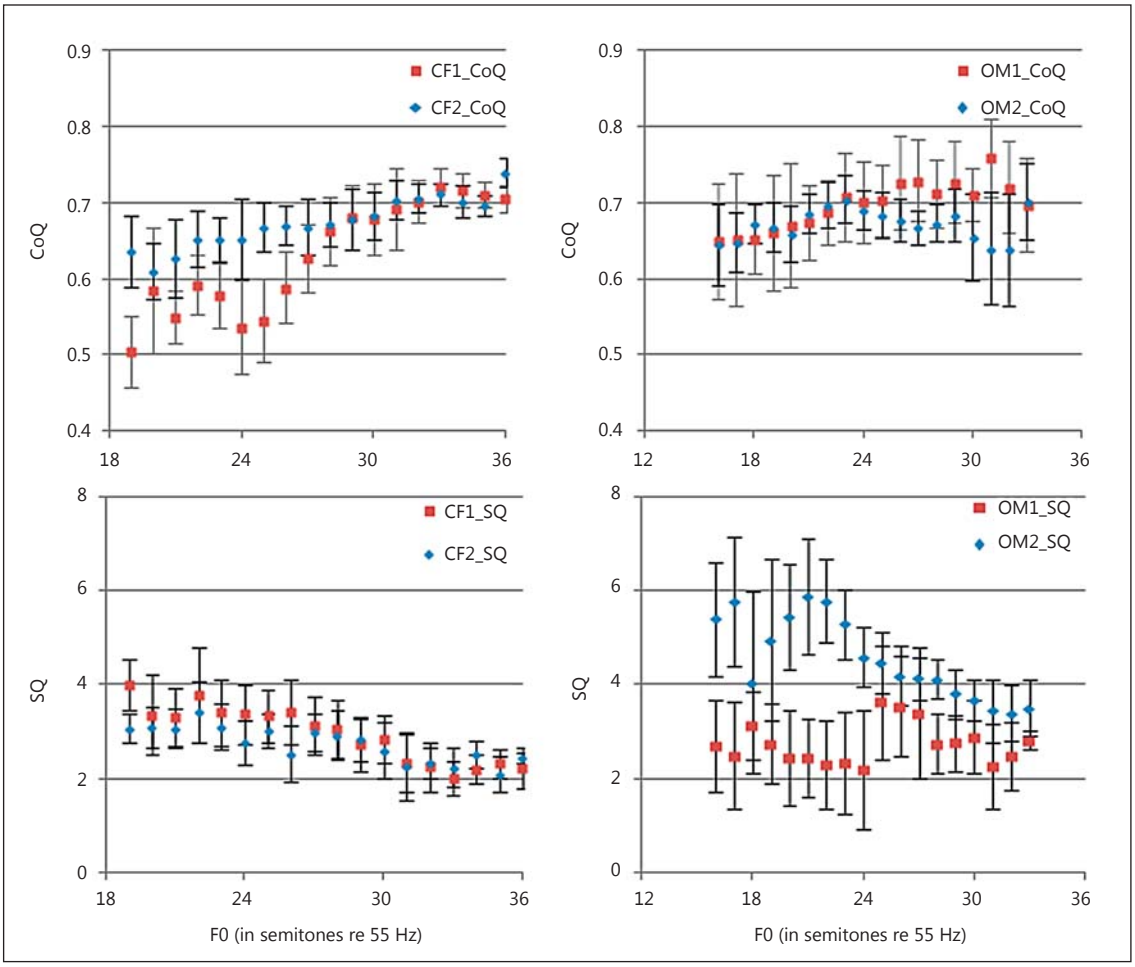


Fig. 4. Mean value and standard deviations of CoQ and SQ as a function of F0 in semitones re 55 Hz.

for CF1, CF2 and OM1. The correlation between H1-H2 and ClQ is negative and significant for CF1, CF2 and OM2, which is consistent with previous research [6, 15].

Linear regression analysis was carried out for parameters that showed a significant correlation. Table 5 shows the slopes. In some cases the slopes clearly differed between roles: MFDR increased more with F0 for OM than for CF; H1-H2 more with AQ for CF than for OM; H1-H2 decreased more with MFDR for CF than for OM, and H1-H2 decreased more with F0 for CF than for OM.

Table 6. Pearson correlation between the indicated EGG parameters

Singer		F0	CoQ	SQ
CF1	F0	1	0.75*	–0.62*
	CoQ	0.75*	1	–0.48*
	SQ	–0.62*	–0.48*	1
CF2	F0	1	0.55*	–0.55*
	CoQ	0.55*	1	–0.23*
	SQ	–0.55*	–0.23*	1
OM1	F0	1	0.40*	0.04*
	CoQ	0.40*	1	–0.02
	SQ	0.04*	–0.02	1
OM2	F0	1	–0.02	–0.54*
	CoQ	–0.02	1	0.19*
	SQ	–0.54*	0.19*	1

F0 in semitones. Correlations significant at the * $p < 0.05$ level (two-tailed).

For each pitch in the range E3–E4 a paired t test was run for each of these parameters between singers. No parameter showed any difference between roles. Singers often show differences in phonation characteristics between high and low notes. Therefore the comparisons with pitch groups seemed relevant. The pitch of A#3 was chosen as the boundary between high and low notes, since the EGG parameters showed break points in the vicinity of this note (fig. 4). A paired t test for each parameter in the same pitch group between singers showed that, in the low pitch range, H1-H2 was significantly higher for the OM role than for the CF role.

EGG Data

Figure 4 shows the mean value and standard deviation of CoQ and SQ at each pitch for 4 singers. CoQ differs between singers of the same role, for the CF role in the low pitch range (E3 to C4, note 19–27, 165–262 Hz), and for OM role in the high pitch range (A#3 to F#4, note 25–33, 233–370 Hz). As for SQ, the 2 CF singers are alike while the 2 OM singers differ, particularly in the low pitch range (A#3 to F#4, note 25–33, 233–370 Hz).

The result of a Pearson correlation between F0, CoQ and SQ is listed in table 6. For singers CF1 and CF2, F0 and CoQ show significant positive correlation; SQ is negatively correlated with F0 and CoQ. For OM1, F0 and CoQ show significant positive correlation. The SQ of OM1 is discontinuous at A#3 (fig. 4). Both below and above A#3, it is significantly and negatively correlated with F0. OM2, by contrast, shows a significant negative correlation and his SQ shows a significant positive correlation with CoQ.

Discussion and Conclusion

Assessing occurrence of formant tuning requires two conditions to be fulfilled: (1) the formant frequency data must be accurate and (2) there must be a convincing formant tuning criterion. In our case, where the F0 distance between the data points is small, another condition could be added: (3) that the separation between the formant and a partial is kept small with changing F0.

Our formant data were derived from inverse filtering, a well-established method in voice research [35]. The method often reflects even small errors in filter settings in terms of drastic flow glottogram deviations from a physiologically realistic shape, or by source spectrum envelope peaks or valleys near formants. Also, independent determination by the second author showed no significant difference from the original formant frequency results. Hence, our formant frequency data can be assumed to be reliable.

Adopting the definition of formant tuning proposed by Henrich et al. [34] we tentatively applied a maximum separation of 50 Hz between partial and formant. This frequency difference corresponds to about 2 semitones in the vicinity of the pitch of E4. At lower F0 this criterion becomes increasingly generous. The F0 range of the OM role goes down to about 140 Hz, so in this case the 50-Hz criterion needs to be applied with caution. For a low F0, such as 150 Hz, the probability that this criterion is met is 67%, while for a high F0, such as 400 Hz, the probability decreases to 25%.

Alternatively, a semitone criterion can be applied. However, this criterion is also associated with problems. For high order harmonics, 1 semitone is too large; for example the separation between partials number 8 and 9 is about 2 semitones, so here any formant frequency would meet the 1-semitone criterion. The semitone criterion is also problematic in the low F0 range since here formant tuning will apply to rather high partials.

A third criterion would be that the formant is systematically shifted between scale tones such that the distance to its nearest partial is constantly kept narrow. This pattern was observed in the case of OM1 in the pitch range B3 to D4. Incidentally, this would have contributed to the lack of correlation between F1 and F0, as shown in table 2. It may also be mentioned that this pitch range is lower

than the passaggio of Western male singers, where formant tuning has been reported [27, 31]. A possible reason why the Kunqu Opera singers apparently avoid formant tuning is that they strive to maintain vowel quality reasonably independent of F0.

Also by applying the 50-Hz criterion we noted formant tuning for singer OM1 between B3 and D4. Here, the maximum distances between formants and partials were 33 and 45 Hz for F1 and F2, respectively. Thus, although the 50-Hz criterion is not perfect, it is acceptable in this pitch range.

NAQ, assumed to reflect glottal abduction [11], showed no correlation with F0 for our Kunqu singers. This suggests that they keep the same degree of glottal abduction throughout their pitch range. By contrast, Björkner et al. [10] noted that Western baritone singers showed higher NAQ values on a high than on a low F0. This suggests that the Kunqu singers produce their high pitches with firmer glottal adduction than what seems common among Western baritone singers. Moreover, normal and pressed voice in normal Western speakers correspond to NAQ values in the range 0.13–0.16 and 0.10–0.15, respectively [9]. The CF and OM singers' NAQ values were remarkably low, ranging between 0.06 and 0.12, thus suggesting use of a more hyperfunctional type of phonation than is common in the Western world.

MFDR, a parameter closely correlated with subglottal pressure and representing the strength of vocal tract excitation, was significantly higher for the OM role than for the CF role. This gives reason to expect that the OM singers were singing more loudly than the CF singers. This expectation is corroborated by data published elsewhere [1]. Furthermore it can be noted that the OM singers increased their MFDR about twice as much with increasing F0 than the CF singers.

H1-H2 has been extensively used in descriptions of voice source characteristics, low values typically being associated with pressed voice and strong high-frequency partials [16–18]. In the low pitch range, the CF singers showed significantly lower H1-H2 values than the OM singers. This is consistent with the LTAS analysis showing that these singers, unlike the OM singers, demonstrated a speaker's formant near 3 kHz [2]. H1-H2 tended to decrease with increasing F0 only for CF voices, suggesting a more hyperfunctional phonation at high pitches. This conclusion is supported also by the CoQ that tended to increase with F0, particularly for the CF singers. On the other hand, as mentioned above the NAQ values showed no variation with F0. It seems that the relationships between these flow glottogram parameters need to be further explored in future research.

Most of the mean CoQ for the 4 singers are higher than the CoQ of pressed phonation reported in previous investigations, which like us have applied a criterion threshold of 35% of the EGG amplitude for defining CoQ [36]. Hence, the current participants used pressed voice when singing. However, as illustrated in figure 4 different singers of the same role behaved somewhat differently. CF1 varied his CoQ substantially in the low pitch range as shown by the large standard deviations. Thus, he apparently varied phonation type between normal and pressed, and sometimes even breathy. On the other hand, CF2 kept his phonation mode more constant. Both OM1 and OM2 tended to increase CoQ with F0 in the low pitch range. Their CoQ values were rather high, suggesting that they both maintained a rather pressed phonation voice.

In conclusion, of the 4 singers only OM1 demonstrated formant tuning. It was observed in the pitch range B3 to D4, which is lower than the passaggio range of Western male opera singers, where formant tuning has been observed. With regard to phonation type, both the CF and the OM role singers showed high CoQ values and low NAQ values in singing, which suggests a rather pressed type of phonation. In the low pitch range, the CF singers seemed to use a more pressed phonation than the OM singers, and they also showed stronger energy in the high-frequency range of the spectrum. For 3 of the 4 singers analyzed, CoQ and H1-H2 were positively and negatively correlated with F0, respectively, which suggests that with rising F0 these singers increased glottal adduction and modified their phonation towards a more pressed type.

Acknowledgments

This research was funded by the National Social Sciences Foundation of China and China Scholarship Council; grant numbers were 10&ZD125 and 201206010134, respectively.

References

1 Dong L, Sundberg J, Kong J: Loudness and pitch of Kunqu Opera. J Voice 2014;28:14–19.
2 Dong L, Kong J, Sundberg J: Long-term-average spectrum characteristics of Kunqu Opera singers' speaking, singing and stage speech. Logoped Phoniatr Vocol 2013, Epub ahead of print. http://informahealthcare.com/doi/abs/10.3109/14015439.2013.841752.
3 Gauffin J, Sundberg J: Spectral correlates of glottal voice source waveform characteristics. J Speech Hear 1989;32:556–565.
4 Fant G: The voice source in connected speech. Speech Commun 1997;22:125–139.
5 Fant G, Liljencrants J, Lin Q: A four-parameter model of glottal flow. STL-QPSR, KTH 4/1985, pp 1–13. http://2.inarchive.com/1206/14/213/GpHUDG.pdf.
6 Sundberg J, Andersson M, Hultqvist C: Effects of subglottal pressure variation on professional baritone singers' voice sources. J Acoust Soc Am 1999;105:1965–1971.
7 Fant G, Kruckenberg A, Liljencrants J, Båvegård M: Voice source parameters in continuous speech: transformation of LF-parameters. Proc ICSLP-94, Yokohama, 1994, vol 3, pp 1451–1454.
8 Alku P, Vilkman E: Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering. Speech Commun 1996;18:131–138.
9 Alku P, Vilkman E: A comparison of the glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. Folia Phoniatr Logop 1996;48:240–254.
10 Björkner E, Sundberg J, Alku P: Subglottal pressure and normalized amplitude quotient variation in classically trained baritone singers. Logoped Phoniatr Vocol 2006;31:157–165.
11 Alku P, Bäckström T, Vilkman E: Normalized amplitude quotient for parameterization of the glottal flow. J Acoust Soc Am 2002;112:701–710.
12 Gobl C, Ní Chasaide A: Amplitude-based source parameters for measuring voice quality. Proc. ISCA Tutorial and Res Workshop VOQUAL'03 on Voice Quality. Geneva, 2003, pp 151–156. http://www.isca-speech.org/archive_open/archive_papers/voqual03/voq3_151.pdf.
13 Airas M, Alku P: Emotions in vowel segments of continuous speech: analysis of the glottal flow using the normalised amplitude quotient. Phonetica 2006;63:26–46.
14 Sundberg J, Thalén M, Alku P, Vilkman E: Estimating perceived phonatory pressedness in singing from flow glottograms. J Voice 2004;18:56–62.
15 Holmberg EB, Hillman R, Perkell JS, Guiod P, Goldman S: Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. J Speech Hear Res 1995;38:1212–1223.
16 Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L: Perceptual and acoustic correlates of abnormal voice qualities. Acta Otolaryngol 1980;90:441–451.
17 Klatt DH, Klatt LC: Analysis, synthesis, and perception of voice quality variations among females and male talkers. J Acoust Soc Am 1990;87:820–857.
18 Hanson HM: Glottal characteristics of female speakers: acoustic correlates. J Acoust Soc Am 1997;101:466–481.
19 Titze IR: Interpretation of the electroglottographic signal. J Voice 1990;4:1–9.
20 Rothenberg M: Some relations between glottal air flow and vocal fold contact area. Proc Conf on the Assessment of Vocal Pathol, ASHA Rep No 11, 1979, pp 88–96.
21 Kong J: On Language Phonation (in Chinese). Beijing, Central Nationalities University Press, 2001.
22 Herbst C, Ternström S: A comparison of different methods to measure the EGG contact quotient. Logoped Phoniatr Vocol 2006;31:126–138.
23 Schutte HK, Miller DG, Duijnstee M: Resonance strategies revealed in recorded tenor high notes. Folia Phoniatr Logop 2005;57:292–307.
24 Titze IR: A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice. J Voice 2004;18:292–298.
25 Titze IR, Worley AS: Modeling source-filter interaction in belting and high-pitched operatic male singing. J Acoust Soc Am 2009;126:1530–1540.
26 Sundberg J, Lã FMB, Gill BP: Professional male singers' formant tuning strategies for the vowel /a/. Logoped Phoniatr Vocol 2011;36:156–167.
27 Neumann K, Schunda P, Hoth S, Euler HA: The interplay between glottis and vocal tract during the male passaggio. Folia Phoniatr Logop 2005;57:308–327.
28 Coffin B: Overtones of Bel Canto. New Brunswick, Scarecrow Press, 1980.
29 Hertegård S, Gauffin J, Lindestad PÅ: A comparison of subglottal and intraoral pressure measurements during phonation. J Voice 1995;9:149–155.
30 Miller DG, Schutte HK: Formant tuning in a professional baritone. J Voice 1990;4:231–237.
31 Doscher BM: The Functional Unity of the Singing Voice. London, The Scarecrow Press, 1994.
32 Sundberg J, Lã F, Gill BP: Formant tuning strategies in professional male opera singers. J Voice 2013;27:278–288.
33 Sundberg J, Titze I, Scherer R: Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. J Voice 1993;7:15–29.
34 Henrich N, Smith J, Wolfe J: Vocal tract resonances in singing: strategies used by sopranos, altos, tenors, and baritones. J Acoust Soc Am 2011;129:1024–1035.
35 Rothenberg M: A new inverse‐filtering technique for deriving the glottal air flow waveform during voicing. J Acoust Soc Am 1973;53:1632–1645.
36 Verdolini K, Druker DG, Palmer PM, et al: Laryngeal adduction in resonant voice. J Voice 1998;12:315–327.

Loudness and Pitch of Kunqu Opera

***Li Dong, †Johan Sundberg, and *Jiangping Kong,** **Beijing, China and †Stockholm, Sweden*

Summary: Equivalent sound level (Leq), sound pressure level (SPL), and fundamental frequency (F_0) are analyzed in each of five Kunqu Opera roles, *Young girl* and *Young woman*, *Young man*, *Old man*, and *Colorful face*. Their pitch ranges are similar to those of some western opera singers (alto, alto, tenor, baritone, and baritone, respectively). Differences among tasks, conditions (stage speech, singing, and reading lyrics), singers, and roles are examined. For all singers, Leq of stage speech and singing were considerably higher than that of conversational speech. Interrole differences of Leq among tasks and singers were larger than the intrarole differences. For most roles, time domain variation of SPL differed between roles both in singing and stage speech. In singing, as compared with stage speech, SPL distribution was more concentrated and variation of SPL with time was smaller. With regard to gender and age, male roles had higher mean Leq and lower average F_0, MF0, as compared with female roles. Female singers showed a wider F_0 distribution for singing than for stage speech, whereas the opposite was true for male singers. The Leq of stage speech was higher than in singing for young personages. Younger female personages showed higher Leq, whereas older male personages had higher Leq. The roles performed with higher Leq tended to be sung at a lower MF0.

Key Words: Equivalent sound level–Sound pressure level–Fundamental frequency–Kunqu Opera–Task–Condition–Singer–Role.

INTRODUCTION

The Kunqu Opera is a traditional performing art in China. It has been handed down orally since the middle of the 16th century and is revered as the ancestor of all Chinese Operas. It is commonly praised for its elegant phrases, wonderful stories, and beautiful melodies and is performed by at least 10 artists, Jing, Guansheng, Jinsheng, Laosheng, Fumo, Zhengdan, Guimendan, Liudan, Fuchou, and Xiaochou, each with a special voice timbre.[1] The roles can be divided into five groups, namely:

1. Sheng (*Young man* roles) recites and sings in both modal and falsetto register. Both Guansheng, who wears an officer's hat, and Jinsheng, who wears a headband, change their voice quality according to the age and identity of the personages. A Guansheng performer acts as a young king or a gifted scholar, and his voice quality has been described as "broad and bright" having "a heavy oral resonance." Jinsheng performers often act in love stories and sing with a brighter, lyrical voice.
2. Dan (*Female* roles) includes Laodan (*Old woman* role), Zhengdan (*Middle-aged woman* role), Guimendan (*Young woman* role), and Liudan (*Young girl* role). To portray their different ages and identities, Dan performers sing with different voice qualities; in general, the older the personage, the greater the proportion of modal voice. Thus, Laodan performers recite and sing with loud modal voice, Liudan performers with falsetto voices, whereas Zhengdan and Guimendan use both these registers.
3. Jing (*Colorful face* roles) performers sing with their faces painted in different colors depending on the identity of the personage. The voice quality has been described as "resonant and vigorous." Often, they use a series of special effects to display different characters, such as voice bursts and "intense resonance."
4. Mo (*Old male* roles), including Laosheng (*Old man* role) and Fumo (second *Old man* role), recite and sing in modal register. Laosheng performers play the roles of middle-aged or elderly gentlemen. The Fumo performer introduces the story at the beginning of the performance.
5. Chou (*Buffoon* roles), including Xiaochou (*Clown* role) and Fuchou (second *Clown* role), recite and sing with register shifts between falsetto and modal. Fuchou pays more attention to expression than to voice. Xiaochou is a comical role performed with a loud and clear voice.

Summarizing, the voice timbres mirror the ages, characters, and identities of the various personages. The voice qualities deviate dramatically from both conversational speech and Western operatic tradition, which have been well described in previous research.[2–4] In contrast, few attempts have been made to describe the acoustic characteristics of Kunqu Opera roles in scientific terms, although these characteristics possess a general relevance from the point of view of voice science, illustrating the flexibility of the human voice and exemplifying how the voice can be used in artistic, musical, and dramatic contexts. The present study investigates the 1) differences among roles; 2) differences among singing, stage speech (also called recitative in Peking Opera[5]), and reading lyrics; 3) intrarole differences between songs; and 4) differences between singers of the same role. The investigation focusses on two primary acoustic properties of the voice, loudness and fundamental frequency (F_0), in five Kunqu Opera roles, two female (*Young girl* and *Young woman*), and three male (*Colorful face*, *Old man*, and *Young man)*.

METHODS

Four female and six male professional performers of Kunqu Opera, aged 25–47 years, volunteered as subjects, two

Accepted for publication July 29, 2013.

From the *Department of Chinese Language and Literature, Peking University, Beijing, China; and the †Department of Speech Music Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden.

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

Journal of Voice, Vol. 28, No. 1, pp. 14-19
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.07.012

TABLE 1.
Ages (y) of the Ten Performers

Roles	Young Girl	Young Woman	Colorful Face	Old Man	Young Man
Singer 1	45	47	27	46	45
Singer 2	41	27	25	44	27

performers in each of five roles (Table 1). Their professional experiences varied between 7 and 27 years. The singers were told to sing just as on stage. As there are no songs that are common to all these roles, the singers were asked to perform three or four songs of their own choice that belonged to their repertoire at the time of the recording. The songs had duration of between 2 and 3 minutes and differed in emotional color. The two *Young girl* singers sang only three songs because one of the songs was very long. The singers also recited a section of stage speech. In addition, all singers read, in modal voice, the lyrics of the songs chosen, duration between 2 and 3.5 minutes. The language differed from Mandarin Chinese but was identical with what they used in their roles on stage, which actually corresponds to ancient Chinese.

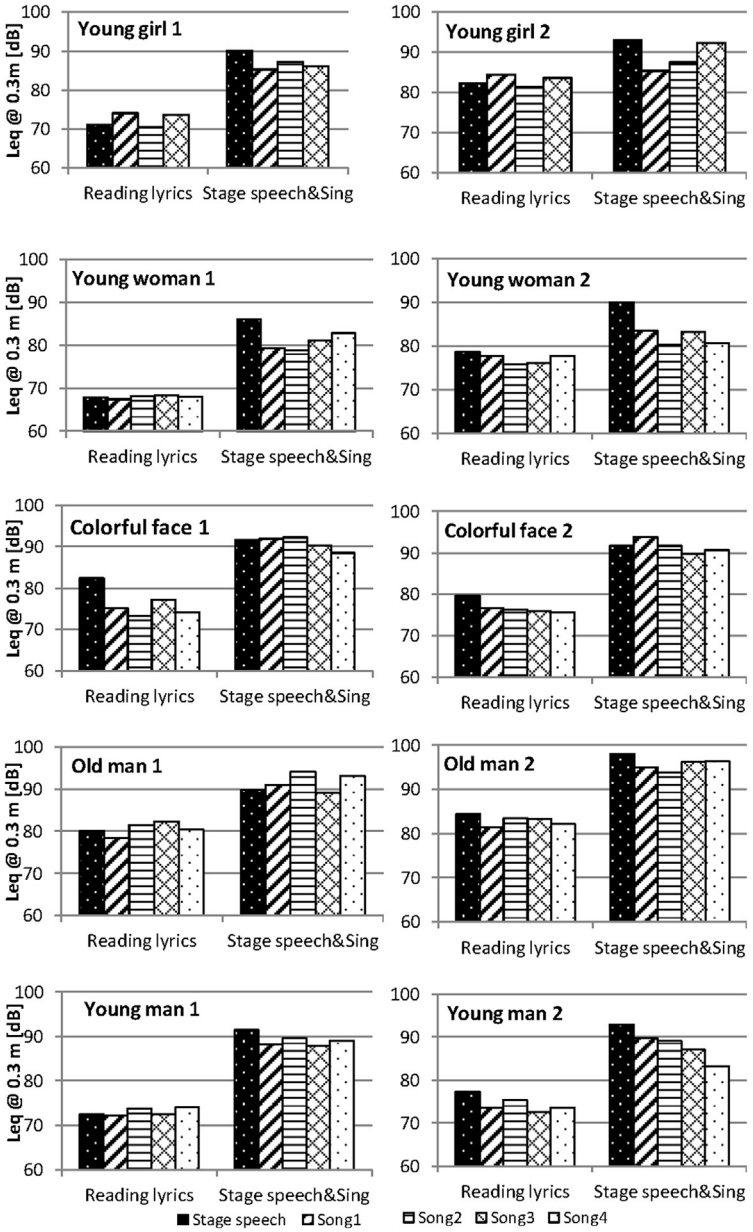


FIGURE 1. Leq at 0.3 m for reading lyrics, stage speech, and singing. In each panel, the left group of columns shows the Leq values of the reading of the different lyrics, and the right group the Leq values of stage speech and three or four songs.

TABLE 2.
The Leq and SPL at 0.3 m Averaged Across the Three or Four Songs Sung and Spoken by the 10 Kunqu Opera Singers

	Reading Lyrics		Singing		Stage Speech	
Subject	$Mean_{Leq}$	$Mean_{SPL}$	$Mean_{Leq}$	$Mean_{SPL}$	Values	$Mean_{SPL}$
Young girl I	73.4	65.52	86.3	77.48	90.2	78.38
Young girl 2	83	73.84	89.3	79.24	93	80.47
Young woman 1	68	60.91	80.7	74.04	86.1	78.39
Young woman 2	77.1	67.8	82.7	74.74	90	81.1
Colorful face 1	75.4	66.42	90.5	83.66	91.7	80.14
Colorful face 2	76.2	68.02	92.1	86.34	91.7	83.61
Old man 1	81.2	71.08	92.6	83.8	89.9	77.52
Old man 2	82.7	72.75	95.4	87.9	98	87.27
Young man 1	73.2	64.06	88.8	81.71	91.5	81.15
Young man 2	73.5	64.85	87.9	79.53	92.9	80.36

The columns marked Reading lyrics refer to the singers' reading of the song lyrics.

Young girl singer 2 and *Old man* singer 2, who both are performers of the Northern Kunqu Opera Theater, could be recorded in an anechoic room, about 3.6 × 2.6 × 2.2 m, as they lived in Beijing, the city where the research was carried out. The other singers, who were performers at the Kunqu Opera Theater of the Jiangsu Province, had to be recorded in an ordinary room, about 4 × 5 × 3 m. Audio was picked up by a Sony Electret Condenser Microphone placed off axis at a measured distance that varied between 15 and 21 cm for the different singers. All sound level data were normalized to 30 cm. The signals were digitized on 16 bits at a sampling frequency of 20 kHz and recorded on single channel wav files into ML880 PowerLab system. Sound pressure level (SPL) calibration was carried out by recording a 1000-Hz tone, the SPL of which was measured at the recording microphone by means of a TES-52 Sound Level Meter (TES Electrical Electronic, Corp., Taiwan, China). This SPL value was announced in the recording file together with respective microphone distance.

Two programs were used for analyzing the recordings. WaveSurfer-1.8.8p3 was used to measure the F_0. After converting the files into the smp format and eliminating pauses longer than 10 milliseconds from the recordings, the *Soundswell Core Signal Workstation 4.0* was used to analyze the equivalent sound level (Leq). The distribution of SPL values was determined by means of the Soundswell Histogram module. Statistic analyses were completed using *SPSS* 18. Given the small sample Leq and mean F_0 ($N \leq 8$), the mean values were compared by *t* test. For the larger sample of time variation of SPL ($N > 360$), a Mann-Whitney *U* test was used.

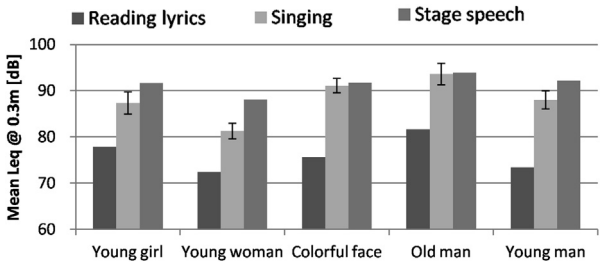


FIGURE 2. Leq at 0.3 m, averaged across subjects for the indicated conditions. The bars represent ±one standard deviation.

RESULTS

Figure 1 shows the Leq for the different singers and tasks. The within-subject averages across read texts and songs are listed in Table 2 together with the values pertaining to stage speech. With regard to the reading of the lyrics, the intrasubject variation was rather small, whereas the variation between the different songs was larger, means 2.5 and 4 dB, respectively. There were clear Leq differences between the songs sung by the same singer, which does not seem surprising because the Leq of singing would depend on the character of the song.

As can be seen in Table 2, the Leq of singing was, on average across subjects, 12.3 dB (standard deviation [SD]: 3.6 dB) higher than that of the reading lyrics. Stage speech showed even higher Leq values, average 15.1 dB (SD: 3.6 dB). For all roles, the Leq differences between reading lyrics and singing were significant, and also between reading lyrics and stage speech ($P < 0.05$). The Leq of stage speech was higher than that of songs for all singers except *Colorful face* 2 and *Old man* 1. However, only *Young woman* and *Young man* roles showed significant differences between singing and stage speech ($P < 0.05$). Thus, the Leq values of singing and stage speech were similar, but both were significantly higher than that of the reading of the lyrics. Also, the variation among singers was greater in singing than in stage speech.

The Leq values for the two performers of the same role varied in many cases. With regard to reading lyrics, the Leq values were significantly different between the two singers of the *Young girl* and of the *Young woman* roles, and with respect to singing the two *Old man role* singers showed significant differences ($P < 0.05$). The female roles who spoke louder in reading the lyrics also recited the stage speech and sang louder. For male roles, in contrast, the Leq of stage speech and singing had little to do with the Leq of reading lyrics.

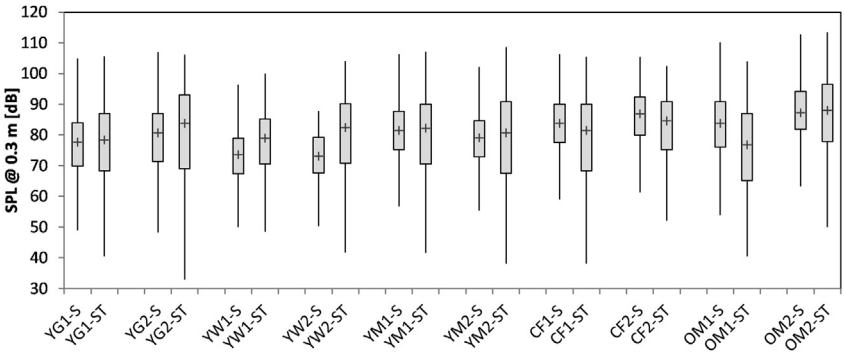


FIGURE 3. Box plot of SPL at 0.3 m. Crosses inside boxes represent the medians. The box represents the value between the first and third quartile locations. The whiskers represent adjacent values. The horizontal axis labels are acronyms of the singers, Y—young, G—girl, W—woman, M—man, CF—colorful face, O—old, S—representing singing, and ST—stage speech.

Comparing the five roles, there were some Leq differences (Figure 2). As a whole, the Leq of male roles were higher than those of the female roles. In both singing and stage speech, *Colorful face* and *Old man* showed the highest values and *Young woman* the lowest, and the differences between roles were much larger in singing. The Leq of most roles differed significantly for singing ($P < 0.05$). Only *Young girl* and *Young man* roles did not show significant difference in singing. With regard to the age of the characters, the younger females and older males had higher Leq.

The mean values of SPL were about 10 dB lower than the Leq values (Table 2). This is not surprising, given the influence of soft phonation and pauses on the SPL. The mean SPL will drop considerably if the recorded signal contains long soft or silent sections, whereas under the same conditions the Leq will remain similar. The reason is that, unlike the SPL average, the Leq is calculated on the basis of linear sound pressure. Therefore, the narrow distribution of SPL values will decrease the difference between the SPL average and the Leq. The SPL of singing had a more concentrated distribution compared with that of stage speech (Figure 3). With respect to the roles, singers of the same role had similar distributions of SPL, whereas singers of different roles showed differing distribution in singing but not in stage speech.

The SPL differences between adjacent voiced segments reflect the time domain variation of loudness. As can be seen in Figure 4, this difference was significantly larger for stage speech than for singing ($P < 0.05$), indicating that the variation

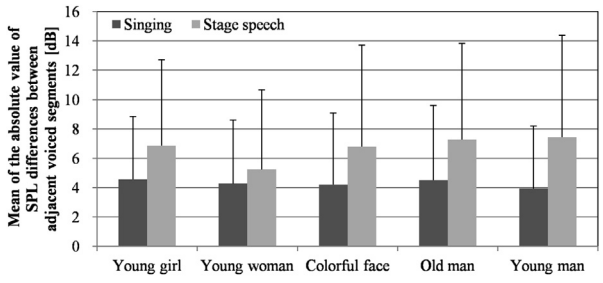


FIGURE 4. Means of the absolute values of SPL differences between adjacent voiced segments. The bars represent +one standard deviation.

was greater in stage speech. For most roles, the variation of SPL with time differed significantly between roles in both singing and stage speech ($P < 0.05$; Table 3).

The average F_0s, MF0, of the different roles are shown in Figure 5. As expected, female roles showed higher means than male roles. For all roles, MF0 was lowest in reading and highest in stage speech. The differences were significant ($P < 0.05$), except for singing and stage speech of the *Old man* role singers. The MF0 for the different roles showed significant differences for singing ($P < 0.05$), although MF0 for *Young girl* and *Young woman* roles were similar. For the male characters, the younger roles used higher MF0. The MF0 differences between singing and stage speech were much larger in the female than in the male roles.

The means and SD of F_0 for each singer are listed in Figure 6. Female singers showed a wider F_0 distribution for singing than for stage speech, whereas the opposite was true for male singers. Between singers, the SD_{F_0} showed great variation for stage speech but small variation for singing, the latter reflecting mainly compositional characteristics. The singers of the same role showed similar SD_{F_0} for the same task except for the *Old man* role. The female singers showed smaller SD_{F_0} than male singers when reciting stage speech.

Comparing the data shown in Figures 2 and 5, interesting relationships between mean Leq and MF0 can be observed for the different roles. Within roles, there was a positive correlation, implying that the MF0 was high when the singers produced a high Leq. In contrast, the roles performed with higher Leq tended to be sung at a lower MF0. It seems likely that these relationships between Leq and MF0 belong to the characteristics of the different roles.

DISCUSSION

Our analyses comprised no more than two singers for each of the five roles. On the other hand, all singers were professional and earned their livelihood from singing, suggesting that they had well-established singing skills and well-controlled voices. Second, four songs with different emotions were enough for reflecting the variations of songs in the same role. The songs of Kunqu Opera could be divided into two groups, the *South song* and the *North song*. Typically, *South song*s are smooth,

TABLE 3.
The *P* values According to a Mann-Whitney *U* test (Significance Level 0.05) of the Difference in Time Variation of SPL Between the Roles in Singing and in Stage Speech

	Singing				Stage Speech			
Role	Young Woman	Colorful Face	Old Man	Young Man	Young Woman	Colorful Face	Old Man	Young Man
Young girl	0.000	0.000	0.000	0.000	0.000	0.014	0.904^{ns}	0.927^{ns}
Young woman		0.000	0.160^{ns}	0.001		0.002	0.000	0.000
Colorful face			0.010	0.105^{ns}			0.025	0.020
Old man				0.184^{ns}				0.831^{ns}

Abbreviation: ns, not significant.

whereas *North song*s are more excited. All roles include both *South song*s and *North song*s except *Colorful face*; at present, most songs of that role are of the *North song* type. Thus, including several song samples for each role should have enhanced the credibility of the results. However, there was only one sample of stage speech for each singer, which might have limited the representatively of the findings. The variance of stage speech should be considered in the future.

As was shown in Figure 1, some singers' Leq was higher when they were reading the text of stage speech than when reading the lyrics of the songs, possibly because they were influenced by the speaking style of stage speech. In fact, they read the texts of stage speech more emotionally than the lyrics. When reading the lyrics of the songs, they perhaps adopted their voice habits of conversational speech.

Considering the age, dialect, and the recording place, some point should be mentioned. In all roles, singer number 1 was older than the singer number 2, particularly for *Young woman* and *Young man*. The younger singers showed higher Leq than the older singers, especially in stage speech. Another factor is the dialect. *Young girl* 2 and *Old man* 2 both came from North China and their dialect was northern mandarin. The other singers came from South China, and their dialect was the Wu dialect, which sounds gentler than Mandarin. The style of north Kunqu Opera is bold, whereas the style of south Kunqu Opera was gentle. Although the singers performed similar plays and used the same language when they were acting, they were probably influenced by their cultures and dialects. Also, it cannot be excluded that the different recording conditions between the north and south groups had an effect. *Young girl* 2 and *Old man* 2 were recorded in a sound-treated booth with an abnormally low reverberation level, which may have caused them to increase vocal loudness. On the other hand, the Leq difference was small between the lyrics reading of *Old man* 2 and *Old man* 1, who were recorded in different rooms.

Previous research has found a strong positive correlation between Leq and MF0 in speech produced at different loudnesses.[6] The correlations differed between roles in singing. The Leq and MF0 were only significantly correlated for the *Old man* and *Young man* roles ($P < 0.05$, $R^2 > 0.9$). The songs sung by each of the singers differed in emotional color, and this is likely to weaken the correlation. In speech, none of the correlations were significant (significance level is 0.05). However, the ranges were narrow both in Leq and in MF0.

In the tradition of Kunqu Opera, the *Young girl* and the *Young woman* roles are performed in falsetto register, whereas the *Colorful face* and the *Old man* roles use modal register. Mean Leq and MF0 were intermediate for *Young man* role, and this role uses modal voice in the lower pitch range and falsetto in the higher. The relationships between vocal register and Leq and MF0 in Kunqu Opera would be worthwhile to study more in detail in the future.

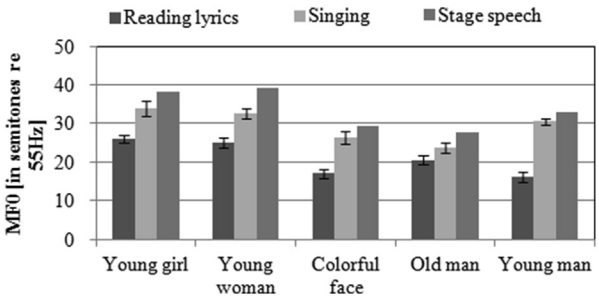


FIGURE 5. The MF0 (in semitones re 55 Hz), averaged across subjects for the indicated conditions. The bars represent ±one standard deviation.

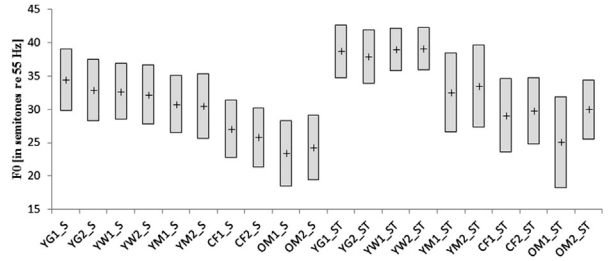


FIGURE 6. Mean and standard deviation of the distribution of F_0 (in semitones re 55 Hz) observed in the 10 Kunqu Opera singers' singing of three or four songs and their stage speech. Crosses inside boxes represent the means. The box represents the positive and negative standard deviation. Y—young, G—girl, W—woman, M—man, CF—colorful face, O—old, S—representing singing, and ST—stage speech.

CONCLUSION

This study explored the differences in Leq, SPL, and F_0 among tasks, singers, conditions, and roles. The interrole difference was larger than the intrarole difference. The singers of the same role showed a similar F_0 concentration, not only for singing but also for stage speech. The variation of SPL with time differed between most roles in both singing and stage speech.

On average, the Leq of stage speech and singing were 15 and 12 dB higher than conversational speech as documented in the singers' reading of lyrics. The Leq of stage speech were higher than singing for all the singers of *Young girl*, *Young woman*, and *Young man* roles. The between-role Leq differences were smaller in stage speech than in singing. In singing as compared with stage speech, the SPL distribution was more concentrated and the time domain variation of SPL was smaller.

The mean Leq and MF0 varied systematically with the sex and age of the singer. Male roles had higher mean Leq and lower MF0 than female roles. The F_0 distribution of singing, expressed in semitones, was wider than that of stage speech for female singers and narrower for male singers. There was not much difference in F_0 concentration between singers while singing. The female singers showed smaller SD_{F_0} than male singers in stage speech. With regard to the ages of the characters, younger female personages showed higher Leq, whereas older male personages had higher Leq. The roles performed with higher Leq tended to be sung at a lower MF0.

Acknowledgments

The authors would like to thank the voice experts for their gentle participation in this investigation. This research was funded by the National Social Sciences Foundation of China and China Scholarship Council, grant numbers were 10&ZD125 and 201206010134, respectively. It was carried out during the first author's stay at the Department for Speech, Music and Hearing at KTH Stockholm, Sweden.

REFERENCES

1. Wu X. *Dictionary of Chinese Kunqu Opera*. Nanjing: Nanjing University Press; 2002.
2. Fant G. *Speech Acoustics and Phonetics*. Norwell, MA: Kluwer Academic Publishers; 2004.
3. Kong J. *On Language Phonation*. Beijing: Central Nationalities University Press; 2001.
4. Sundberg J. *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press; 1987.
5. Sundberg J, Gu L, Huang Q, Huang P. Acoustical study of classical Peking Opera singing. *J Voice*. 2012;26:137–143.
6. Gramming P, Sundberg J, Ternström S, Leanderson R, Perkins WH. Relationship between changes in voice pitch and loudness. *J Voice*. 1988;2:118–126.

Logopedics Phoniatrics Vocology, 2014; 39: 72–80

informa
healthcare

ORIGINAL ARTICLE

Long-term-average spectrum characteristics of Kunqu Opera singers' speaking, singing and stage speech

LI DONG[1], JIANGPING KONG[1] & JOHAN SUNDBERG[2]

[1]*Peking University, Department of Chinese Language and Literature, Beijing, China, and* [2]*Department of Speech Music Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden*

Abstract

Long-term-average spectrum (LTAS) characteristics were analyzed for ten Kunqu Opera singers, two in each of five roles. Each singer performed singing, stage speech, and conversational speech. Differences between the roles and between their performances of these three conditions are examined. After compensating for Leq difference LTAS characteristics still differ between the roles but are similar for the three conditions, especially for Colorful face (CF) and Old man roles, and especially between reading and singing. The curves show no evidence of a singer's formant cluster peak, but the CF role demonstrates a speaker's formant peak near 3 kHz. The LTAS characteristics deviate markedly from non-singers' standard conversational speech as well as from those of Western opera singing.

Key words: *Kunqu Opera, LTAS, role, singer's formant cluster, speaker's formant*

Introduction

The voice timbres of Kunqu Opera singers are supposed to mirror the ages, characters, and identities of the respective roles, which have been described elsewhere (1). In our previous investigations, Kunqu Opera singers' stage speech, singing, and conversational speech were found to differ with regard to equivalent sound level (Leq) and fundamental frequency (F0) (1). These parameters were somewhat higher for stage speech than for singing, and both were significantly higher than for conversational speech. They also differed between roles. However, Leq and F0 differences would not be enough for describing all relevant acoustic characteristics of the specific voices of the different Kunqu Opera roles. Also spectrum differences would be important. Already Leq differences are typically accompanied by frequency-dependent effects on the voice source spectrum (2–5). Furthermore, at high F0 singers may vary the formant frequencies and the distances between them (6–8). This affects the levels of formant peaks in the spectrum and hence also the voice timbre. Therefore, an exhaustive description of the vocal style of Kunqu Opera singing needs to analyze also spectrum characteristics.

The long-term-average spectrum (LTAS) is an effective tool for voice analysis. It represents the overall spectral characteristics of a voice and typically stabilizes after 30–40 seconds of running speech (9–14) and singing (15–18). The LTAS contour reflects both the voice source and the vocal tract resonance characteristics. In singing as well as in speech an LTAS typically shows a peak near 0.5 kHz. The reason is that F1 is frequently located in this range. Classically trained Western singers, such as bass, baritone, and tenor singers, typically display another pronounced peak in the high-frequency part of an LTAS, between about 2.5 and 3.3 kHz (8,15). This peak has been referred to as the singer's formant cluster and has been explained as the result of clustering formants 3, 4, and 5 (15). For professional voice users, such as actors and country singers, a prominent peak often occurs at a slightly higher frequency, near 3.5 kHz. It has been called the speaker's formant (12–14,19). It has been explained as the result of the closeness of F3 and F4 (14).

Correspondence: Professor Johan Sundberg Department of Speech Music Hearing, School of Computer Science and Communication, KTH, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden. Fax: + 46 (8) 790 78 54. E-mail: pjohan@speech.kth.se

(Received 29 April 2013; accepted 2 September 2013)

ISSN 1401-5439 print/ISSN 1651-2022 online © 2014 Informa UK, Ltd.
DOI: 10.3109/14015439.2013.841752

Also the singer's formant has been explained as the consequence of a reduction of the frequency distance between higher formants. Acoustic theory of voice production (7) predicts that the levels of two formants generally increase by 6 dB each if the distance between them is halved. Likewise, vowels with a high first formant, such as /a/, or a high second formant, such as /i/, have strong singer's formants, and vice versa. Formant frequencies are determined by vocal tract shape. For example, the singer's formant is highly dependent on the physiological configuration of the vocal tract, particularly the shape of the larynx tube and the area ratio between the larynx tube opening and the pharyngeal tube at the level of this opening (15).

The amplitudes and frequencies of the LTAS peaks just mentioned are influenced also by voice source. The amplitudes of the voice source partials depend mainly on the maximum flow declination rate which occurs during the closing of the glottis (7). If the rate is slow, the amplitude of the partials in high frequency will be low, and vice versa. The type of closure also influences the amplitude of the partials. For example, in 'breathy' phonation, in which the vocal folds fail to close the glottis completely, the amplitudes of the upper partials are decreased, which reduces the prominence of the singer's formant.

In this investigation, voice characteristics of Kunqu Opera performers of five traditional roles, Young girl (YG), Young woman (YW), Young man (YM), Colorful face (CF), and Old man (OM), are analyzed in terms of LTAS. The aim was to investigate 1) whether the LTAS of Kunqu Opera singers are similar in conversational speech, singing, and stage speech; and 2) whether the Kunqu Opera singers demonstrate a singer's formant or speaker's formant LTAS peak. Comparisons of LTAS of classically trained Western singers and normal speakers and those of Kunqu Opera singers are made to illustrate the differences.

Method

Four female and six male professional performers of Kunqu Opera used in our previous study (1) were subjects also for the experiment (Table I). The singers were told to sing three to four songs just as on stage. The total duration of the songs, which differed in emotional color, was 6–18 minutes. The singers also recited a section of stage speech, which lasted for 1–3 minutes. In addition, all singers read, in modal voice and in the style of conversational speech, the lyrics of the recorded songs. This reading, henceforth referred to as reading, took between 2 and 3.5 minutes. The language differed from Mandarin Chinese but was identical with what they used in their roles on stage, which actually corresponds to ancient Chinese in Ming Dynasty.

Table I. Information of ten performers.

Roles	Singer	Age (years)	Professional experiences (years)	Gender
Young girl	1	45	25	Female
	2	41	21	Female
Young woman	1	47	27	Female
	2	27	8	Female
Colorful face	1	27	9	Male
	2	25	7	Male
Old man	1	46	27	Male
	2	44	25	Male
Young man	1	45	25	Male
	2	27	8	Male

YG singer 2 and OM singer 2, who work at the Northern Kunqu Opera Theater, were recorded in an anechoic room, about $3.6 \times 2.6 \times 2.2$ m. The other singers, who are performers of the Kunqu Opera Theater of Jiangsu Province, had to be recorded in a quiet living room, about $3.5 \times 5 \times 3$ m; the background noise was 35 dB(A), and the reverberation time was about 0.3 s. Although the room acoustic was quite different from a typical Kunqu Opera stage, none of these highly experienced singers complained about difficulties to control their voices. A Sony Electret Condenser Microphone, placed off axis at a measured distance that varied between 15 and 21 cm for the different singers, was used to record the audio signals (critical distance of the room was about 75 cm). The signals were digitized on 16 bits at a sampling frequency of 20 kHz and recorded on single-channel wav files into ML880 PowerLab system. Sound pressure level (SPL) calibration was carried out by recording a 1 kHz tone, the SPL of which was measured at the recording microphone by means of a TES-52 Sound Level Meter (TES Electrical Electronic Corp., Taiwan, ROC) and then announced in the recording file together with the respective microphone distance. All sound level data were normalized to 30 cm.

The LTAS analysis of the wav files was accomplished using the WaveSurfer software (1.8.8p3). The FFT window length was set to 128-point, the bandwidths of the analysis filters to 303 Hz, and the frequency range to 0–10 kHz. After eliminating pauses longer than 10 ms from the recordings, LTAS were computed for each singer's entire recording in each condition. The recordings of singing were long, and those of reading lyrics and of stage speech was rather short (1–3 minutes). Therefore, for each singer, LTAS was computed for each 40-second section of the recordings of singing so as to allow analysis of

variation. Since the main sound energy appeared in the frequency range 0–5 kHz, the analysis was limited to this range. The curves for reading and stage speech were adjusted so as to compensate for Leq differences. This compensation was realized by multiplying the level values by the LTAS mean gain factors reported in previous research for different frequency bands (1,5). The gain factor increases with frequency in the low-frequency range, keeps stable in the middle range (from 1.3 to 3 kHz) at 1.4 for male singers and at 1.6 for female singers, and decreases in the high-frequency range. For instance, to compensate a difference in Leq of 10 dB between two voice samples of a male singer, the LTAS level of the voice with lower Leq is increased by $10 \times 1.0 = 10$ dB in the 500 Hz frequency band, while the LTAS level in the 3000 Hz frequency band is increased by $10 \times 1.4 = 14$ dB. To obtain a quantitative measure of LTAS similarity, correlations (linear regression) were calculated between pairs of LTAS curves, using SPSS 18.

F0 was extracted using the WaveSurfer software. The extraction method was ESPS (Entropic Speech Processing System), using the algorithm of ACF (Auto Correlate Function); F0 was limited from 60 to 900 Hz; the analysis window length was 0.0075 s; and the frame interval was 0.01 s. The description statistics were accomplished using SPSS.

Results

After the LTAS had been compensated for Leq differences (1,5), the differences between them for the three conditions were substantially diminished, especially for the CF and OM roles (Figure 1). The LTAS curves for the three different conditions differ in a similar way for the two singers of the same role. For the female singers stage speech showed considerably less energy in the low-frequency range, up to about 0.6 kHz. This would depend on their elevated F0 range. On the other hand, for the CF and OM roles, the LTAS curves of all three conditions are quite alike. For YG, YW, and YM roles, the maximum peak in stage speech is located near or somewhat higher than in singing, and the peak is also narrower. The stage speech curve exhibits several peaks. Their center frequencies are close to harmonic. For example, for YG1, the center frequencies of the second, third, fourth, and fifth peaks of the stage speech appear at 1.8, 2.4, 3.1, and 4.2 kHz, i.e. close to 3, 4, 5, and 7 times 600 Hz.

Pairwise LTAS comparisons of conditions are listed in Table II in terms of the determination coefficients. After the compensation for Leq differences, the data show higher correlations than the original data, especially between reading and singing and between reading and stage speech. This suggests that Leq variation was an important reason for the differences between three conditions. With regard to the correlations between the compensated data, all of them were significant, and for most singers reading lyrics and stage speech showed the lowest similarity; the spectrum level of reading lyrics and singing were highly correlated ($R^2 > 0.9$ in 8 of the 10 singers). Thus the LTAS curves of Kunqu Opera singers' singing and reading show high similarity.

The voice timbres differ between roles (1), and LTAS curves can reflect the voice timbre. Thus, it also seems relevant to examine how the LTAS differ between the roles. Although in the present study no more than two representatives of each role were analyzed, the average LTAS for a role seems worthwhile to study. It should be borne in mind that our subjects were professional representatives of the respective roles and hence their voice must contain typical characteristics of that role. Furthermore, such an average LTAS will reduce the salience of individual characteristics. For example, of the two OM singers, one showed a marked peak near 3000 Hz, while the other did not, so this peak is rather weak in the average LTAS. On the other hand a marked peak appeared in this frequency range in both CF singers' LTAS, so it became prominent in the average LTAS, thus suggesting that this may be a typical property of this role.

The left and right panels of Figure 2 show the average LTAS for each of the roles for singing and stage speech. All roles display a main peak between 0.7 and 1.1 kHz; for the CF and OM roles it appears at somewhat higher frequencies than for the other roles, for both singing and stage speech. The curves differ in steepness in the octave above the main peak. In singing it is more than 16 dB/oct for the CF and OM roles and much less for the three young roles, no more than 4 dB/oct for the YW role. In stage speech the spectrum slope in this octave is 8 dB/oct for the YM role, 12 dB/oct for the YG, YW, and CF roles, and 17 dB/oct for the OM role. A second peak can be observed at 3 kHz. It is particularly marked for the CF role and the stage speech of the YM role.

To see the LTAS characteristics of Kunqu Opera singers' singing and stage speech, it is relevant to compare their LTAS with that of standard conversational speech, which has been reported in a previous study (5). Figure 3 shows how the Kunqu Opera singers' LTAS curves deviate from this reference. For both singing and stage speech, the LTAS level around 1 kHz is higher than the reference. This applies to all roles. In the female roles' singing, the LTAS level between 1 and 2 kHz is much stronger than the reference. A marked valley occurred in the vicinity of

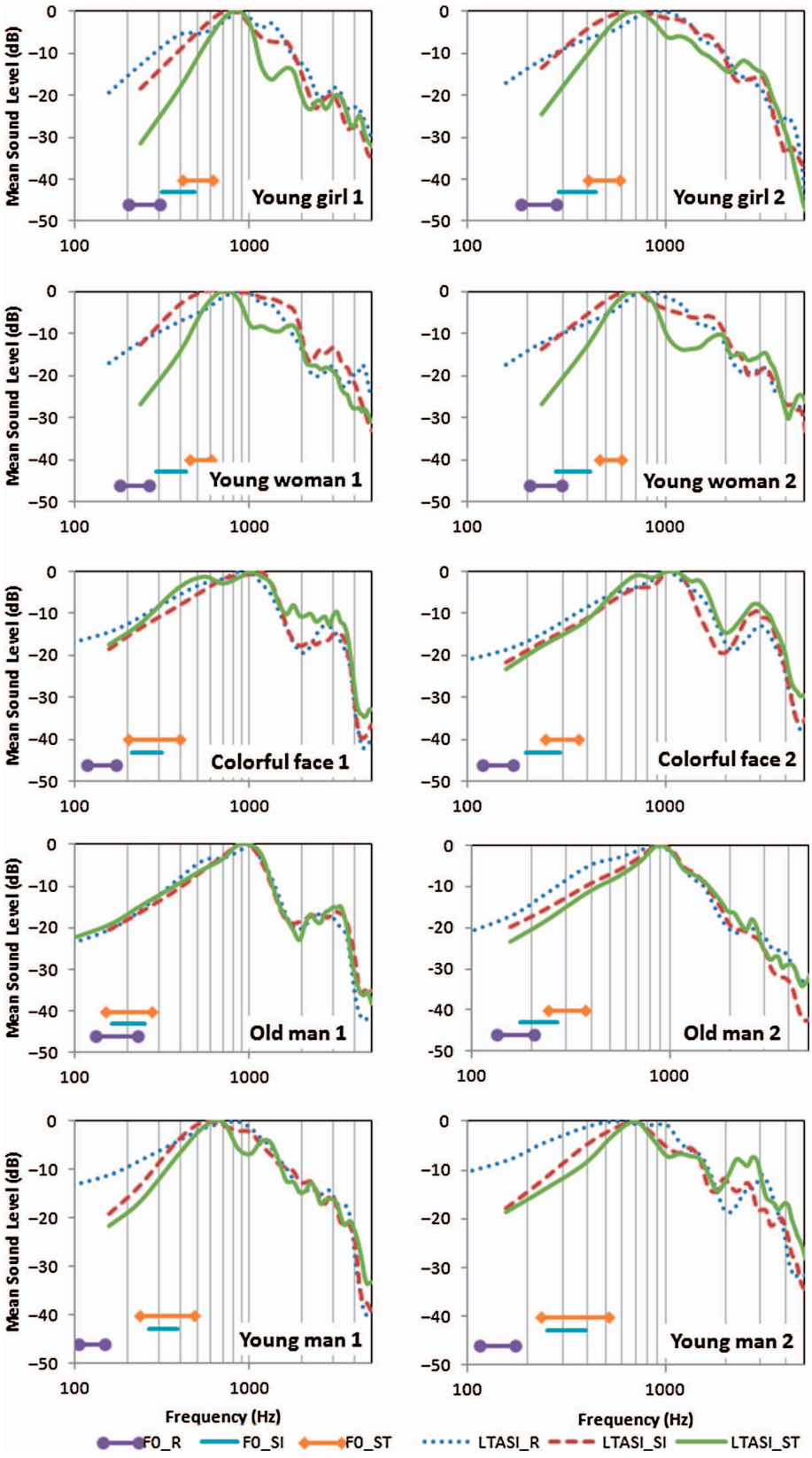


Figure 1. LTAS curves for lyrics reading, singing, and stage speech (R, SI, and ST, respectively). The horizontal lines correspond to the separation of the first and third quartiles of the F0 distribution.

Table II. Coefficients of determiation for the correlations between three conditions for ten singers before and after compensation of the Leq differences (Original and Compensated, respectively) (5). All correlations are significant.

	Reading and singing				Reading and stage speech				Singing and stage speech			
	Original		Compensated		Original		Compensated		Original		Compensated	
Singers	R^2	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2	Slope
Young girl 1	0.78	1.00	0.97	1.16	0.40	0.61	0.76	0.81	0.78	0.76	0.83	0.72
Young girl 2	0.91	1.05	0.93	1.04	0.73	1.01	0.86	1.03	0.84	0.99	0.88	0.96
Young woman 1	0.49	0.84	0.76	1.15	0.32	0.63	0.74	1.01	0.84	0.85	0.87	0.83
Young woman 2	0.92	0.99	0.96	1.10	0.47	0.64	0.73	0.70	0.67	0.74	0.77	0.70
Colorful face 1	0.86	1.03	0.97	0.92	0.77	0.90	0.95	0.82	0.96	0.90	0.96	0.89
Colorful face 2	0.74	1.05	0.93	0.92	0.67	0.90	0.91	0.83	0.94	0.87	0.94	0.88
Old man 1	0.93	0.87	0.97	0.77	0.96	0.92	0.97	0.83	0.97	1.02	0.98	1.07
Old man 2	0.79	1.24	0.96	1.24	0.71	0.96	0.92	0.96	0.95	0.80	0.96	0.78
Young man 1	0.81	1.03	0.98	0.94	0.71	0.81	0.93	0.72	0.96	0.82	0.96	0.77
Young man 2	0.64	0.84	0.87	0.84	0.37	0.54	0.76	0.58	0.86	0.79	0.89	0.70

2 kHz for OM and CF roles. Between 1.5 and 4.5 kHz there are between one and three peaks for most singers. The CF shows a positive deviation from the reference between 2.5 and 4.5 kHz, and for the YM role a peak, particularly marked for stage speech, can be seen around 4 kHz. Less clear peaks can be observed near 3 kHz for the YG, YW, and YM roles.

Figure 4 compares the LTAS curves of different Kunqu Opera singers with those of comparable Western opera singers (8), using similarity in pitch range as criterion and the akronyms SI for singing and ST for stage speech: alto for YGSI, YGST, YWSI, and YWST; baritone for CFSI, CFST, OMSI, OMST, and YMSI; and tenor for YMST. In both singing and stage speech, the main peak of Kunqu Opera singers' LTAS curves appears at higher frequency than for the Western opera singers, and the LTAS level below the main peak frequency is clearly lower. However, this may be because the LTAS curves of the Western opera singers were derived from commercial recordings in which the singers were accompanied by an orchestra. In the female roles' singing, the LTAS level between 1 and 2 kHz is much stronger than in the case of Western altos. The female Kunqu Opera singers and the Western altos both display an LTAS peak near 3 kHz, which is somewhat higher in frequency and less marked in the Kunqu Opera singer voices. The LTAS curves of the CF role show a peak similar to that of Western baritone singer's formant cluster, even though its center frequency is higher. Its level is comparable for stage speech but clearly weaker in singing. The LTAS curves of OM role's singing and stage speech and YM role's singing show no obvious peak in this frequency range. In YM role's stage speech, two small peaks present between 2 and 3 kHz, while Western tenor singer's formant cluster appears at higher frequency and is more marked.

The standard deviations associated with the LTAS curves for the ten subjects' singing are shown in Figure 5. This standard deviation (SD_{LTAS}) varies considerably between roles and singers. It is particularly wide for YW2 and particularly narrow for the OM and CF roles. For the female roles, the SD_{LTAS} between 1 and 2.5 kHz is similar to the difference

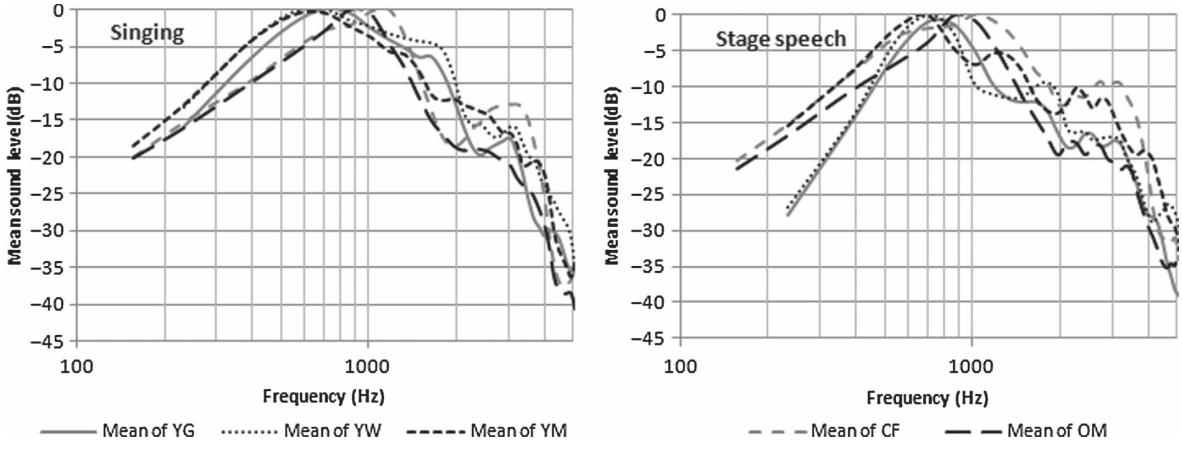


Figure 2. Mean LTAS of the two singers of the indicated roles.

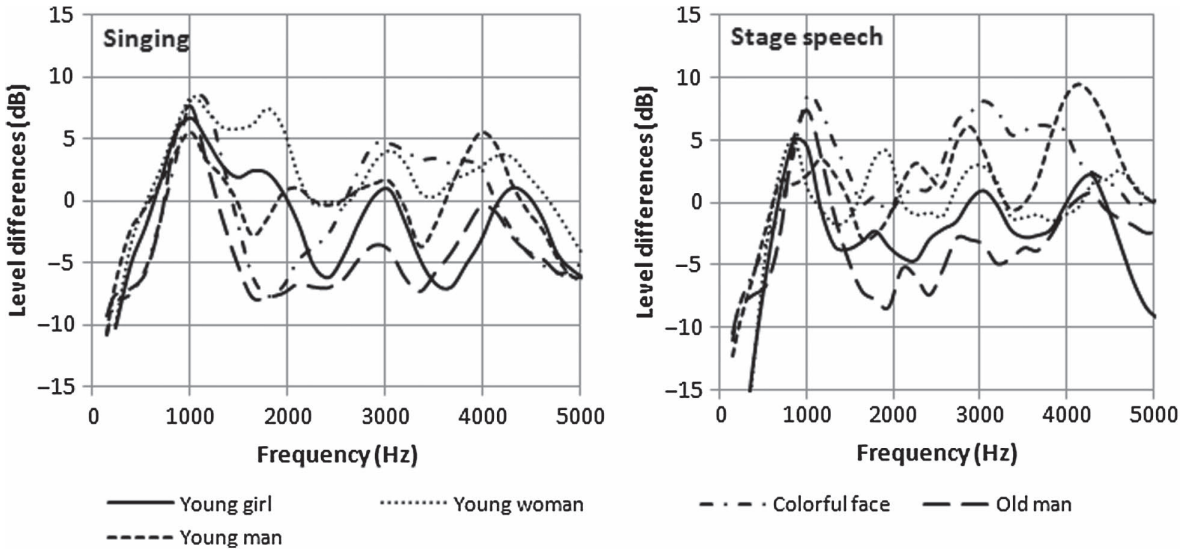


Figure 3. Differences between the LTAS of singing, stage speech of Kunqu Opera singers, and standard conversational speech (5). The LTAS of standard conversational speech was compensated for the Leq difference for the different singers.

between their LTAS for singing and the LTAS for standard conversational speech (Figure 3). This indicates that for these voices the LTAS curves vary considerably depending on what segment is chosen for analysis.

The SD_{LTAS} in the frequency range of the singer's formant cluster is relevant for determining whether or not a voice possesses a singer's formant cluster; a low SD_{LTAS} would imply that the spectrum level in the corresponding frequency range shows a small variation. In the case of the CF role, particularly in the case of singer 2, the SD_{LTAS} is quite narrow in the frequency range of the singer's formant cluster. This means that these singers tended to produce strong partials in this frequency region. The three young roles, especially YW2 and YM2, show large values of SD_{LTAS} near 3 kHz.

Traditional Kunqu Opera singing is performed without sound amplification and typically accompanied by a solo Kun bamboo flute. The singer's formant cluster in Western operatic singing seems to have been developed in response to the sound quality of Western orchestra, enhancing partials in a frequency range where the competition with the accompaniment is moderate. It is then relevant to ask if a similar relationship exists between the timbral quality in Kunqu Opera singing and the Kun bamboo flute. Figure 6 shows LTAS curves, measured over several minutes of playing of the Kun bamboo flute for two types of music, 'south song' and 'north song'. Both demonstrate three peaks below 5 kHz. The main peak appears in the low-frequency range, near 700 and 1200 Hz. Both show secondary peaks between 2 and 3 kHz and between 4 and 5 kHz.

Discussion

LTAS curves of most Kunqu Opera singers show one or more peaks in the high-frequency range. Clear peaks in an LTAS curve may reflect either of three conditions or combinations of them: 1) stable formants frequencies; 2) narrow formant bandwidths; and 3) partials in the corresponding frequency region. Since the frequencies of the higher formants are rather constant, the first condition is mostly met. Regarding the second condition, a long closed phase will make the bandwidths narrow, and, with respect to the third, a high F0 implies wide separation of spectrum partials, so that the peaks at high frequencies may reflect both harmonic partials and formants. Conversely, an LTAS peak will be a sign of a stable formant when the F0 average is low or when the variation of F0 is great. Compared with the CF role, the YG, YW, and YM roles, who all sing in a high F0 range, showed lower spectrum level at high frequencies. This may be a combined effect of formants and partials.

A tendency to cluster two formants will result in a peak at the center frequency of the cluster surrounded by valleys. The singer's formant is produced by clustering F3, F4, and F5, and the center frequency of the peak appears between 2.5 and 3.3 kHz, depending on the voice classification. According to Bele (14), the speaker's formant is produced by lowering of F4 such that it approaches F3, and the center frequency is between 3.15 and 3.7 kHz. Both CF singers and one of the OM role singers show peaks near 3 kHz surrounded by valleys, while the other singers do not. The peak has wider bandwidth and lower level than the singer's formant in Western

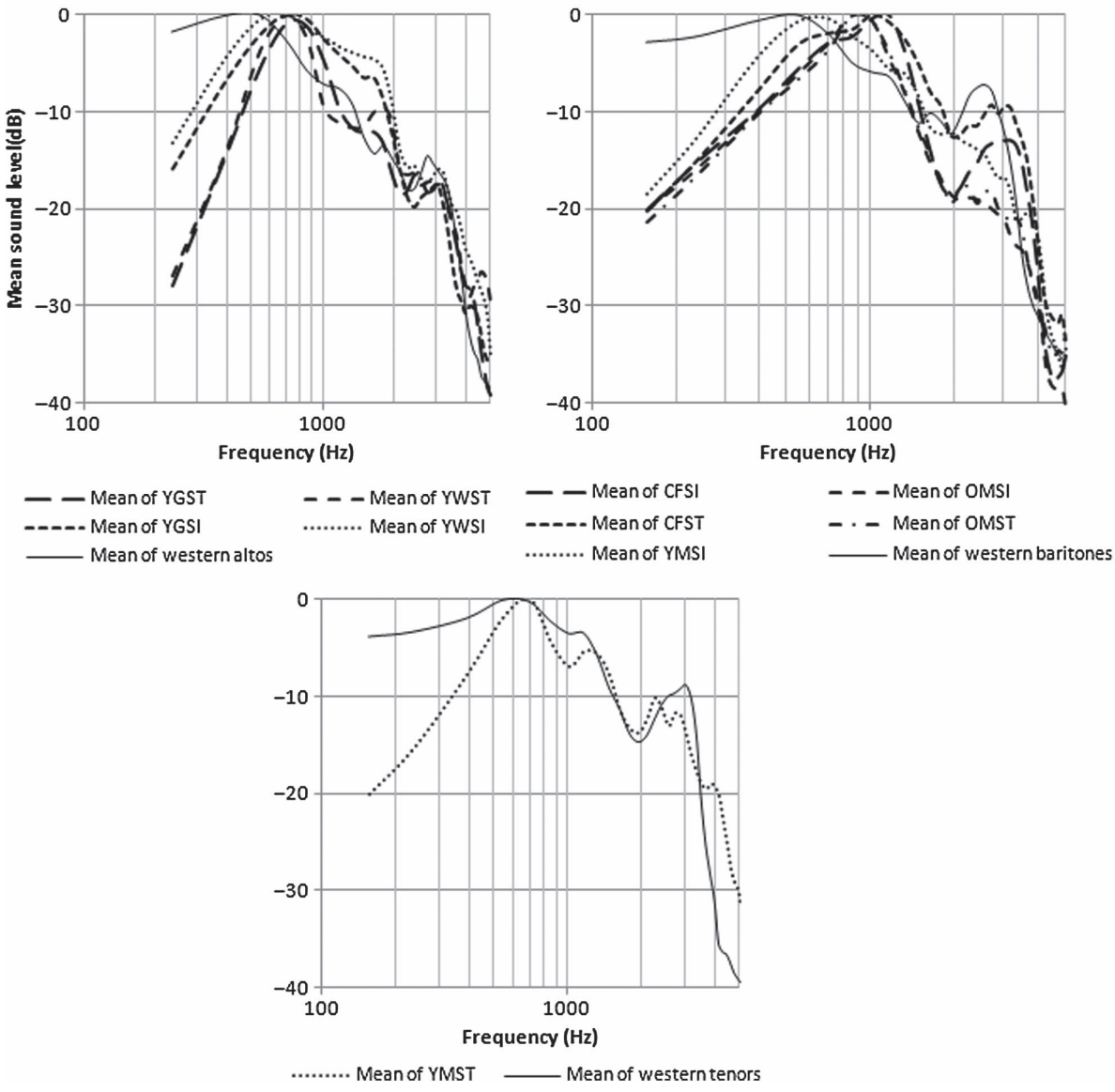


Figure 4. LTAS of singing, stage speech of YG, YW, CF, OM, and YM roles and singing of Western opera singers. SI: singing, ST: stage speech.

baritones' LTAS. Thus, it is not comparable to the singer's formant cluster but similar to the speaker's formant.

Formant frequencies affect the shape of the LTAS curve, as mentioned. For Kunqu Opera singers, F2 in low vowels, e.g./a/ and/ɑ/, produce a strong spectrum peak which tends to extend the main peak up to 2 kHz. By contrast, the center frequency of the main LTAS peak in previously published studies of conversational speech and of Western opera singing is typically located in a lower frequency range, about 500 Hz. In Kunqu Opera singers' front vowels, F2 in singing is up to 2.5 kHz and close to F3. This will raise the level of the second marked LTAS peak and form a valley between the main peak and the second peak, as in the case of the CF singers (Figure 1).

There may be several reasons for the absence of the singer's formant cluster in Kunqu Opera: 1) The presence of a singer's formant cluster reduces the differences between vowels, and text intelligibility may be particularly important in Kunqu Opera; and 2) the singer's formant cluster boosts the sound of the singer's voice so it can be heard over an accompanying orchestra. However, Kun bamboo flute, which is the most common accompaniment for Kunqu Opera, shows a peak in the frequency range of the singer's formant cluster (Figure 6). Thus, it has an LTAS curve totally different from that of a Western opera orchestra, which shows a rather low level around 3 kHz. Hence, a speaker's formant would be more effective than the singer's formant cluster to boost the singer's voice. For female roles,

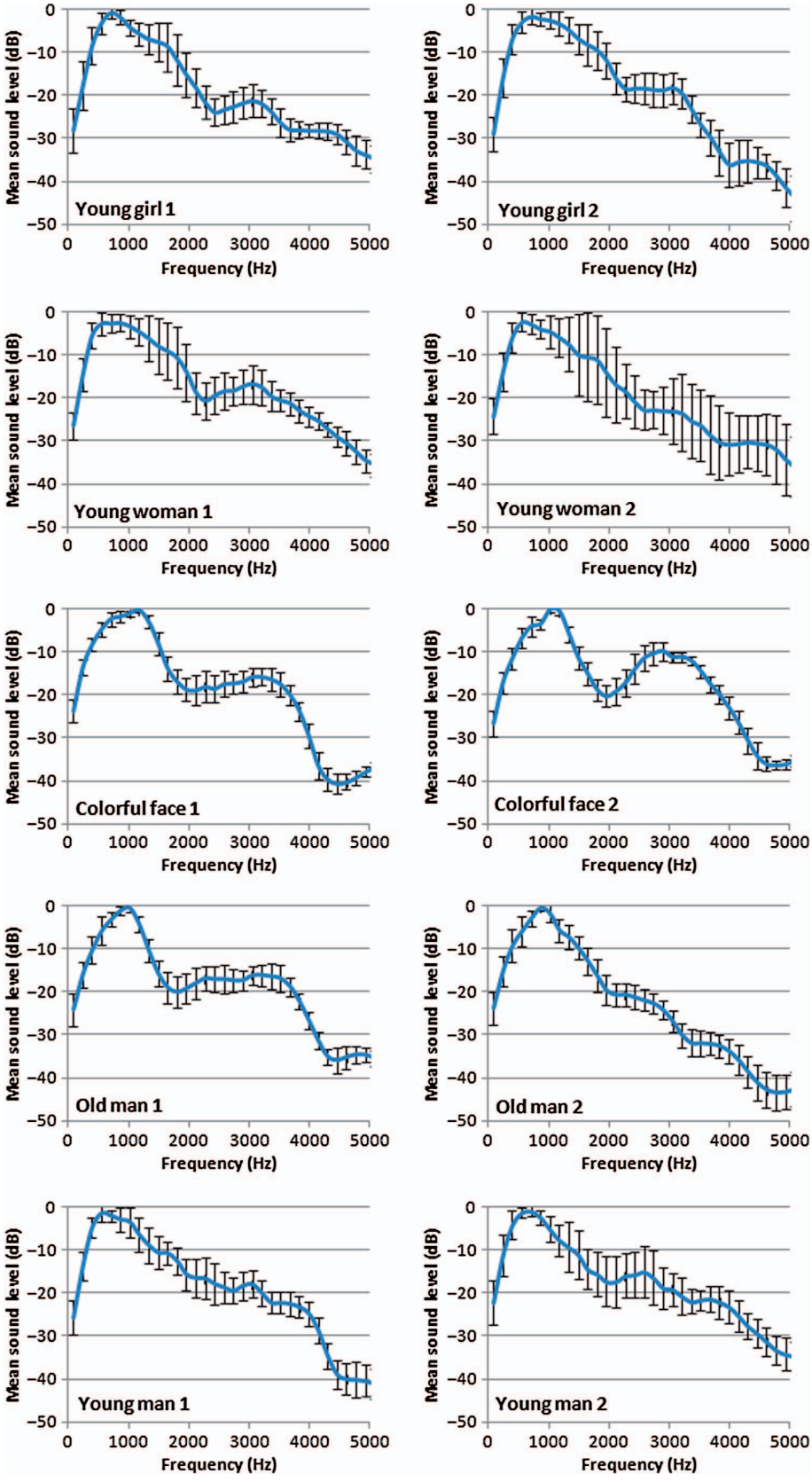


Figure 5. LTAS curves and standard deviations of the different singers' singing.

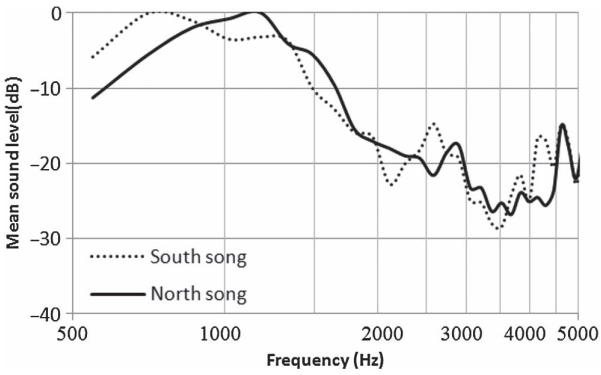


Figure 6. LTAS of Kun bamboo flute.

which show lower Leq than the male roles, the LTAS levels between 1.5 and 2 kHz are higher than that of the bamboo flute. This may help the female roles to cut through the sound of bamboo flute.

Conclusion

LTAS characteristics of Kunqu Opera performers of the roles YG, YW, YM, CF, and OM were found to differ between the roles. The CF role demonstrated a speaker's formant peak in their LTAS curves. In singing, the LTAS curves for the performers of the three young roles showed a great variability near 3 kHz between consecutive parts of the song, as reflected in terms of large values of SD_{LTAS}. This implies a great variation of voice timbre and/or vocal loudness. None of the roles showed a singer's formant cluster. For all roles the main LTAS peak showed wider bandwidth and appeared at a higher frequency in singing and stage speech than in non-singers' standard conversational speech. The singers' reading differed considerably from their singing and stage speech, but the substantially lower Leq seemed to be an important reason for this difference. Thus, after compensating the LTAS curves for this difference, the characteristics of reading, singing, and stage speech became strikingly similar, particularly for the CF and OM roles. For all roles the similarity was particularly high between reading and singing.

Acknowledgements

The authors would like to thank the voice experts for their gentle participation. Special thanks should go to Oh Hanna for her help with recording. The work on this article was carried out during the first author's stay at the Department for Speech, Music and Hearing at KTH.

Declaration of interest: This research was funded by the National Social Sciences Foundation of China and China Scholarship Council, grant numbers 10&ZD125 and 201206010134, respectively.

References

1. Dong L, Sundberg J, Kong J. Loudness and pitch of Kunqu Opera. J Voice 2014;28:14–9.
2. Cleveland T, Sundberg J. Acoustic analysis of three male voices of different quality. In: Askenfelt A, Felicetti S, Jansson E, Sundberg J, editors. Proceedings of the Stockholm Music Acoustics Conference (SMAC 83): I. Stockholm: Royal Swedish Academy of Music, Publ No. 46:1; 1985. p. 143–56.
3. Bloothooft G, Plomp R. The sound level of the singer's formant in professional singing. J Acoust Soc Am. 1986; 79:2028–33.
4. Hollien H. The puzzle of the singer's formant. In: Bless DM, Abbs JH, editors. Vocal fold physiology. Contemporary research and clinical issues. San Diego: College-Hill; 1983. p. 368–78.
5. Nordenberg M, Sundberg J. Effect on LTAS on vocal loudness variation. Logoped Phoniatr Vocol. 2004; 29:183–91.
6. Sundberg J. Formant technique in a professional female singer. Acustica. 1975;32:89–96.
7. Fant G. Acoustic theory of speech production. Haag: Mouton; 1960.
8. Sundberg J. Level and center frequency of the singer's formant. J Voice. 2001;15:176–86.
9. Kitzing P. LTAS criteria pertinent to the measurement of voice quality. J Phonetics. 1986;14:477–82.
10. Löfqvist A, Mandersson B. Long-time average spectrum of speech and voice analysis. Folia Phoniatr. 1987; 39:221–9.
11. Novak A, Dlouha O, Capkkova B, Vohradnik M. Voice fatigue after theater performance in actors. Folia Phoniatr. 1991; 43:74–8.
12. Leino T. Long-term average spectrum study on speaking voice quality in male actors. In: Friberg A, Iwarsson J, Jansson E, Sundberg J, editors. Proceedings of the Stockholm Music Acoustics Conference (SMAC 93). Stockholm: Royal Swedish Academy of Music, Publ No. 79; 1993. p. 206–10.
13. Nawka T, Anders LC, Cebulla M, Zurakowski D. The speaker's formant in male voices. J Voice. 1997;11:422–8.
14. Bele I. The speaker's formant. J Voice. 2006;20:555–78.
15. Sundberg J. Articulatory interpretation of the 'singing formant'. J Acoust Soc Am. 1974;55:838–44.
16. Cleveland T. Acoustic properties of voice timbre types and their influence on voice classification. J Acoust Soc Am. 1977;61:1622–9.
17. Dmitriev L, Kiselev A. Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities. Folia Phoniatr. 1979; 31:238–41.
18. Sundberg J, Gu L, Huang Q, Huang P. Acoustical study of classical Peking Opera singing. J Voice. 2012;26: 137–43.
19. Cleveland T, Sundberg J, Stone RE. Long-term-average spectrum characteristics of country singers during speaking and singing. J Voice. 2001;15:54–60.

2014 4th International Conference on Applied Social Science (ICASS 2014)

Research Methods for Tibetan Speech Physiological Multimodal Study

Yonghong Li

Key Lab of China's National Linguistic Information Technology, Northwest University for Nationalities, Lanzhou, Gansu, China

email: lyhweiwei@126.com

*Yonghong Li

Keywords: Tibetan, Lip signal, EPG signal, Voice signal, Respiratory signal, Multimodal

Abstract. This paper has used the modern biological technology, from the perspective of multimodal, to study on Tibetan speech physiology aspect. Using lip video signals to study the Tibetan consonants lip changing; using EPG signals to study the tongue palatal contraction when Tibetan language pronunciation; using voice signal to study the Tibetan voice types; using respiratory signals to study the reading breathing way for Tibetan language. This study can improve the basic theory research of Tibetan language level and application level in the information science, in Tibetan speech physiology synthesis, speech pathology, language teaching and so on also has a broad application prospect.

藏语语音生理多模态的研究方法

李永宏

西北民族大学中国民族语言文字信息技术重点实验室，兰州，甘肃，中国

Email: lyhweiwei@126.com

*通讯作者:李永宏

关键词: 藏语; 唇位信号; 动态腭位信号; 嗓音信号; 呼吸信号; 多模态

中文摘要. 本文利用现代生理技术，从多模态的角度进行藏语言语生理研究。通过唇位视频信号，研究藏语辅音的唇形变化；通过动态腭位信号，研究藏语发音时舌腭接触情况；通过嗓音信号，研究藏语的发声类型；通过呼吸带信号方面，研究藏语朗诵的呼吸方式。本研究可以提高藏语的基础理论研究水平和在信息科学方面的应用水平，对藏语语音生理合成、言语病理矫治、语言教学等方面也都有着广阔的应用前景。

1. 引言

藏族是我国人口众多、历史悠久，拥有灿烂文化的古老民族。藏语是汉藏语系中藏缅语族的一支，分为三大方言：拉萨方言、康方言和安多方言。方言之间的差别主要集中在语音方面，其次是语法和词汇。语音上的差异主要体现在三个方面：声母有无清浊的对立、辅音

978-1-61275-066-8/10/$25.00 ©2014 IERI

韵尾的多寡、有无声调。虽然三大方言在口语上有比较大的差异性，但是他们基本上与文字各有一套整齐的对应规律，因此三种方言都可以通用的用来拼读文字。

藏语声学在上世纪末有一些研究（胡坦，1980，2002；华侃，1985，1986；鲍怀翘，1992；瞿霭棠，1995；孔江平，1991，1995；江荻，1996，1997）。西北民族大学中国民族语言文字信息技术重点实验室，购买了一批国内外一流的语音生理设备，积累了大量的语音语料库和研究经验，取得了一系列的成果，是国内目前进行藏语声学和生理研究最为深入和全面的单位[1]。

语音多模态研究主要是指对某种语言进行语言学、语音学、语音声学和语音生理学的全方位研究。在理论上，注重语音产生理论的研究和语音共性的研究。在方法上，尽可能采用声学、生理、心理学的研究方法，采集各种声学、生理信号和心理信号[2]。本文通过对藏语方言的语音、嗓音、腭位、唇位和呼吸等多模态信号的采集，建立的语音生理数据库，并根据信号类型提取生理参数，研究藏语多模态研究的数字化方法和技术标准。

2. 唇位研究

嘴唇是发音器官的重要组成部分，而且它是唯一视觉上完全可见的发音器官，因此在语音学研究的早期，人们就开始关注唇形与语音之间的关系。通过唇形参数，我们可以研究藏语发音时开口度大小、发音时间的长短，辅音成阻和除阻状况。

二维唇形主要通过对发音过程进行摄像，标记每帧图片的内外唇线（如图1所示）。二维唇形研究中提取参数共有11个，具体包括：外唇宽度（w1）、内唇宽度（w2）、上唇外轮廓开口度（h1）、下唇外轮廓开口度（h2）、上唇内轮廓开口度（h3）、下唇内轮廓开口度（h4）、唇角闭合曲线开口度（h5）、人中凹陷程度（xoff）、等。通过以上参数可以量化不同音位的唇位变化（如图2所示）。

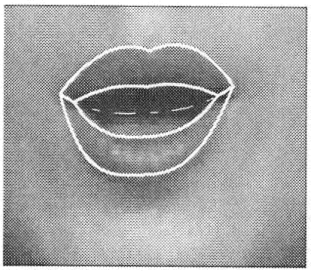


图1 二维唇形信号处理

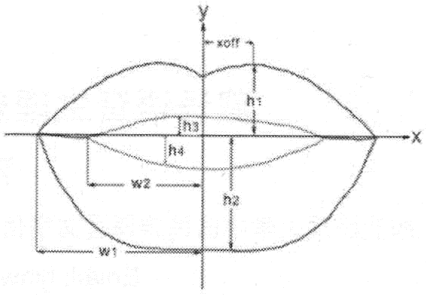


图2 二维唇形参数定义

下面以藏语夏河话的复辅音声母发音时的唇形变化为例进行研究，夏河话复辅音声母主要以下面三种形式：

（1）前置辅音为鼻音：[nd]、[ndz]、[ndʐ]、[ndʑ]、[mb]、[ŋg]

（2）前置辅音为擦音：[hm]、[ht]、[hn]、[hɕ]、[hk]、[hɳ]、[hŋ]、[hts]、[htʂ]、[htɕ]、[ɦm]、[ɦd]、[ɦn]、[ɦz]、[ɦl]、[ɦʑ]、[ɦj]、[ɦɳ]、[ɦdʑ]

（3）后置辅音为[w]：[hw]、[ɦw]、[kw]

从表1可以得出以下结论：

（1）外唇宽度平均值的变化弧度在30.22-36.28之间，内唇开口度平均值变化弧度在0.95-5.82之间。外唇宽度和开口度的变化弧度差别不大，但外唇宽度变化弧度还是稍大于开口度变化弧度；

（2）结合所选音节发音时的内唇宽度和开口度图，我们可以看到比起外唇宽度，内唇开口度更能体现复辅音声母音节发音时的特点；

（3）内唇开口度和外唇宽度之间并无绝对的对应关系。在复辅音音节中，外唇宽度并不一定会随着内唇开口度的增大而减小。

表1 外唇宽度平均值与开口度的平均值表

类型	复辅音	外唇宽度平均值	开口度平均值	类型	复辅音	外唇宽度平均值	开口度平均值
前置辅音为鼻音	nd	36.28	3.62		hk	32.11	1.82
	ndz	30.58	1.52		htɕ	33.43	2.2
	ndʑ	32.74	2.26	前置辅音为浊擦音	ɦm	33.57	2.47
	ndz̩	34.55	5.82		ɦn	32.33	1.65
前置辅音为清擦音	hm	34.44	1.16		ɦɳ	31.71	4.66
	hn	32.29	0.95		ɦz	32.83	3.29
	hɳ	33.1	4.21		ɦʑ	33.25	5.53
	hŋ	33.37	2.79		ɦl	31.92	4.2
	ht	32.1	1.32	后置辅音为鼻音[w]	ɦw	31.9	4.02
	hts	30.95	1.69		hw	32.8	0.99
	htʂ	30.22	2.44		kw	32.56	1.56
	htɕ	33.1	2.07				

3. 嗓音研究

发声（phonation）是指声带在气流的作用下，以不同的振动方式而产生的声源，声源主要包括了声带振动的频率，即振动的快慢，以及声带振动的方式。通过对嗓音声源的研究，可以了解发音时的声带振动情况，从而使我们能够更好地认识语音发声的生理机制、语音发声的微观运动、各种发声类型的特性和语音声学信号的关系。

电子声门仪（Electroglottograph Model）也称喉头仪，主要用于言语嗓音及与言语病理相关的科学研究和诊断。其原理是将一对电子感应片分别固定在喉结两边贴紧甲状软骨，发声时一个非常微弱高频信号从一个电子感应片发送，被另一个接受。当声带完全接触，即声门完全关闭时，阻抗值最小；当声带分开，即声门完全开启时，阻抗大大增加（如图3所示）。

EGG信号主要包括基频、开商、接触商、速度商四类参数，具体算法如下：

1）基频（Pitch）= 1/周期（A）；

2）开商（Open Quotient，简称OQ）= 开相（C）/周期（A）×100%；

3）接触商（Contact Quotient，简称CQ）= 闭相（B）/周期（A）×100%；

4）速度商（Speed Quotient，简称SQ）= 开启相（E）/关闭相（D）×100%。

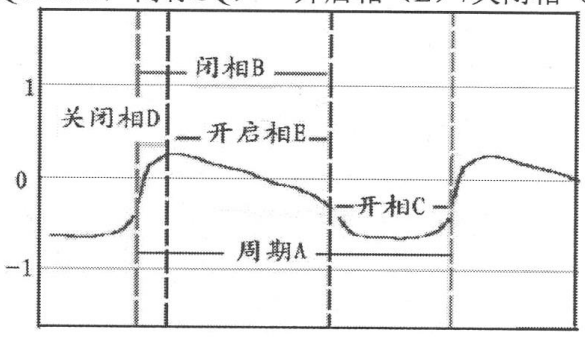


图3 EGG信号分析图

由公式可知，开商和接触商是一组相对的概念，由于开相和闭相的和为周期，示意图上表示为B+C=A，故可得OQ+CQ=1（如图3）。目前，关于嗓音发声类型方面的研究，主要见北京大学孔江平教授在《论语言的发声》，书中对多个民族语言的发声类型进行了探究[3]。以藏语夏河话为例，取/a/,/i/,/ə/，/o/四个元音，表2为元音嗓音参数。

表2 元音嗓音参数

参数类型	a	e	ə	o
基频	119.2	122.4	111	109.9
开商	53.7	58.9	55.0	52.0
速度商	206.0	214.2	212.0	216.1

4. 动态腭位研究

EPG(electropalatography)是一种在发音状态下测定舌腭接触的技术。EPG能真实反映语流中舌腭接触的细节变化，并可以和许多其它仪器联合使用。按照发音人口腔形状做成电子假腭（内含62/96个银质电极），根据发音时舌头与硬腭接触时通电与否，舌腭的接触位置在屏幕上显示出来（1秒种内拍摄100张舌腭接触照片），给研究者以视觉反馈。通过相关软件，提取AC、PC、VC、Ant、Pos、CA、CP、CC等参数。

TC（舌腭接触最大帧的接触总面积比）=接触的电极数(n)/总电极数(62)

AC（齿龈区接触面积）=齿龈区接触的电极数(n)/齿龈区电极总数(22)

PC（硬腭区接触面积）=硬腭区接触的电极数(n)/硬腭区电极总数(24)

VC（软腭区接触面积）=软腭区接触的电极数(n)/软腭区电极总数(16)

Ant（前腭接触面积）=前腭区接触的电极数(n)/前腭区电极总数(30)

Pos（后腭接触面积）=后腭区接触的电极数(n)/后腭区电极总数(32)

靠前性指数（CA）

$CA=(log(1*(R(8)/8)+9*(R(7)/8)+81*(R(6)/8)+729*(R(5)/8)+6567*(R(4)/8)+59049*(R(3)/8)+531441*(R(2)/8)+3587227*(R(1)/6)+1))/(log(4185105))$

靠后性指数（CP）

$CP=(log(1*(R(1)/6)+1)+9*(R(2)/8)+81*(R(3)/8)+729*(R(4)/8)+6567*(R(5)/8)+59049*(R(6)/8)+531441*(R(7)/8)+3587227*(R(8)/8))/(log(4185105))$

趋中性指数（CC）

$CC=(log(1*((C(1)+C(8))/14)+17*((C(2)+C(7))/16)+289*((C(3)+C(6))/16)+4913*((C(4)+C(5))/16)+1))/(log(5220+1))$

以藏语夏河话为例，单辅音能与元音/a/相拼的共有20个，其中舌面前j为半元音，有元音的性质，/k/、/kh/、/h/发音部位靠后，超出EPG采集范围，所以这四个音不在研究之中。其余16个辅音/th/、/s/、/z/、/ʦ/、/ʦh/、/ʂ/、/tʂ/、/tʂh/、/ʐ/、/ɕ/、/tɕ/、/tɕh/、/l/、/n/、/ɳ/、/r/均由舌腭接触造成阻碍或阻塞调音，选取舌腭接触最大帧的那一帧作为目标帧（如图4所示）。

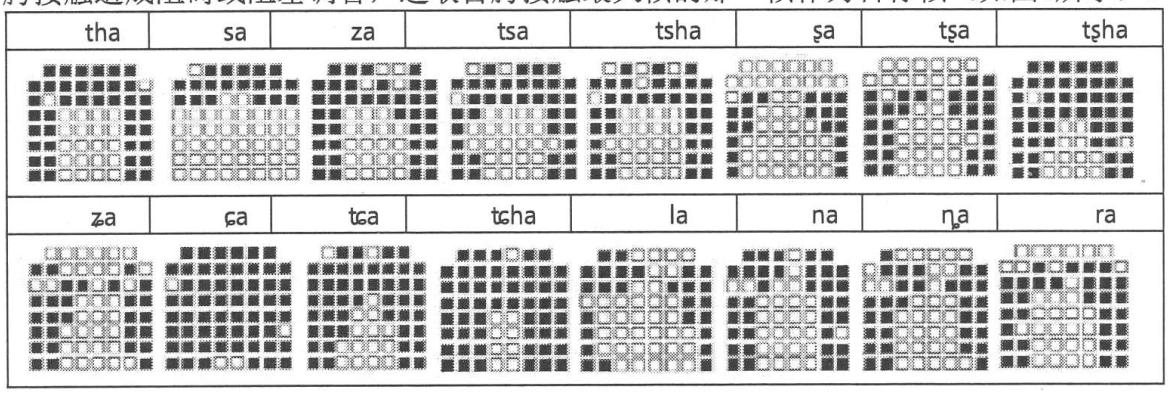


图4 藏语夏河话单辅音舌腭接触图

塞音因为发音部位靠前，在前腭区有较多的舌腭接触，靠前性强，发音时完全阻塞趋中性也较强。擦音因为不完全阻碍一般要比同部位的塞擦音的接触电极少。边音发音时舌尖形成阻碍，气流从舌边流出，因而齿龈区、前腭区的舌腭接触面积大，靠前性、趋中性强，靠后性比较弱。鼻音发音时口腔里形成的阻碍完全闭塞，各个参数值也都比较大。颤音发音时，舌面参与发音，舌腭接触电极较多，靠后性、趋中性也比较大[4]。

表3 藏语夏河话单辅音腭位参数表

	TC	AC	PC	VC	Ant	Pos	CA	CP	CC
塞音	+-	++	-	-	+	+-	++	++	+
擦音比塞擦音	-	-	-	-	-	-	-	-	-
边音	-	++	0	0	++	0	++	-	+
鼻音	+	++	+	+	+	+	++	++	+
颤音	+	++	+	+-	++	+-	++	++	++

注：80-100为++，60-80为+，40-60为+-，40一下为-

5. 言语呼吸研究

呼吸信号的采集使用PowerLab生物信号采集处理系统，包括16通道的采集器、2根呼吸带传感器，采集软件使用Chart7。呼吸带传感器可以测量呼吸导致的胸腹部收缩扩张的变化，将一根呼吸带传感器系在发音人的胸部，另一根系在发音人的腹部，由压电传感设备检测出发音时呼吸带长度的变化，从而获得两个通道的呼吸节奏信号。

一个完整的呼吸过程，称之为呼吸周期（Breath circle），一般情况下包括一个吸气相(Inspiration Phrase)和一个呼气相(Expiration Phrase)。胸呼吸信号曲线上升为吸气过程，一般对应于语音信号的静音段；信号下降表示呼气过程，一般对应于语音信号的语音段。胸呼吸参数见图5，横坐标为采样点，纵坐标为呼吸信号的幅度，其中ID为吸气相时长，P为峰值，ED为呼气相时长，V为谷值。

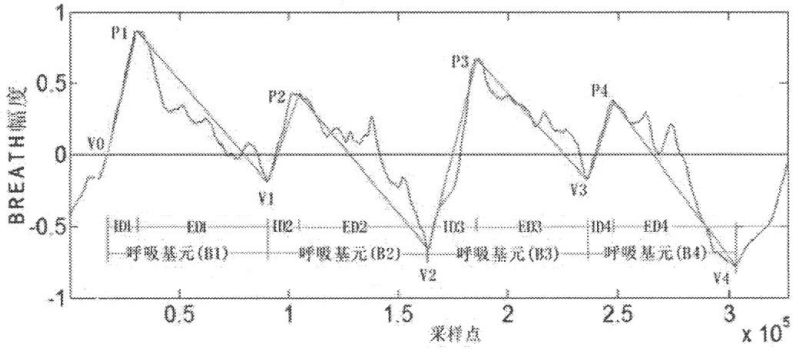


图5 呼吸参数示意图

取出藏语40首诗歌的峰值点（P1、P2、P3、P4）和谷值点的（V1、V2、V3、V4）的实际数值，绘制折线图，从宏观上来看吸气和呼气单元的幅值在诗歌朗读中的分布（如图6所示）。用呼吸基元的时长和幅度的平均值建立诗歌呼吸信号的模型，用线性方程进行模拟。每句诗歌的吸气和呼气单元都包括斜率（IK、EK）和截距（IB）两个参数，整首诗歌共有16个参数，结果如图7所示[5]。

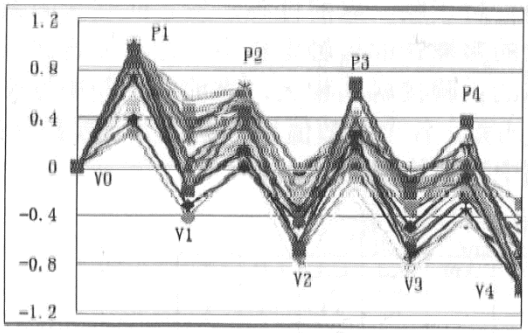


图6 诗歌呼吸信号幅值图

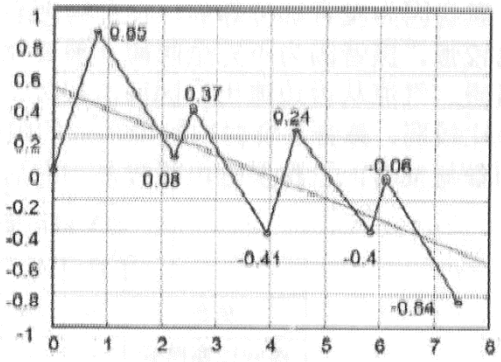


图7 诗歌呼吸信号线性模型图

6. 结束语

从目前的研究现状来看，语言生理的研究主要集中在国外的研究机构，语种主要是英语、日语、法语、汉语等大的语种，而具有多种语言现象和语音生理机制的少数民族语言由于条件的限制难以系统的进行研究。本文利用现代生理技术，从多模态的角度进行藏语言语生理研究，可以提高藏语在信息科技领域的理论研究水平和在信息通讯方面的应用水平，对藏语语音参数合成、言语病理矫治、语言教学与学习、聋哑儿童的语言习得、虚拟主播等方面也都有着广阔的应用前景。

整体来看，语音生理研究主要是对某个生理仪器或者局部的发音部位的研究，很少进行整个发音过程的整合。随着研究技术的不断进步，我们可以不断的推进语音生理研究的深化，不断扩大研究的内涵和外延。因此，我们还需继续加强对语音生理仪器的应用，扩展仪器的应用范畴，以此来帮助我们对言语产生的深入研究，促进现代语音学的发展和进步，进而推动我国语音学理论和应用研究的发展。

致谢

本文为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)和国家民委2012年科研项目《藏语语音多模态数字化方法研究》的阶段性成果之一。

References

[1] Li Yong Hong, Hu Axu, Lv Shilang, Research of Modern Phonetic Instrument and the Physiological Phonetics, *Northwest University for Nationalities*, vol.2, pp30-39, 2012.6

[2] K. Jiangping, The study on the speech multi-mode and diversifying phonetics [J]. *Chinese Journal of Phonetics*, vol.1, pp1-7, 2008

[3] K. Jiangping. Language Phonation[M], The Central University for Nationalities Press, 2001

[4] Li Yonghong, Wang Jianbin, liu sisi, The Research on the Constraint Degree of Tibetan Plosives, Fricatives and Affricates, *2012 2nd International Conference on Future Computers in Education*, vol.22, pp138-142, 2012.6

[5] Li Yonghong, Yang Yangrui, GuoLei, Yu Hongzhi, The Linear Model Study of Tibetan Six-character Pomes' Respiratory Signal, *2012 International conference on medical physics and biomedical engineering*, vol.23.pp919-926, 2012.5

2014 4th International Conference on Applied Social Science (ICASS 2014)

Digitalized Multi-model Methods Study on Mongolian Long-tune Folk Songs

Yonghong Li

Key Lab of China's National Linguistic Information Technology, Northwest University for Nationalities, Lanzhou, Gansu, China

email: lyhweiwei@126.com

*Corresponding author: Yonghong Li

Keywords: Mongolian Long-tune; Speech signal; Voice signal; Respiratory signal; Multi-modal

Abstract. Using types of advanced phonetic speech instruments, this paper from the speech physiological multi-model perspective has collected the speech signals, the voice signals and the respiratory signals from the Mongolian long-tune folk songs. With the speech signal, this paper has extracted the length, pitch, intensity, quality and other parameters; with the voice signal, it has studied the Long-tune's own special phonation types; with the respiratory signal, this paper has studied the breathing mechanism of chest and belly while performing. With the qualitative research methods of computer technology, it has explained the inner mechanism including the acoustic and physiological phonation ways in order to build some foundations for the more scientific standards of the digitalized protection for oral cultures.

蒙古长调的多模态数字化研究方法

李永宏

西北民族大学中国民族语言文字信息技术重点实验室，兰州，甘肃，中国

email：lyhweiwei@126.com

*通讯作者:李永宏

关键词: 蒙古长调；语音信号；嗓音信号；呼吸信号；多模态

中文摘要. 本文从语音生理多模态的角度出发，利用多项先进的语音生理设备，采集长调演唱时的语音信号，嗓音信号和呼吸信号。语音方面，通过音长、音高、音强和音质等参数，研究长调的韵律特征；嗓音方面，研究长调演唱中特殊的发声类型；呼吸方面，研究长调演唱中气息的使用和胸腹呼吸的机制。用计算机量化的研究方法解释长调的声学和生理的内在机制，为建立科学和规范的口传文化数字化保护标准建立基础。

978-1-61275-066-8/10/$25.00 ©2014 IERI

1. 引言

中国是一个注重文化典籍挖掘、整理和保护的国家，这是我们的文化传统。但中华口传文化一直都没有很好的被记录和传承。普通文字、语音、图像和录像已不能满足不同口传文化的记载和传承，要引入更多声学、生理和心理的信息，如国际音标、声带的发声方式、呼吸方式、语音的情感、认知心理等信息，只有这样才能更全面地记录中华口传文化的内容和形式[1]。北京大学孔江平教授首次提出了建立“口传文化的多模态研究”。这是一项跨学科的研究，涉及了人文科学中的语言学、语音学、田野调查方法和自然科学中的言语声学、言语生理学和信息科学。

长调，蒙古语称“乌日图道”，是对气息悠长、历史久远、节奏舒缓、意境开阔的民蒙古歌的一种传统称谓[2]。蒙古族长调民歌是北方草原游牧文化中具有代表性的音乐艺术品种，更是被国际权威机构认定的具有世界级文化地位的艺术文化品牌，2005年，蒙古族长调民歌成功入选世界非物质文化遗产名录。蒙古长调在传统的曲艺、内容、风格方面研究的比较多，声学和生理研究近两年才开始，例如蒙古语长调民歌的呼吸信号生理分析[3]，蒙古语长调民歌《圣主的两匹骏马》声学分析[4]。

为了展现全面的研究长调的演唱机理，选择采集的原始信号包括：音频信号、视频信号、动态电子腭位信号、电子声门信号、视频唇形信号和朗读呼吸节奏信号，是全面反映蒙古语的生理声学的基础原始资料。我们将重点从语音特征、嗓音发声类型和呼吸方式三个方面进行研究。

2. 语音声学研究

声音信号是口传文化最直接的信号，也是人类感知中最直观的信号，容易记录好保存，是长期以来最为广泛的保护和传承方式。

2.1 实验设备

语音信号的采集设备一般包括：高质量领夹式话筒、调音台、外置声卡和笔记本电脑。考虑到声乐的基频变化比较大，频谱范围比较广，因此信号采样频率一般为48kHz，量化精度为16bit，录音格式的没有压缩的PCM编码的“.wav”文件。如果只录制单通道的声音信号，Audition或者其他的录音软件都可以，分析时可以使用Praat或者Multispeech软件。

2.2 时长特征

蒙古长调一直以“字少腔长”为主要表现特点。音符在时间上的区分度很大。音长即为音的持续时间。可以提取演唱时每一个音符和每一段的时间长度，主要体现节奏在时间上的表现，一般以秒(s)或者毫秒(ms)为单位。以《圣主的两匹骏马》为例，按照长调歌词的音位单元来进行切分，整首长调共得到30个音段。根据音段时长大小，可分为两组，8个长音和21个短音，其中长音基本均为主节拍的末尾拖腔音。长音平均时长为4.5s，短音平均时长0.6s，长音是短音的7倍多。

2.3 基频特征

讲话时的语音基频和长调的语音基频是完全不同的，一般长调的要略低一些，而拖腔中则是相反，往往拖腔的基频会高于长调的基本基频，这也说明拖腔在长调中扮演着释放情感的功能。长调女声的基频范围从200Hz到550Hz，男声的基频分布范围较窄，在200-500Hz之间，平均基频约为350Hz，女声的基频分布范围则相对宽，在200-600Hz，平均基频约为410Hz。女声的平均基频比男声高60Hz，且男声的音域范畴比女声小100Hz左右。

颤音是长调中使用最为广泛的一种艺术方式，发出颤音时，气流会快慢相间，声带有节奏的时快时慢，喉部肌肉会剧烈跳动。图1 为取男声一段颤音的基频变化图。

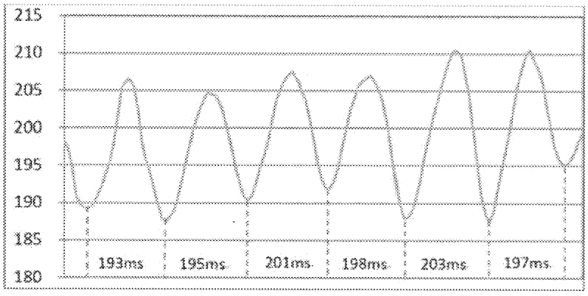


图1 颤音基频周期变化图

颤音的基频随着时间呈波浪方式上下变化。颤音基频的上限值为210Hz,下限值为187Hz，每隔200ms左右有一次重置，我们称之为基频变化的周期性或者为颤音单元。一个颤音单元约为200ms，变化幅度约20Hz左右。颤音的平均基频以200Hz计算，那就是说声带的振动周期为5ms。那么在一个颤音单元内，声带振动了40个周期(200ms/5ms)，每个周期基频的变化越为1Hz，也就是说，连续的周期的时间增加或者减少约为1ms。这也反映出了人类控制声带的微观方式。

2.4 音强特征

音强的变化是声乐表现的重要因素之一,任何一种具有很强表现力的旋律往往都有着丰富而细腻的力度变化。一般来说,音强的变化有逐渐增强、逐渐减弱、由弱到强再到弱、或是极强、极弱、由极强到极弱、由极弱到极强等。以颤音的能量变化为例，图2中上图为选取的一段颤音的语音波形图，下图为颤音的三维语谱图和能量变化曲线。

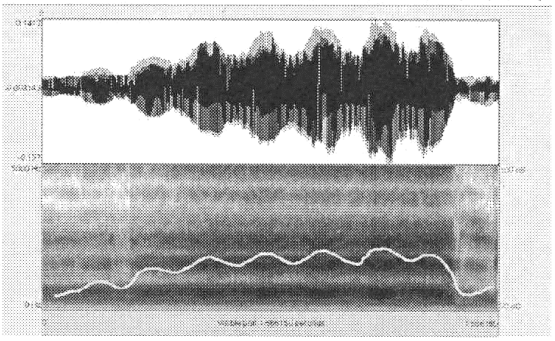


图2 颤音能量变化图

整体上来看，颤音的能量从弱逐渐变强的，并且呈周期式变化，与颤音的基频类似。经过测量，颤音的每一个能量周期时间长度平均在190ms-200ms之间，中间的平稳周期峰值到谷值的幅度变化约为4dB，也就是说稳定的颤音能量在200ms内的能量浮动约为4dB。峰值幅度最小值为60dB，最大值约为70dB，也就是说在颤音的起始到高潮整体上越有10dB的提升。

2.5 频谱分析

从共鸣特征来看，共鸣在歌唱发声中是很重要的因素，也是歌唱艺术中的重要表现手段，它能使人的声音变得清脆、明亮，更富有穿透力。很多研究歌唱的结论中都有歌唱共振峰的提法，认为在2400-3200Hz频率附近有一个比较高能的共振区，对音质的贡献比较大。我们选取同一个内容的长调和念白，计算长时平均功率谱，如图3所示。

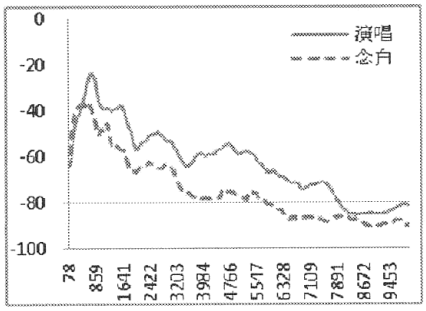


图3 长时平均功率谱

可以看出，800Hz以下的低频区域，演唱和念白频率基本一致，800Hz-8000Hz频率段，演唱的能量要明显高于念白。在3500Hz-5000Hz范围内演唱要比念白高20Db,高频区域的能量提升，也是演唱声音透彻、洪亮、圆润的主要原因之一。

3. 嗓音发声类型

在口传文化中，由于语言发声类型的不同，唱腔表现出多样性，使口传文化具有丰富的艺术感染力。从言语声学和生理学的角度，嗓音的发声类型和唱腔的变化可以用不同的声学和生理参数来定量描述和定量分析，从而科学地定义嗓音的语言和艺术功能。

3.1 实验设备

电子声门仪（Electroglottograph Model）也称喉头仪，主要用于言语嗓音及与言语病理相关的科学研究和诊断。其原理是将一对电子感应片分别固定在喉结两边贴紧甲状软骨，发声时一个非常微弱高频信号从一个电子感应片发送，被另一个接受。当声带完全接触，即声门完全关闭时，阻抗值最小；当声带分开，即声门完全开启时，阻抗大大增加。根据声门阻抗信号，测出声门的关闭点和开启点，计算出开商和速度商等参数（Fabre1957）。

3.2 嗓音参数

EGG信号主要提取基频、开商、接触商、速度商四个参数，具体算法如下：

（1）基频（Pitch）= 1/周期（A）；

（2）开商（Open Quotient，简称OQ）= 开相（C）/周期（A）×100%；

（3）速度商（Speed Quotient，简称SQ）= 开启相（E）/关闭相（D）×100%。

由公式可知，开商和接触商是一组相对的概念，由于开相和闭相的和为周期，国内主要是北京大学孔江平教授在《论语言的发声》对民族语言的嗓音发声类型讨论的比较多[5]。

3.3 嗓音类型

在长调的演唱过程中，为了展示特殊的艺术效果，有着丰富的、复杂的嗓音发声方式，比较典型的有以下几种形式，见图4：

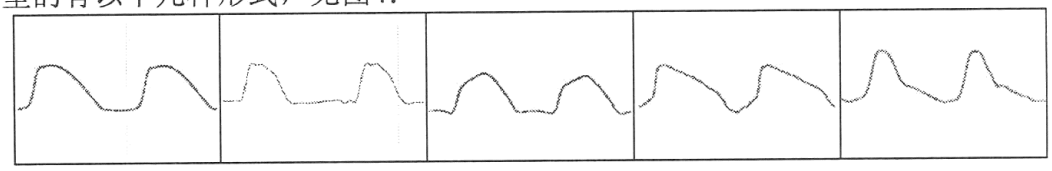


图4 嗓音类型图

从图4的几种EGG信号来看，嗓音的变化是比较大的，声带的接触方式也是根据演唱的风格变化进行调整。图1中声门关闭段小，有稳定的闭合阶段，声门打开段时间比较长，有稳定的开相。图2与图1比较接近，但开相比较大。图3声门关闭段时间比较长，关闭的过程是先快后慢。图4接近图1，但没有稳定的开相。图5声门开启相比较复杂，先快后慢，没有稳定的开

相。另外在实际的演唱中会有更复杂的声带振动方式出现，需要对发音生理进行更为深入的研究。

4. 呼吸动力研究

歌唱的呼气动作是在吸气和呼气两大肌群相对抗的情况下稳健完成的。为了能够反映不同演唱风格在呼吸上的差异性，我们引入了呼吸带来实时的采集呼吸的变化。

4.1 实验设备

呼吸信号的采集使用的PowerLab生物信号采集处理系统，包括16通道的采集器、2根MLT1132呼吸带传感器，采集软件使用Chart7。呼吸带传感器可以测量呼吸导致的胸腹部收缩扩张的变化，将一根呼吸带传感器系在发音人的胸部，另一根系在发音人的腹部，由压电传感设备检测出发音时呼吸带长度的变化，从而获得两个通道的呼吸节奏信号。

4.2 呼吸参数设计

呼吸参数的设计，需要根据研究的需要进行定义。一个完整的呼吸过程，称之为呼吸周期（Breath circle），一般情况下包括一个吸气相(Inspiration Phrase)和一个呼气相(Expiration Phrase)。胸呼吸信号曲线上升为吸气过程，一般对应于语音信号的静音段；信号下降表示呼气过程，一般对应于语音信号的语音段。胸呼吸参数见图5，腹呼吸和胸呼吸的相互关系还需要深入研究。

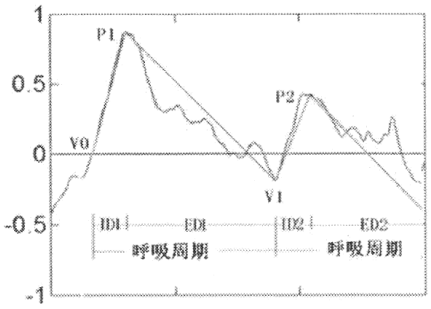


图5 呼吸参数示意图

ID为吸气相时长P为峰值；ED为呼气相时长V为谷值，吸气量IQ为吸气相从谷值变化到峰值的幅度差，表示人体吸入气流量的大小；呼气量EQ为呼气相从峰值降到谷值的幅度差，表示演唱时所呼出的气流量，气流总量BQ代表呼吸单元的气流总量，而持续段没有气流量变化，为零值。斜率代表吸气量和呼气量的变化速度，斜率的绝对值越大，气流速度越快[6]。

4.3 呼吸类型

歌唱呼吸主要通过增加呼吸的深度等途径来加大气息量。歌唱时的呼吸运动是在意识与非意识的双重作用下进行的。在一定的限度内，呼吸的深度和频率是可以用意识加以改造。观察长调的胸呼吸和腹呼吸关系可以发现，基本上表现出胸腹式、胸主式和腹主式三种呼吸模式，如图6所示。

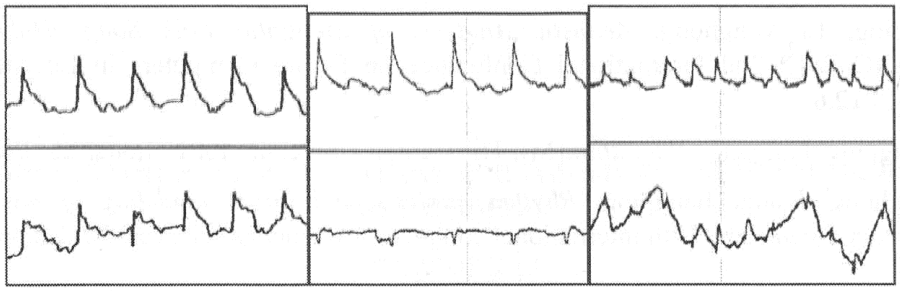


图6 胸腹呼吸图

图6中，上半图为胸呼吸信号，下半图腹呼吸信号，根据胸腹之间关系得出三种呼吸模式。

（1）胸腹式

胸呼吸信号和腹呼吸信号共同作用，走向基本保持一致。吸气时胸腹同时扩张，横膈膜下降，信号急剧上升；歌唱时胸腹快速收缩，信号急剧下降，然后后半段胸腹信号均趋于水平直线，此时胸腹肌肉都比较紧张，胸腔内的空气压力持续下降（见图6左）。

（2）胸主式

歌唱时主要以胸呼吸为主，腹部变化较小。吸气时胸部扩张，信号急剧上升，横膈膜下降，腹部略微扩张；歌唱时胸部快速收缩，信号急剧下降，而腹部基本保持不变（见图6中）。

（3）腹主式

歌唱时主要以腹呼吸为主，胸呼吸信号比较稳定，横膈膜上升下降幅度比较大，腹呼吸信号变化比较剧烈（见图6右）。

在实际的长调歌唱中，发音人可以灵活的调整自己的气息变化和肌肉的运动，采集到胸呼吸信号和腹呼吸信号的变化如何反映气息的真实变化是一个需要深入研究的课题。

5. 结束语

本文从多模态的角度对长调口传文化的数字化传承与保护的方法进行了论述。目前，处于积极的探索阶段，还不能完全建立标准的、规范的和统一的数字化保护方法，还需要多方面的深入研究。从宏观上来看，声学、生理和心理信息的保存，能更全面地记录中华口传文化的内容和形式。不同的口传文化在局部的艺术形式上有所差别，但都可以采用同一种采集系统进行统一的采集和整理。但随着研究技术的不断进步，我们可以不断的推进保护方法的深化，不断扩大保护的内涵和外延。

致谢

本文为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)的阶段性成果之一。

References

[1] K.Jiangping. *The interaction of humanities and science and technology*-Linguistics lab. of Peking University and the oral cultural heritage protection and digital research, journal of Peking University (philosophy and social sciences edition), 2012.04

[2] W. Lanjie. T*he Mongolian Music*, Hohhot, Inner Mongolia people's publishing house, 1998

[3] L.Yonghong, Fang Huaping.etc *Physiological Analysis of Mongolia Folk Long Song Breathing Signals* [C]. 2012 International Conference on Research Challenges in Social and Human Sciences(ICRCSHS 2012), 2012.6, pp261-267

[4] F.Huaping, Li Yonghong. *Acoustic Analysis of Mongolia Folk Song 'Holy Lord Two Horses'*[C].2012 2nd International Conference on Future Computers in Education (ICFCE 2012). 2012.6

[5] K. Jiangping. *Language Phonation*[M], The Central University for Nationalities Press, 2001

[6] L.Yonghong, Kong, Jiangping. *Rhythm analysis and linear modeling of metrical poetry respiratory signal*. The 17th international congress of phonetics sciences, 2011.8,pp1222

普通话圆唇研究*

潘晓声　孔江平

提要　本文目的在于讨论汉语普通话圆唇特征的定义问题。本研究从汉语普通话录像中提取出元音发音时关键帧的唇形内外轮廓线，并进一步计算得到唇圆度、宽度和开口度等参数，用于分析圆唇元音、非圆唇元音与上述参数之间的关系。实验结果表明，唇圆度和唇开口度并不具备区分圆唇元音和非圆唇元音的功能，而唇宽度具有区别意义的功能。完整发音过程可以分为张嘴与闭嘴两个阶段，二者有着不同的唇形运动模式。综合此二阶段唇形的变化规律，本文主张使用内外唇宽度之和为特征，用其变化表示圆展唇运动的动态过程。实验还发现普通话发音的唇部生理运动具有一定的随意性，但基本上总能保持同一发音人圆唇元音比非圆唇元音的唇宽度更窄这一特性。

关键词　汉语普通话　圆唇　唇宽度　唇开口度　唇圆度　唇突度

1　引言

圆唇是语音学和音系学的一个基本概念，在各种教学参考书中都可以见到，但各家的描述并不统一，本文希望通过实验语音学的方法找到关于圆展唇运动特征的最合适定义。

音系学研究中，有学者（Chomsky, Halle 1968; Ladefoged 1975）将圆唇作为一种区别性特征，用于区分语音类别。语音学研究的结果认为元音音质是由舌位前后、高低和是否圆唇所共同决定，这三者都通过对共鸣腔形状的影响来改变元音的共振峰结构。另外，只有圆展唇具有语言学意义，语音学家们在研究唇的协同发音时（Hardcastle, Hewlett 1999），一般也只关注圆唇元音和非圆唇元音

* 本文得到国家社科基金重大项目（13&ZD132）、国家自然科学基金支持项目（61073085）、国家社科青年基金项目（12CYY031）和教育部社科青年基金项目（12YJC740082）的资助。

之间过渡时唇形的变化规律。

《普通语言学纲要》（罗常培、王均 2002）关于圆唇元音的定义是“双唇撮敛成圆形，不圆唇元音是两唇舒展，成扁平形或保持自然状态”;《语音学教程》（林焘、王理嘉 1992）中指出“嘴唇圆的是圆唇元音，嘴唇不圆的是不圆唇元音”；而《现代汉语》（北京大学中文系现代汉语教研室 2009）中的定义则是“嘴唇向前伸，呈圆形，是圆唇元音”。然而某些不属于圆唇元音的语音，比如［ɻ］，在发音时，其嘴唇形状也很圆，因此本文认为唇形状上呈现圆形并不能用于区分圆唇元音和非圆唇元音。此外，上述教材都明确指出圆唇元音的定义主要考虑的是唇形呈圆形，但都没有进一步说明圆唇指的是嘴唇外轮廓还是内轮廓呈现圆形。值得注意的是，除了在《现代汉语》中指出圆唇元音的发音与突唇动作相关之外，另几种教材都仅仅是以嘴唇形状是否是圆形作为判断圆唇元音的标准。在一些语言中，唇突度可以作为圆唇元音的区别特征。比如Lindau（Lindau 1978）和Ladefoged等人（Ladefoged，Maddieson 1990）提出使用垂直敛唇（vertical lip compression）和水平撮唇（horizontal lip protrusion）来区分瑞典语的圆唇元音。胡方（2007）提出可以使用水平撮唇（horizontal protrusion）和垂直撮唇（vertical protrusion）来区分宁波方言的两个前高圆唇元音［y］和［ʏ］。Chomsky等人（1968）在《英语音型》一书中，专列了一项关于唇的区别性特征：圆唇和非圆唇，并指出圆唇由唇孔变窄所形成，非圆唇则否。Ladefoged（1975）的区别性特征系统中也包括了圆唇化这一特征，但他用两嘴角水平宽度的倒数来定义圆唇程度。Chomsky和Ladefoged定义圆唇特征的区别在于，前者使用唇形内轮廓的宽度，而后者使用唇形外轮廓宽度。Abry等人（1986）在研究中指出，所有语言中区分圆唇元音和非圆唇元音的参数是嘴唇宽度，即开口时的水平宽度，它和所有关于嘴唇突度的参数成反比。前人的研究从个人的主观经验出发，观察的角度不同，则使用的参数也不同。唇圆度、唇突度、唇宽度和唇开口度都有人使用，因此

各家对圆唇的定义也并不一致。

《实验语音学概要》（吴宗济、林茂灿 1989）中，鲍怀翘提取了五位被试的唇形参数，包括唇突度、唇开度和唇形面积。结果表明展唇元音的开口度大于圆唇元音的开口度，舌高度相同的元音，展唇元音的唇形面积大于圆唇元音。另外，作者认为唇突度特征对于区分元音圆展作用不大。但由于当时的实验条件，无法提取到发音时嘴唇的动态变化，所采集的唇形参数的精度也受实验器材和算法的限制。今天，我们可以利用数字化技术提取到精度更高的唇形参数，本文对唇形参数的提取和分析将做进一步细化处理。

鉴于圆唇的定义对于语音学、音系学都是一个重要的基础问题，各家对于圆唇特征定义不同的主要原因可以归为两点：1、圆唇只是一个概念，在舌位相同时人们发现圆唇的变化对元音的声学特征有影响，因此用它作区别性特征。究竟用唇圆度还是其他参数来定义并不重要，因此没有必要精确的提取参数作分析。2、受技术手段和实验条件的限制，无法提取某些唇形参数，或提取的参数精度不够，无法对唇形进行量化的描述。

一般认为，发圆唇元音时，唇形呈现撮拢状态，而非圆唇元音则否。因此，本文假设圆唇元音比非圆唇元音的唇宽更窄，唇形更圆，唇开口度更小。通过对8个发音人汉语普通话圆唇元音和非圆唇元音样本的唇形参数进行统计，以验证上述假设是否合理，并希望找出最合适的参数用于定义圆唇特征。

2 实验方案

2.1 实验材料

本文希望作为样本的元音，发音时唇形动作不受到辅音的影响。汉语普通话的单元音中，/ə/、/a/和/ɑ/都不能单念，一定要和辅音相配合发音，且配合不同发音部位的辅音时，这几个元音所

表现的唇形不同，即元音唇形受到辅音发音动作的影响，因此将其排除。/i/的两个变体/ɿ/和/ʅ/也必须和辅音相结合才能发音，但/ɿ/只能与舌尖前辅音/ts/、/ts^{h}/和/s/相配合，/ʅ/必须要与舌尖后辅音/tʂ/、/tʂ^{h}/、/ʂ/和/z̩/相配合，它们的唇形受辅音的影响基本相同，且不存在不受影响的形式。因此本文最终选择8个汉语普通话零声母元音加上两个非零声母单元音/ɿ/和/ʅ/，共10个音节进行圆唇特征研究。其中元音［ʅ］和［ɿ］分别配合声母［tʂ］和［ts］发音。/ɛ/和/o/平时使用较少，一般作为普通话的语气词存在。最终所使用的发音样本如表1所示：

表1　汉语普通话元音的国际音标及对应汉字和拼音标注

国际音标	ɑ	ɤ	ɛ	ɚ	i	o	u	y	ʅ	ɿ
汉字	阿	饿	欸	儿	衣	哦	屋	淤	之	资
汉语拼音	a	e	ê	er	i	o	u	ü	zhi	zi

为了使样本来源具有更广泛的代表性，而不仅仅是代表受过专业语言学训练的发音人，本文的8位发音人中并没有全都受过语言学专业训练，但所有被试人员都受过高等教育。

此外，不是所有发音人都认识国际音标，因此实验不要求发音人读这些元音所对应的国际音标，而是给出这些音节所对应的汉字与拼音标音，并求发音人按拼音方案中的标音朗读。发音人具体朗读的汉字以及拼音标音见表1。

8位发音人，10个音节，每人朗读所提示的汉字各三次，本文共得到240段录像用于提取实验参数。

为更精确的提取唇形参数，我们要求发音人正面朝向录像机，在发音时头部基本保持静止不动，以免得到的参数受点头运动和发音人身体晃动的影响。发音从闭唇状态开始，发音完成后最终恢复到闭唇状态，在两个音节之间停顿1到2秒，由此可以尽量减少前后音节的协同发音影响。要求被试发音持续一段时间，以保证可以

从录像视频中提取到代表此元音发音动作稳定段的唇形关键帧。

实验数据的采集在上海师范大学语言研究所的专业隔音室录像室中进行。实验使用专业录像机采集发音人的唇形数据，录像速率为30帧每秒。

2.2 唇形参数提取

本实验采用的是北京大学中文系语音乐律实验室二维唇形处理平台提取唇形参数，该平台在Windows平台下用Matlab语言编程实现。

Liew等（2000）假设唇形是左右对称的，通过对唇形的变形可以得到任意形状的唇形轮廓，并用双曲线表示唇形的外轮廓。但在语言学研究中，唇形的内轮廓与语音的相关性往往要高于唇形外轮廓。由此，本文在Liew研究工作的基础上，加入了描述唇形内轮廓的参数。

根据内唇的开合状态，本文将唇形分为三大类：闭合状态和半开状态和全开状态。唇形的三种开合状态具体如图 1 所示：

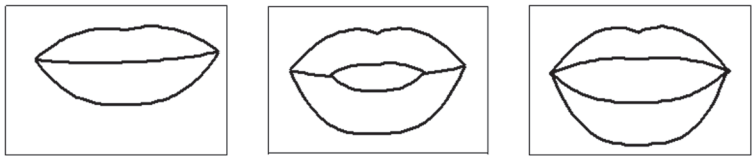


图1（a） 闭合状态　图1（b） 半开状态　图1（c） 全开状态

在闭合状态时，内唇的上下轮廓线重合为一条曲线。事实上，可以把闭合状态看成是半开状态的一个特例。如果半开状态的内唇宽度不断变小，极端情况是两个内唇唇角完全重合，即内唇宽度为0的半开状态就是闭合状态。另外，全开状态也看成半开状态的一个特例。即内唇的两个唇角与外唇的两个唇角完全重合的半开状态就是全开状态。由此，闭合状态和全开状态都可以看成半开状态的特例，三种不同的唇形状态可以用一个模型来表示。

本文在Liew的算法基础之上，加入了描述唇形内轮廓的参数，

建立自已的唇模型，具体如图 2 所示。

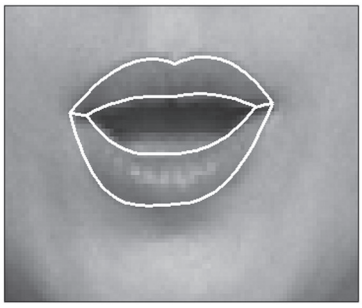


图 2（a） 使用唇模型重构的唇形轮廓

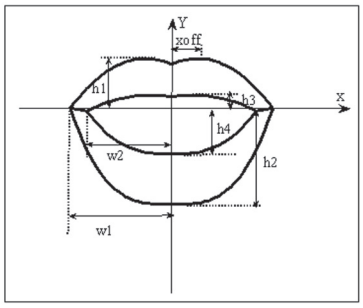


图 2（b） 去除伴随动作重构的唇形

本文称用于重构唇形内外轮廓的参数为重构参数。重构参数具体可以分为两类，分别是在图 2（b）中可以标出的参数，如h1，h2，w1，w2等，以及部分无法在图 2（b）中标出的参数，比如歪嘴程度等。

重构参数共有11个，具体包括：左侧外唇宽度（w1）、左侧内唇宽度（w2）、上唇外轮廓开口度（h1）、下唇外轮廓开口度（h2）、上唇内轮廓开口度（h3）、下唇内轮廓开口度（h4）、人中凹陷程度（xoff）、下唇圆弧度（δ）、唇闭合处曲率（ql）、头部倾斜程度（qx）和歪嘴程度（s）等。

图 2（a）是通过手工改变唇模型的重构参数所重构的唇形轮廓。可以发现，无论是唇形的内外轮廓，本文的唇模型都能精确描述。

人们发音时总是不自觉的会带有歪嘴和头部倾斜等伴随动作，因此图 2（a）中的两嘴角高度并不处在同一水平线上。头部倾斜幅度越大，两嘴角坐标的垂直距离就相差越大，两嘴角坐标的水平距离也会相应的变小，此时两嘴角之间的水平宽度并不能真正代表唇宽度。为避免发音时伴随动作对唇形特征造成影响，本文在提取唇形参数时，将表示伴随动作的参数值归零，即去除了表示歪嘴和头部倾斜的伴随动作。将图 2（a）的唇形轮廓去除伴随动作后如

图 2（b）所示。

从图 2（a）中可以提取去除伴随动作之后的唇形轮廓几何特征，本文称之为唇形几何特征参数。实验所提取的唇形几何特征参数共有10个，分别为：1.外唇水平宽度：外唇两嘴角之间的水平距离，宽度为两倍的左侧外唇宽度（w1）。2.内唇水平宽度：内唇两嘴角之间的水平距离，宽度为两倍的左侧内唇宽度（w2）。3.外唇开口度：唇形外轮廓在垂直方向上的最大距离，高度为唇外轮廓开口度（h1）和下唇外轮廓开口度（h2）之和。4.内唇开口度：嘴唇内轮廓在垂直方向上的最大距离，上唇内轮廓开口度（h3）和下唇内轮廓开口度（h4）之和。5.外唇面积：嘴唇外轮廓所包含的区域内像素点的总数。6.内唇面积：嘴唇内轮廓所包含的区域内像素点的总数。7.外唇周长：嘴唇外轮廓线上的像素点总数。8.内唇周长：嘴唇内轮廓线上的像素点总数。9.外唇圆度：嘴唇外轮廓接近圆形的程度。10.内唇圆度：嘴唇内轮廓接近圆形的程度。

一些文献在假设唇形是一个椭圆的前提下，用椭圆的长短轴之比来表示唇圆度。但严格来说，唇形并不是椭圆，而是一个更加复杂的几何形状。如果仍然使用长短轴之比表示唇圆度的话就不够精确。比如正方形A和圆B，如果A的边长等于B的直径，则二者的宽高比相同，但它们的圆度完全不同。因此，本文采用数学上定义圆形度的方法来表示唇圆度（Baddeley et al. 2009）：

$$\text{圆形度} = 4\pi * \text{面积} / \text{周长的平方}$$

形状越圆，圆形度的取值越大，当唇形为绝对圆形时，圆形度达到最大值1。

具体提取唇形参数的步骤如下：首先由手工从中录像中提取可以代表元音音节发音时的嘴唇轮廓形状的关键帧；然后手工调整重构参数，通过目测使重构的唇形轮廓和关键帧的嘴唇内外轮廓线之间的误差最小；之后将代表头部倾斜程度的重构参数qx与歪嘴程度的重构参数s置零以消除伴随动作，最后再次重构唇形轮廓，并从中提取唇形几何特征参数。

2.3 实验方案

每个人的唇形宽窄厚薄都由于生理条件的差异而各有不相同，另外发音的动作习惯也都有所区别。因此在发音时，唇形轮廓的几何参数将具有个人生理特征。同一个发音人在发不同元音时，其唇形可以反映元音的部分特性，唇形特征之间的比较可以被认为是元音特征的比较，而不同发音人的唇形特征受个人生理的制约，不具可比性。

实验1：本文假设圆唇元音比非圆唇元音的唇宽更窄，唇形更圆，唇开口度更小，分别对8个发音人录像样本的元音关键帧进行统计分析。通过比较同一发音人的圆唇元音和非圆唇元音发音特征的唇形参数，以找出其中的一个或多个作为圆唇的区别特征。

实验2：本文通过统计张嘴与合嘴阶段内外唇参数的变化范围，来说明嘴唇发音动作在不同发音阶段有着不同的运动模式。并综合两个发音阶段的规律，提取出适合描述圆展唇动作动态过程的特征。

实验3：本文通过计算得到各个发音人的唇形参数的最小值和最大值，二者之差就是此发音人发音时唇形参数的变化范围。将此变化范围做归一化处理，可以消去不同发音人唇形参数的个人生理特征，但保留了个人的发音动作习惯。从归一化之后的唇形参数分布可以说明发音人的发音动作习惯是否具有共性。

3 实验结果及分析

每个发音人都有30个语音的唇形样本，其中包括9个圆唇元音样本和21个非圆唇元音样本。以外唇宽度为例，如果满足实验1的假设，每个圆唇元音样本的外唇宽度就会小于其它非圆唇元音样本的外唇宽度。9个圆唇元音的外唇宽度最多会出现189次满足实验1假设的情况。针对每个发音人，本文提取内外唇宽度、内外唇圆度和内外唇开口度等6个唇形几何参数，分别进行圆唇元音和非圆唇

元音间的相同参数的比较，统计出每个参数实际满足实验1假设的次数以及其占最大满足假设次数的百分比，结果如表2所示。

表2 圆唇元音和非圆唇元音唇形参数比较

	外唇宽度	内唇宽度	外唇圆度	内唇圆度	外唇开口度	内唇开口度
	W2＜W1	W4＜W3	R2＞R1	R4＞R3	H2＜H1	H4＜H3
发音人1	100%	98%	41%	32%	85%	96%
发音人2	100%	99%	81%	47%	74%	88%
发音人3	100%	95%	75%	80%	61%	61%
发音人4	99%	98%	54%	79%	63%	76%
发音人5	95%	98%	39%	66%	67%	90%
发音人6	99%	99%	89%	70%	46%	72%
发音人7	100%	99%	87%	76%	52%	83%
发音人8	100%	97%	44%	77%	81%	81%
平均	99%	98%	64%	66%	66%	81%

表2中W1、W3、R1、R3、H1和H3分别表示非圆唇元音的外唇水平宽度、内唇水平宽度、外唇圆度、内唇圆度、外唇开口度和内唇开口度；W2、W4、R2、R4、H2和H4分别表示圆唇元音的外唇水平宽度、内唇水平宽度、外唇圆度、内唇圆度、外唇开口度和内唇开口度。表格内的值为每个参数实际满足实验1假设的次数与最大满足假设次数的百分比，结果四舍五入取整数，计算公式为：

round（count（参数满足实验1假设次数）/189）*100%

从表2可得知，平均99%的样本，外唇宽度满足实验1的假设。平均98%的样本，内唇宽度满足假设。其它唇形参数中，只有内唇开口度达到81%的样本满足实验1的假设，而内外唇圆度和外唇开口度满足实验1假设的比例都不到70%。实验1结果表明外唇宽度和内唇宽度都适合作为圆唇的区别特征，而内外唇圆度和开口度无法有效区分圆唇元音和非圆唇元音。

为描述圆唇运动的动态过程，本文将完整发音过程分为闭嘴与张嘴两个阶段。闭嘴阶段包括发音的准备动作阶段和发音完成后的复位阶段，此时内唇宽度、圆度和开口度的值都恒为0。而由于逆向协同发音的作用，外唇宽度在未发音之前就开始逐渐变窄，它

在闭嘴阶段能有效描述圆唇运动的发音动作。张嘴阶段主要是有声段，所有作为唇形样本的关键帧均来自于有声段。实验2分别对每个发音人圆唇元音和非圆唇元音内外唇宽度的变化范围进行统计平均，结果如表3所示。由表可知圆展唇元音内唇宽度的平均变化范围远大于外唇宽度的平均变化范围。另外，有声段语音的声学特性和内唇的开口形状更加相关。因此，在张嘴阶段内唇宽度的变化更适合表示圆唇运动。综合考虑发音的两个阶段，本文提出使用内外唇宽度之和为特征，其变化可以良好的描写圆唇运动的动态过程。

表3 发音人圆唇元音及非圆唇元音内外唇宽度的平均变化范围（单位：像素）

发音人	1号	2号	3号	4号	5号	6号	7号	8号
内唇宽度平均变化范围	37	52	40	52	41	43	49	39
外唇宽度平均变化范围	11	17	13	17	8	17	23	9

个人生理特征和发音动作习惯的不同造成了不同发音人的唇形参数不能直接相互比较。根据实验1和实验2的结果，实验3使用内外唇宽度之和为特征，对其变化范围进行归一化处理，以消除不同发音人嘴唇宽窄各有不同的个人生理特征，并保留了个人的发音动作习惯。归一化之后，内外唇宽度之和的分布如图 3（a）所示。对所有发音人的元音样本唇宽参数取均值，其分布如图 3（b）所示。

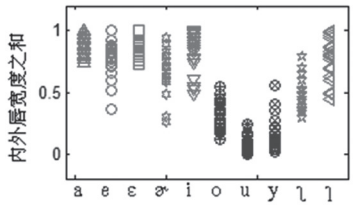


图 3（a） 所有发音人元音样本

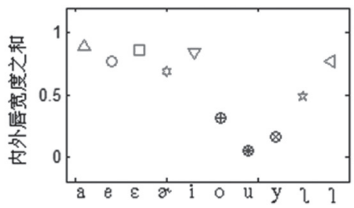


图 3（b） 所有发音人元音样本均值

从图 3（a）可知每一个元音的内外唇宽度之和变化范围大，说明不同发音人的发音习惯具有一定的随意性，有人发音动作幅度较大，有人发音动作幅度较小。而从图 3（b）可知，虽然发音动作

具有随意性，但总体上保持非圆唇元音的内外唇宽度之和大于圆唇元音。事实上，从每个发音人的样本也基本上得到了相同的规律。

4 讨论

本文提出圆唇的本质是唇宽变窄，使用内外唇宽度之和为特征可以描述圆展唇的运动过程，发音动作幅度有着较大的随意性，但同一发音人总会保持圆唇元音的唇宽小于非圆唇元音的唇宽。

一些文献使用唇突度为圆唇的特征，本文不对唇突度进行讨论有如下原因：1.本文希望找到一种简单可计算的圆唇的特征。使用录像机采集侧面人脸提取唇突度的方法有着较高的技术难度。其他三维数据采集设备虽然可以实时采集唇突度，但价格昂贵，数据预处理复杂，距大范围应用还有一定距离。2.本文提出以内外唇宽度之和为特征已经可以良好的描述圆展唇运动。3.鲍怀翘的研究已经证明唇突度对于区分元音圆展作用不大。

由于提取唇形关键帧仍没有一定的标准，因此本文所有的数据及推论都是建立在我们对关键帧的提取标准之上，即元音共振峰稳定段中，发音动作保持基本不变时的某一帧。在唇形录像样本中提取关键帧的位置不同，会影响采集到的唇形几何特征，导致实验结果发生变化。而事实上，本实验中如果没有事先作保持发音动作一段时间的约定，从实际数据中我们会发现在元音共振峰的稳定段中，发音动作仍然处于变化之中。

王志明、蔡莲红（2002）使用算法来自动提取唇形关键帧，可以实现大量数据的快速处理，本文的方法则是对语音的共振峰和录像数据进行人工判断，由此决定关键帧的位置。王志明的方法好处在于提取的所有关键帧都是按相同的标准所得。本文的方法优点在于更多的利用了语音学的知识，缺点是同时也具有更多的主观性，在唇形相似的相邻帧中，选择关键帧往往具有随意性。另外，手工提取关键帧速度较慢，无法做到大量快速提取。

关于圆唇特征定义学界并没有取得一个统一观点，目前仍没有标准定义，这造成研究成果之间不具备相互比较的基础。本文对6种唇形参数进行统计分析后，分别指出其单独做为圆唇特征的不足之处，并提出使用内外唇宽度之和作为圆唇特征，可以区分圆展唇元音，其变化能描述动态的圆展唇发音动作。

5 总结

语音学还是一门发展中的学科，很多术语的概念也没有形成多数人接受的共识，对于圆唇的定义也是如此。本文通过实验采集各种唇形参数，利用统计的方法，提出圆唇的本质是唇宽度的变窄，而非嘴唇的形状变圆。在生理上，发音是一个动态过程，本文提出内外唇宽度之和的变化可以描写圆展唇发音动作的变化。此外，根据实验结果，虽然普通话发音时，唇部发音动作具有较强的随意性，但基本上总能保持同一发音人圆唇元音比非圆唇元音的唇宽度更窄这一特性。

参考文献

北京大学中文系现代汉语教研室 （2009）《现代汉语》（重排本），商务印书馆，北京。

胡 方 （2007） 论宁波方言和苏州方言前高元音的区别特征——兼谈高元音继续高化现象,《中国语文》第 5 期，商务印书馆，北京，455—465 页。

林焘、王理嘉 （1992）《语音学教程》，北京大学出版社，北京。

罗常培、王均 （2002）《普通语音学纲要》，商务印书馆，北京。

王志明、蔡莲红 （2002） 汉语语音视位的研究,《应用声学》第 3 期，科学出版社，北京，29—34 页。

吴宗济、林茂灿 （1989）《实验语音学概要》，高等教育出版社，北京。

Abry, Boë （1986） “Laws” for lips. *Speech Communication*, 5/1, 97-104.

Baddeley, Jayasinghe, Lam, Rossberger, Cannell & Soeller （2009） Optical single-channel resolution imaging of the ryanodine receptor distribution in rat cardiac myocytes. *Proceedings of the National Academy of Sciences*, 106/52, 22275-80.

Chomsky, Halle （1968） *The sound pattern of English*. New York: Harper & Row.
Hardcastle, Hewlett （1999） *Coarticulation: Theory, Data and Techniques*. Cambridge, Cambridge University Press.
Ladefoged （1975） *A course in phonetics*. New York: Harcourt Brace Jovanovich.
Ladefoged, Maddieson （1990） Vowels of the world's languages. *Journal of Phonetics*, 18/1, 93-122.
Liew, Leung & Lau （2000） Lip contour extraction using a deformable model. *Image Processing 2000*, 2, 255-258.
Lindau （1978） Vowel features. *Language*, 54/3, 541-563.

（潘晓声：上海，上海师范大学信息与机电学院计算机系 itol_xs@shnu. edu. cn；
孔江平：北京，北京大学中文系 / 中国语言学研究中心 jpkong@pku. edu. cn）

安多藏语塞音的 VAT 研究

桑塔　姚云　兰正群

摘　要　VAT（Vocal Attack Time）是指声带开始抖动到声带接触的时间，主要分析嗓音起始端的特征，是通过 SP 和 EGG 两路信号的时间差计算。本文考察了 VAT 与藏语安多语的塞音之间的相关性。最后的数据显示，清送气多半是正值；清不送的声带开始振动和到声带接触几乎是同时的。浊音的 VAT 多半为负值，且其值较大，这说明在嗓音的起始端声带有一个长时间的闭合。同时，浊音的一部分 VAT 与清送气重合，这一点有可能是许多语言或方言塞音清化的一个生理基础。

关键词　嗓音，发声类型，VAT，藏语，塞音

VAT Measurement of Amdo Tibetan Plosives.

Sangta, Yao Yun, Lan Zhengqun, Phonetics Lab, Peking University, Beijing, China, 100871

Abstract Vocal attack time(VAT) is the time lag between the growth of the sound pressure signal and the development of the physical contact of vocal folds at the vocal initiation, which can be calculated with the sound pressure(SP) and electroglottograph(EGG) parameters. By measuring the relation between VAT and Amdo Tibetan plosives, the study reveals that most voiceless aspirated sounds have positive values, as vocal folds start vibrating at the same time of their physical contact. While voiced sounds have negative and large values, indicating the long closure of vocal folds at the beginning of the voice. In the meantime, part VAT of the voiced sound is overlapped with the voiceless aspirated token, which could be a physiological trigger of devoicing in certain languages or dialects.

Key words Voice, phonation, VAT, Tibetan, plosives

1. 引言

1.1 VAT 的提出

在病因学里，嗓音的起始特征是一项重要研究的内容。一般情况下嗓音的起始特征分为两个阶段：1）发声调整阶段，包括肌肉一定的紧张度、闭合以及气压方面的调整；2）接触阶段，包括声带抖动的起始和声音的产生（Robert et al 2007）。其中第二个阶段是嗓音起始特征重要的内容。为了比前人更加客观地观察和研究嗓音起始的状态，嗓音医生 R. J. Baken 和他的同事（2007）通过同步的声音信号（SP）和声门阻抗信号（EGG）提出了一套计算 VAT（Voice Attack Time）的非侵入型的方法。所谓的 VAT 就是声带最初开始抖动到两片声带接触的时间间隔，一般以毫秒为单位。这里所谓的抖动是指声带并没有完全接触之前的波动。VAT 的值可以是正的，也可以是负的或者是零。在他们通过对美国人 VAT 的测量后，把嗓音的起始状态分为三种，即软启动、正常启动和硬启动。

1.2 VAT 的研究

这种方法提出以后，人们从不同的角度和目的进行了一些研究。Rick M. Roark（2010）等人测量了健康的年轻人最自然的音高和音强状态下的 VAT 后，发现性别间的 VAT 有显著的差异，即女性的 VAT 要比男性的小，并且 25—29 岁男性

的 VAT 最大。Ben C. Watson 等人(2013)研究了声带振动的频率和 VAT 之间的关联后，发现高频条件下的 VAT 值要比中频和低频的要小，中频和低频之间的 VAT 值没有显著差异。

在语言学研究方面，Estella P.^-^M (2011) 等人观测了粤语中声调对于 VAT 的影响。发现了 VAT 在性别间的差异，即女性的 VAT 值要比男性的小，这与 Rick M. Roark et al (2010) 的结论相吻合。在音位层面上，粤语的两组调类有显著的不同，即三个平声调 (level tone) 的 VAT 要比和三个轮廓调 (contour tone) 的要小。在汉语的研究方面，潘晓声和孔江平 (2008) 研究了汉语普通话零声母的嗓音起始状态和 VAT，发现普通话零声母音节的中低元音的 VAT 大多小于 2，而高元音起始的零声母的 VAT 大多大于等于 2。这说明中低元音在嗓音起始时声带紧张一点，而高元音相对较松弛一点。

上述这些 VAT 方面的研究没有涉及到发音方法对 VAT 的影响。藏语安多方言塞音根据不同的发音方法可以分为送气、不送气和浊音。本文主要去探讨这三组辅音的 VAT 及其相关的问题。

2. 实验方法

2.1 词表

根据塞音的发音方法，本文的实验词表由五组单音节词组成，具体是清不送气、清送气和浊音，其中不送气可以分为单辅音和复辅音，复辅音指主要辅音加前置辅音/h－/。浊塞音也可以分为单辅音和复辅音，其中复辅音指主要辅音加前置辅音/n－/。本文将要论述的复辅音的 VAT 指的是复辅音中主要辅音的 VAT。所以，根据这些特征可以把塞音分为单浊塞音、复浊塞音、单送气塞音、单不送气塞音、复不送气塞音。每一组塞音大概选了 40 个左右的词，由于数量较多，所有的词项不在此一一列举。同时，这些塞音在发音部位上可以分为三种，分别是双唇、齿龈和舌根（具体见表 1 所示）

表 1　　实验词表的分类

发音方法		发音部位			例词
		双唇	齿龈	舌根	
清	不送气(单)	p	t	k	[ta] 现在
	不送气(复)	^{h}p	^{h}t	^{h}k	[^{g}tax] 老虎
	送气	p^{h}	t^{h}	k^{h}	[p^{h}o] 男性
浊	单	b	d	g	[ga] 高兴
	复	^{n}b	^{n}d	^{n}g	[^{n}ba] 羊叫

2.2 信号采集

本次测量的数据来源于青海省同仁县的 29 岁的安多藏语母语者，并且没有(过)任何言语产生、感知方面的障碍。语音信号 (SP) 是用 SONY ECM－44B 话筒录制，采样频率为 44kHZ，采样精度为 16 位；声门阻抗信号 (EGG) 是由 Kay 公司制造的 Real-Time EGG Analysis (型号 5138) 采集。

2.3 步骤

首先，同步的语音信号和 EGG 信号在北京大学语言实验室录制。然后通过 SP 和 EGG 的时间差，用全自动的方法计算 VAT 数据 (Robert F. Orlikoff et al, 2007)。同时计算出 FOM (Figure of Merit) 来检测其 VAT 的有效性。FOM 值越接近于 1 说明 VAT 越可靠，低于 0.75 的 VAT 一般被视为无效 (Rick M. Roark et al 2010)。FOM 值不会大于 1。最后观察有效地 VAT 值和不同塞音之间的关系。

3. 结果与分析

表 2 列出了五组词的 VAT 值（包括

均值、范围以及中位数）、FOM值（包括均值和中位数）以及基频（包括均值和范围）。

3.1 复辅音的VAT分析

不送气塞音和浊塞音分别有单辅音和复辅音的对立。不送气塞音的复辅音由主要辅音带前置辅音/h/组成。通过数据可以看出，包括复辅音在内的所有的不送气塞音的FOM的均值是0.99，中位数是0.98，说明其VAT值是很可靠的。

表2　安多藏语塞音的VAT值、FOM值和基频数据

		清			浊	
		送气	不送气（单）	不送气（复）	单辅音	复辅音
词项数		46	63	83	63	59
VAT	均值（SD）	7.92（6.52）	−0.75（2.11）	0.5（2.2）	−82.05（79.55）	−66.11（78.52）
	范围	−1.68 to 68.7 30.11	−6.49 to 2.54	−3.83 to 5.28	−174.79 to 17.53	−179.96 to 12.36
	中位数	5.4	−0.7	0.62	−127.075	3.65
FOM	均值（SD）	0.99（0.02）	0.99（0.01）	0.99（0.02）	0.83（0.09）	0.84（0.08）
	中位数	0.99	0.97	0.98	0.84	0.83
F0	均值（SD）	165（8.04）	170（16）	175（14）	112（13）	122（15）
	中位数	167	172	178	107.81	117.99

单辅音的VAT的范围在−6.49到2.54毫秒之间，而复辅音的VAT在−3.83到5.28之间。虽然它们彼此的范围有所偏移，但是它们的中位数都是接近于0的。因此，不送气的单辅音和复辅音之间是没有显著差异的。这一点通过图1可以看得更直观。不送气的单辅音和复辅音的VAT是相互重合的，说明不送气塞音的前置辅音/h/对于VAT是没有影响的。

浊塞音的复辅音可以带一个前置鼻音/n/，其单辅音的VAT的范围在−174.79到17.53毫秒之间，复辅音分布在−179.96到12.36毫秒之间。很明显它们的VAT也是完全重合的。这也跟前面的情况是相同的，说明浊塞音的前缀/n/同样对其VAT也没有任何影响。因此，当我们在下文讨论清不送气塞音和浊塞音的VAT时我们是包括单辅音和复辅音的。单辅音和复辅音之间的VAT降不加以区分。

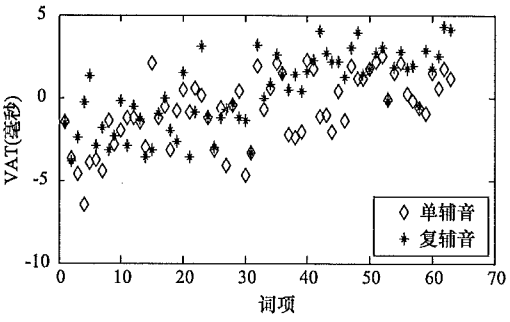


图1　不送气塞音的VAT

3.2 浊塞音的VAT分析

现在我们来看看浊塞音情况。浊塞音的平均FOM值都比清音的FOM要小的多。122个浊塞音的20%的FOM值都低于0.75，所以只使用了FOM值在0.75以上的VAT。被使用的FOM的均值和中位数也低于清塞音的FOM。具体见图2所示。

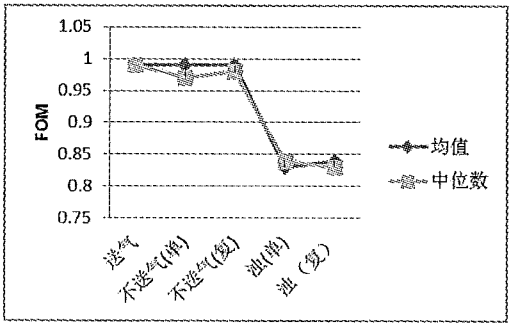


图2 不同塞音VAT的FOM值

FOM值越小，说明其VAT的有效性相对较差。同时，浊塞音的VAT比清塞音的要复杂。这种复杂性主要表现在两个方面。一方面是它范围跨度很大，大概在－179到17.53毫秒之间（见图3所示）。

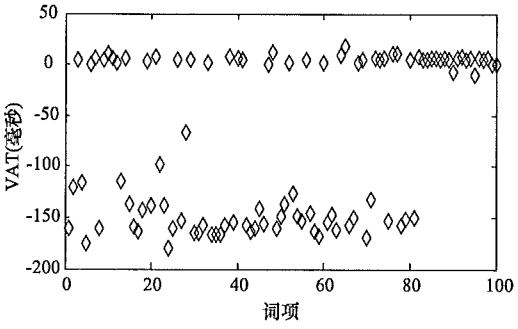


图3 浊塞音的VAT分布

另一方面是它的负值都很大，几乎都在－100毫秒以下。但有意思的一点是它集中在两个区域：一个是在－150毫秒上下；另外一部分正好与清送气塞音的VAT相互重合（见图4所示），而没有与不送气塞音的VAT重合，这部分的VAT值比较大。

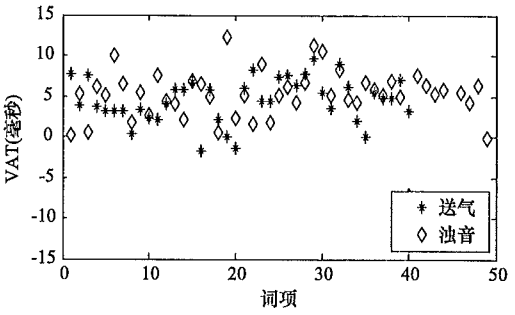


图4 送气塞音和一部分浊塞音的VAT

3.3 清塞音的VAT分析

送气塞音的VAT值在－1.7毫秒到68.7毫秒之间。由于30毫秒以上分布的数据较少，所以可以剔除这些分布在外围的数据，最后得到了一个更为客观的范围：－1.68毫秒到30.11毫秒之间。送气塞音的FOM的均值是0.99，其标准差是0.02，说明其VAT值是非常有效的。其中87%的VAT都分布在10毫秒以内。用这些送气塞音的VAT和不送气塞音的VAT相对比时，它们之间除了少数的VAT有所重合以外，大多数VAT之间还是泾渭分明的。送气塞音的VAT值比不送塞音的要大（具体见图5所示）。

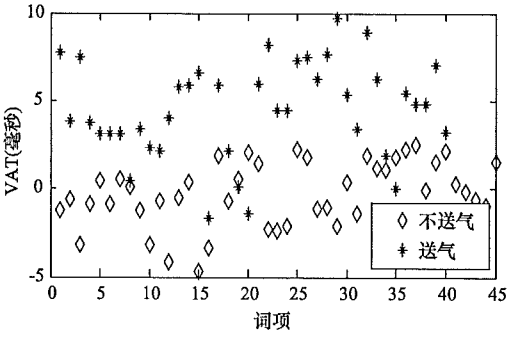


图5 不送气塞音和送气塞音的VAT对比

我们可以根据R.J.Baken等人(2007)提出的嗓音的三种不同的启动类型来作为坐标看一下藏语塞音的VAT的格局。他们提出的软启动（'breathy' on-

set）的VAT的范围在7.6到38.0毫秒之间；正常启动（‘comfortable’ onset）的VAT在－1.4到9.6毫秒之间；而硬启动（‘hard’ onset）则是负的，在－9.5到－1.7毫秒之间。图6是这三类VAT的均值。

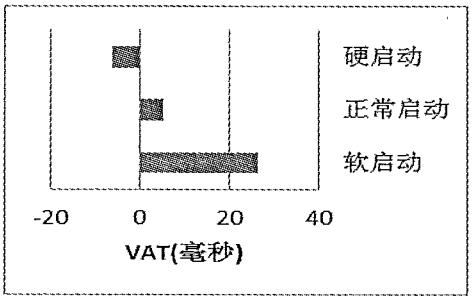


图6 基于VAT的三类嗓音启动类型
(R. J. Baken et al 2007)

藏语塞音的三种VAT并不是很好地与这三种启动方式相呼应。整体上藏语的VAT都偏小。藏语的浊音的VAT的均值都比硬启动的要小的多。不送气清音的VAT几乎都分布在零的上下，比正常启动的值也要小。送气的VAT的均值也比软启动要小。造成这种差异的主要原因是因为R. J. Baken等人提出的三分法主要的基于纯粹的语音学实验，是让发音人发语义上没有意义的/a/。而在藏语中，这些词都是有意义的语素。因此，影响这些VAT变量是很多的，有个人的，、语言的以及生理的因素。所以不易于做深入的比较。这里比较的目的仅仅用这三种启动方式作为参照来观测一下藏语塞音的VAT。

本次试验中，基频也可能是影响VAT的因素之一。送气的塞音比不送气的塞音的基频要低，大概相差5赫兹。而送气塞音的VAT比不送气塞音的VAT要高。这与Ben C. Watson et al（2013）的研究是相吻合的，基频和VAT在一定的范围内是反比关系。浊塞音的基频和VAT之间的关系比较复杂，不易于直接对比，需要用很多维度的数据来比较。基频和VAT之间的关系方面有孔江平和张锐峰也在做深入的研究（发表中）。

4. 讨论

4.1 VAT的生理机制

VAT在送气和不送气之间的差异在生理上是完全可以解释的。送气的VAT的值都比不送气的要大，这说明声带在接触之前有一个较长的扰动时间。造成这种情况的主要原因应该是气流。在声带接触之前需要一股较强的气流从声门出去。这股气流扰动了声带边缘。然而在发不送气塞音时就没有这样强的气流可以使声带在接触之前被扰动。所以造成了这种显著的VAT差异。

通过上面的数据发现，浊音的VAT值表现得比较复杂。通过图3可以看出，浊音的VAT分布在两个区域。正值的VAT都和送气的重合，它的值要比不送气的VAT值要大。造成这种情况最主要的一个原因是因为藏语中的浊塞音在口腔除阻之前就开始振动，其VOT（Voice Onset Time）是负的。引起声带振动主要有两个方面的因素，一个是气流。另外一个是声带自身的调整，发浊音时的气流肯定没有送气时的气流强。那么引起这种VAT值的主要原因是由于声带自身的调整，即声带要调整到一个相对于清音较松弛的状态以便于在除阻之前振动。而清塞音不需要这种调整，因为它在口腔除阻之后声带才开始振动。同时，在提取浊塞音的开商时，我们发现元音之前塞音段的开商是比较大的。这说明声带更加气化一点。气化嗓音的VAT就是R. J. Baken等人所说的软启动嗓音，软启动嗓音的VAT都比较大的。因此，相对松弛的声带应该是造成这类浊塞音的VAT的格局的主要原因。

另外一部分浊塞音的VAT是负值。从图3中可以看出，这些负的VAT都分布在－150的左右。负的VAT说明声带在开始正常振动之前是有一个较长时间的闭合。EGG信号在SP信号之前就产生了。首先，声带振动的一个重要条件是声门上下有一个气压差，一般至少是2cm水柱的压力差。这些气流才能穿过声门使声

带振动。发浊塞音正好对这种压力差产生了一个阻力。因为发浊塞音时，口腔内的阻塞对气流形成阻力使声门上压加大。为了减少声门上压力，喉头只能下移来制造一个压力差。喉头的这种下移应该是造成声带长时闭合的一个原因。因为它需要做一个较复杂的调整来发浊塞音。浊塞音改变发声类型（Titze，2000）这一点也证明声带在发浊塞音时的这一特征。总之，发浊塞音时，声带在发声前和发声中有一个比清音较复杂的活动。

4.2 VAT在音变方面的意义

通常情况下，我们在考察塞音的声学表现时，一般从塞音的VOT、闭塞段时长以及气流气压等参数来研究的。通过本次测量，我们也发现塞音起始的VAT也是一项重要的参数。VAT的特征在不同的语言中可能表现出不同的模式。

浊塞音的VAT值的二分引起了我们的注意。主要是浊塞音的一部分VAT值与送气塞音的VAT相吻合。虽然引起这样的VAT值的生理机制在两者之间有可能不一样，但这一特征完全有可能成为共同的语音演变的生理基础。比如说，中古汉语的浊塞音的一部分在普通话中变为清送气塞音；有些汉语方言中的古代浊音甚至全部变为送气音；还有藏语拉萨话中的古藏语单辅音的浊塞音也变成同部位的清送气音（见表3所示）。

表3 藏语拉萨话中的来源于古浊塞音的送气清塞音示例

藏文	<ga>	<da>	<ba>
拉萨	/k^ha12/	/t^ha12/	/p^ha12/
汉译	‘哪’	‘现在’	‘牛’

因此，我们可以提出大胆的假设。在没有其它条件制约的情况下，由于浊塞音和送气塞音的VAT相同，浊塞音在清化的过程中受到其VAT的影响而变成清送气。也就是说，浊塞音的VAT起初仅仅是一个很微观的伴随特征，之后，在音变的过程中，浊塞音保留了这一嗓音起始特征，这种起始特征又反过来影响了嗓音起始之后的音变轨迹，即浊塞音的VAT使其演变为送气塞音。当然，引起语音演变的因素是非常复杂的、多元的。这里提出的VAT的音变机制是其可能性之一，它应该属于机械性的演变，在没有其它音变条件（如系统、社会、接触等）制约的情况下的一种音变。

5. 结语

总之，通过这些塞音的VAT的值，我们可以把安多藏语塞音的嗓音起始类型分为三种，送气、不送气与浊音。送气的VAT是最大的，不送气VAT则在0的附近，而浊音的一部分很整齐地和送气的VAT相重合，另一部分则是负的，且值也比较大（见图7所示）。复辅音的前缀对VAT没有影响。

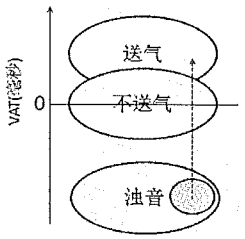


图7 安多藏语塞音VAT的格局

VAT目前主要应用于嗓音医学领域，在语言学领域研究的还比较少。通过对安多藏语塞音VAT的研究，我们发现在研究塞音方面我们又多了一个视角。在语言学领域，VAT的研究也可以给我很多启发，譬如上面讨论的音变方面的启发。还有声调的研究方面，除了基频、开商和速度商这几个维度的物理参数之外，调类之间在VAT方面也可能呈现出不同的模式，值得我们做研究。还有不同的语言的不同的发音方法之间的VAT的格局可能是不同的，这也值得我们继续做这方面的研究。

中国语音学报 第5辑，2014年，北京

6. 致谢

本研究得到了中国哲学社会科学重大投标项目“中国有声语言及口传文化保护与传承的数字化方法研究和基础理论研究”（项目批准号：10&ZD125）的经费支持。在研究和写作过程中得到了导师孔江平教授和著名嗓音学家 Ron. J. Baken 先生的指导。

7. 参考文献

[1] Ben C. Watson, R, J. Baken, Rick M. Roark, Stephanie Reid, Melissa Riberio and Weiln Tsai, “Effect of Fundamental Frequency at Voice Onset on Vocal Attack Time,” *J. Voice*, Vol. 27, 2013, pp. 273－277.

[2] Estella P. ma, R. J. Baken, Rick M. Roark and p.－M. LI, “Effect of Tones on Vocal Attack Times in Cantonese speakers,” *J. Voice*, Vol. 10, 2011, pp. 1－6.

[3] Ingo R. Titze, *Principles of Voice Production*, Lowa City: National city for Voice and Speech, 2000, p. 321.

[4] Huakan, *Zangyu Anduo Fangyan Cihui* [*A Vocabulary of AmdoTibetan*], Lanzhou: Gansu Minzu Press, 2002, pp. 40－275.

[5] Orlikoff RF, Deliyski D. D, Baken RJ, Watson BC, “Validation of a glottographic measure of vocal attack,” *J. Voice*, Vol. 23, 2009, pp. 164－168.

[6] Roark RM, Watson BC, Baken RJ, Brown DJ and Thomas JM, “Measures of Vocal Attack time for healthy young adult,” *J. Voice*, Vol. 33, 2011, pp. 1－6.

[7] Roark RM, Watson BC, Baken RJ, “A figure of merit for vocal attack time measurement,” *J. Voice*, Vol. 26, 2012, pp. 8－11.

[8] 孔江平：《论语言发声》，中央民族大学出版社 2001 年版，第 23 页。

[9] 袁家骅：《汉语方言概要》，文字改革出版社 1983 年版。

[10] 石峰、冉启斌：《塞音的声学格局分析》，载《现代语音学前沿文集》，商务印书馆 2009 年版，第 46 页。

[11] 潘晓声、孔江平：《VAT 和汉语普通话零声母嗓音起始状态研究》，载《第八届中国语音学学术会议暨庆贺吴宗济先生百岁华诞语音科学前沿问题国际研讨会论文集》，中国社会科学院语言所 2008 年版，第 5 页。

桑 塔 北京大学中文系语言学实验室 100871
姚 云 北京大学中文系语言学实验室 100871
兰正群 上海师范大学 E 语言研究所 200234

方言 2014 年第 3 期 206—214 页(2014 年 8 月 24 日出版于北京)

河南禹州方言声调的声学及感知研究*

张锐锋[1]　孔江平[1,2]

(1,北京大学中文系 北京　ruifeng_72@163.com;

2,北京大学中国语言学研究中心 北京　jpkong@pku.edu.cn)

提要　本文对中原官话禹州方言的声调进行了实验研究,用声学方法确定了该方言四声的调值,并通过感知听辨发现,在禹州方言声调感知中基频模式和发声类型都起作用,前者起主要作用,后者起补偿作用。在基频区别力弱时,发声对感知的作用较大,在基频区别力强时,发声对感知的贡献就变小。

关键词　声调　范畴感知　挤喉音　发声　基频　开商　速度商

壹　引言

1.1 禹州市原为禹县,隶属于河南省许昌市,位于郑州市南偏西约一百公里处,北临登封、新密,南接郏县、襄城,西面是汝州,东面是长葛、许昌。河南省的方言可分为两类:一为晋语,包括安阳等 19 个县市,有入声;一为中原官话,包括郑州、开封等 111 个县市,无入声。禹州方言属于中原官话。邵文杰等(1995:3-6)把河南境内的中原官话分为郑汴、洛嵩、蔡汝、信潢和陕灵五片,禹县归郑汴片。贺巍(2005:136-140)分中原官话为八片,河南境内有郑开、洛嵩、南鲁、漯项、商阜、信蚌、兖荷七片,禹州归南鲁片。

张启焕等(1993:53)、李淑娟(2004:7)和屈颜平(2010:4)都记录禹县话的四个声调为:阴平[˩]24、阳平[˧˩]42、上声[˥]55、去声[˩˧]31。

1.2 传统的方言调查主要通过调查者听辨并记录方言的音位系统,包括声调的调位。然而,受母语声调感知范畴、声调斜率和声调时长等的影响,有时人们不能准确记录声调的调值,因此,将实验语音学手段用于方言调查,目的是弥补听感的局限性,帮助得到准确的音位系统。语音实验包括声学分析和感知听辨两个方面,声学分析是纯物理性的,感知听辨则关系具体的音位系统。本文就是对禹州声调作这两方面的实验。

本文的材料是作者于 2012 年 5 月至 2013 年 5 月间多次实地调查所得。

贰　禹州方言四声调值的声学研究

2.1 本研究使用的字表包含了 6 组 24 个字:[tɑ]搭达打大 | [ti]低敌抵地 | [tu]督毒赌肚 | [tʂʰu]出除础处 | [tʰu]秃图土兔 | [tʂʅ]支直纸志。

每组的字声韵母相同,声调则分别为阴平、阳平、上声、去声。发音人包括两男两女:男 A,46 岁,来自禹州县城,乡村教师,只会说禹州方言,不会讲普通话,录音完成于北京大学中文系录音室;男 B,23 岁,来自禹州市小吕乡;女 A,22 岁,女 B,21 岁,均来自禹州市苌庄乡。他们

* 本文得到国家社会科学基金重大招标项目资助,项目号:10&ZD125。初稿曾提交"2013 汉语方言类型研讨会"(2013 年 8 月,北京)。感谢《方言》匿名评审专家给本文提出的宝贵意见。

也都讲地道的禹州方言,录音完成于河南师大院内一个安静的房间里。要求发音人尽量在同一个强度上读字表,每个字念三遍,共得到288个声音样本,每个调类有样本72个。所有声音样本均符合声学分析要求。

2.2 在发音人所读的去声字末尾都出现了比较特殊的发声现象。图1是用Praat的常规设置所做的女A的一个去声字样本"大"[tɑ]的波形图、语图和基频线。

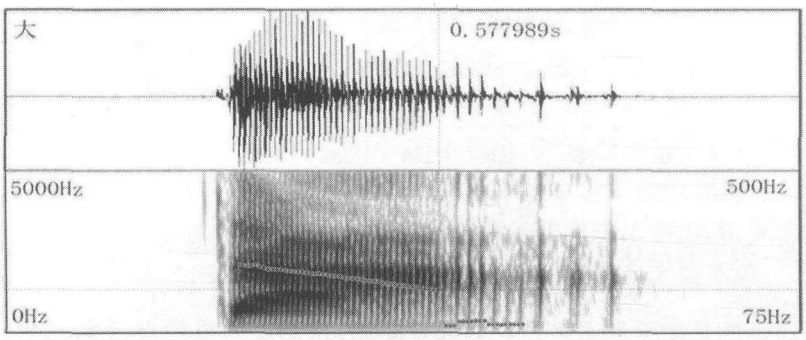


图1 发音人女A"大"字的波形图、语图和基频线

从波形图可见,在0.577989秒之前,声波振幅很大,脉冲间距离规则而均匀地小幅增大,显示为匀速的降调。而在该时间点之后,振幅急剧变小,脉冲骤然变得稀疏。语图上此时点之前的声门脉冲间隔也是均匀规则地变宽,其后突然变得稀疏而模糊。相应地,基频线在该时点断裂,消失几毫秒之后,在很低的75Hz附近重新出现。听感上,在0.577989秒之前显得清晰而饱满,此后则明显带有吱嘎声。这是一种特殊的发声类型(phonation type)——挤喉音(creaky voice。朱晓农2010:93-94称为"嘎裂声")。在有的声调语言中,如汉语、藏语等,有些人在发低调(如普通话上声)时常常使用挤喉音(孔江平2001:179),其特点是基频低且不稳定,开商最小(Edwin M-L.Yiu 2013:174)。在禹州方言中,阳平和去声均为降调,去声的末尾压到最低时使用了挤喉音。在praat中,挤喉音的基频有时不易显现,可以通过调节advanced pitch settings中的voicing threshold等设置来提取挤喉音的基频值。例如上述音节,把voicing threshold由0.45改为0.1,可在基频线中断处得到一个F0值123.83Hz。我们从发音人女A的18个去声样本提取出36个挤喉音基频,它们的最大值为132.19Hz,最小值为50.94Hz,均值为85.82Hz,标准差为19.66Hz。

2.3 本文用praat软件提取基频值。声调的起点从语图上元音的第二个脉冲算起;升调的终点定在基频峰点处;一般降调的终点是在宽带图上的基频直条有规律成比例的间隔结束处(朱晓农2010:281-282);音节末尾出现挤喉音的样本,通过调整praat的advanced pitch settings来提取挤喉音段的基频值,每个样本提取两点。从每个声音样本等间隔地提取11个基频值,代表基频线的走向。然后,将每位发音人的72个声音样本按四个调类分别作基频值平均,并按以下公式求出每位发音人在自己调域内的五度值:

$$T=\{[lgx-lg\ (min-SDmin)]/[lg\ (max+SDmax)-lg\ (min-SDmin)]\}\times 5$$

式子中的x代表测量点的频率均值,min-SDmin是各测量点平均值中的最小值减去该点全部数据的标准差,max+SDmax是各测量点平均值中的最大值加上该点全部数据的标准差。取常用对数是为了让求出的五度值跟人的听感吻合(石锋2010:1-14)。把四位发音人的数据求平均数,即得到禹州方言的声调图,见下页图2。其中的声调曲线排除了性别之间及性别内部个体之间的差异。各声调在音长上不存在显著性差别:阴平平均为285.08毫秒,阳平为

249.31 毫秒，上声为 275.33 毫秒，去声为 236.56 毫秒。

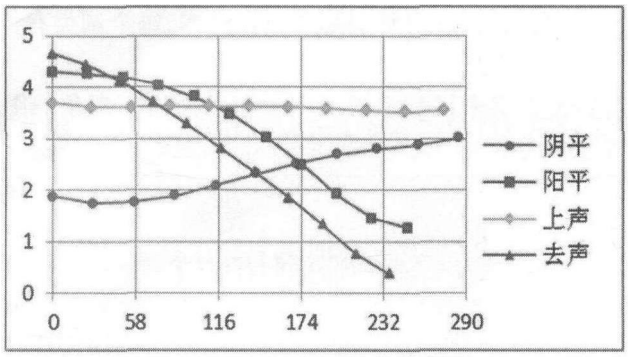


图 2　禹州方言的声调

从图 2 可见，阴平起点在 2 度音高区间的上部，经过一段小幅下倾后，又升至 3 度区间，可定为[223]。上声是一个平调，位于 4 度区间之中部稍上处，是[44]。

需要特别讨论的是阳平和去声。它们都是降调，从图 2 可见两点区别：第一，去声从 5 度区间的上部斜直滑向 1 度区间的中下部，是典型的[51]调，而阳平开始于 5 度区间的中点稍下处，终点落在 2 度区间的中下部，去声的跨幅明显大于阳平；第二，阳平起始后，并没有直直地滑向终点，而是先平缓下降，经过了一段弧度后，才滑向终点，可记为[542]。

叁　禹州声调的合成和感知实验

3.1 禹州方言的四个声调可形成 6 种对立形式：阴平-阳平、阴平-上声、阴平-去声、阳平-上声、阳平-去声、上声-去声。由于一个声调的感知范畴并不是固定的，它取决于声调系统中所对立的声调（孔江平 1995:56-64），因此有必要将这 6 种组合（对立组）全部进行合成，以探索禹州方言四声的感知全貌。

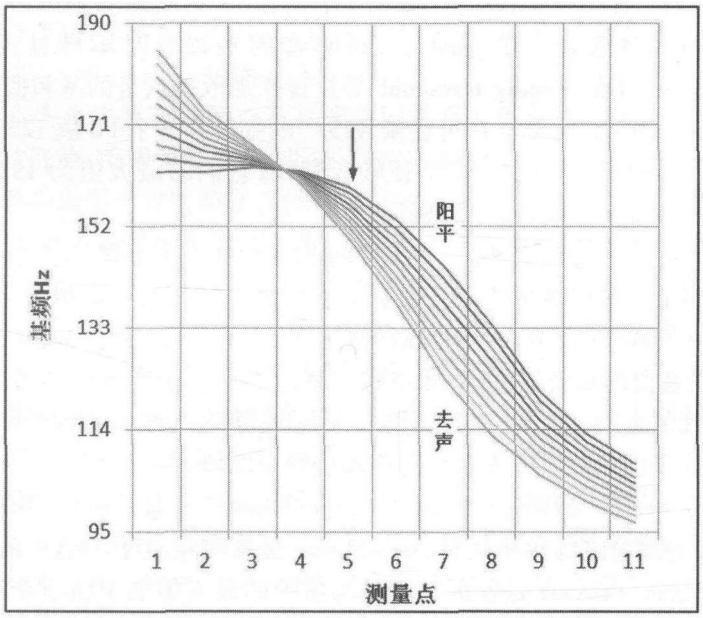


图 3　阳平-去声对立组的合成设计

图 3 所示是为阳平与去声对立组的感知实验做的合成工作：用 praat 的基音同步叠加法（PSOLA）将一对声调中的第一个样本阳平“达”[ta^{542}]，从原声均匀分九步改变基频逐步合成

到[51],得到10个样本,其第一个样本是"达"的原声,后面9个由改变其基频而得来;再利用第二个调去声"大"[tɑ^{51}],从原声均匀分九步合成为[542],所得亦10个样本,此时第一个样本是"大"的原声,后面9个由改变其基频而得来。一来一去用的是同一套基频值,所以纯从基频上说,前10个样本和后10个样本是相同的。如是操作,每个对立组可得到20个合成样本,6个对立组共得样本120个。各样本的声韵母结构均为[tɑ]。

3.2 听辨方法一般有两种,一种是辨认测验(identification tests),即每次给出一个样本,让听辨者指出它是两个目的词中的哪一个,两者必选其一;另一种是区分测验(discrimination tests),即以ABX的方式给出样本,X要么是A,要么是B,让听辨者判断X是A还是B。本文的听辨实验采用第一种方式。首先编写DMDX脚本程序,让它控制每组中的20个合成样本。实验开始时,20个样本会按随机顺序在电脑上播放,每个样本以1秒的间隔播放三遍后,电脑屏幕上会立刻出现A、B两个选项,听辨者根据听到的声音决定选A还是B,两者必选其一,不可漏选,且一定要在5秒钟之内做出选择。连续听完六组合成样本大概需要二十几分钟。听辨者年龄在15至70岁之间,年龄分布合理,都出生和生活于禹州境内,只讲禹州方言,不带其它口音。听辨工作以走访形式在禹州的不同地点进行。

3.3 根据最终得到的27人的听辨数据,计算出每组的20个样本被听成两个目的词的百分比数,见图4和图5所示。可以看出,禹州方言四个声调两两之间都有着明显边界的范畴感知,即在范畴内部无论基频如何变化都只能被感知为同一个声调,一旦变化跨越范畴边界就被感知为另一个声调。

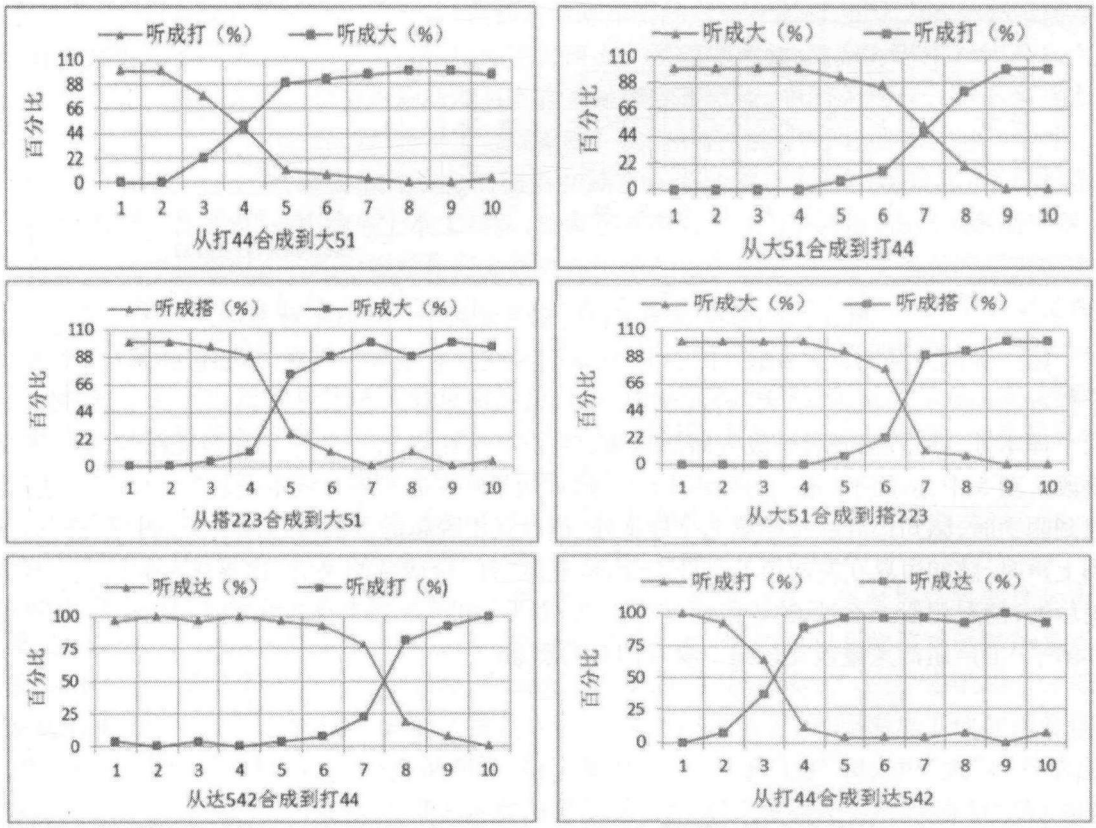


图4 上:上声对去声;中:阴平对去声;下:阳平对上声

这6组又可分两种情况。图4的3组为第一种情况:从图4上可看到,在从上声“打”合成到去声的情况下,感知边界在左起第四个样本处,在从去声“大”合成到上声的情况下,感知边界在左起第七个样本处,也就是说,不管是从上声合成到去声,还是从去声合成到上声,感知边界是重合的。图4中,不论是阴平“搭”合成到去声,还是去声“大”合成到阴平,感知边界也都大致在同一个点上。图4下也基本如此。

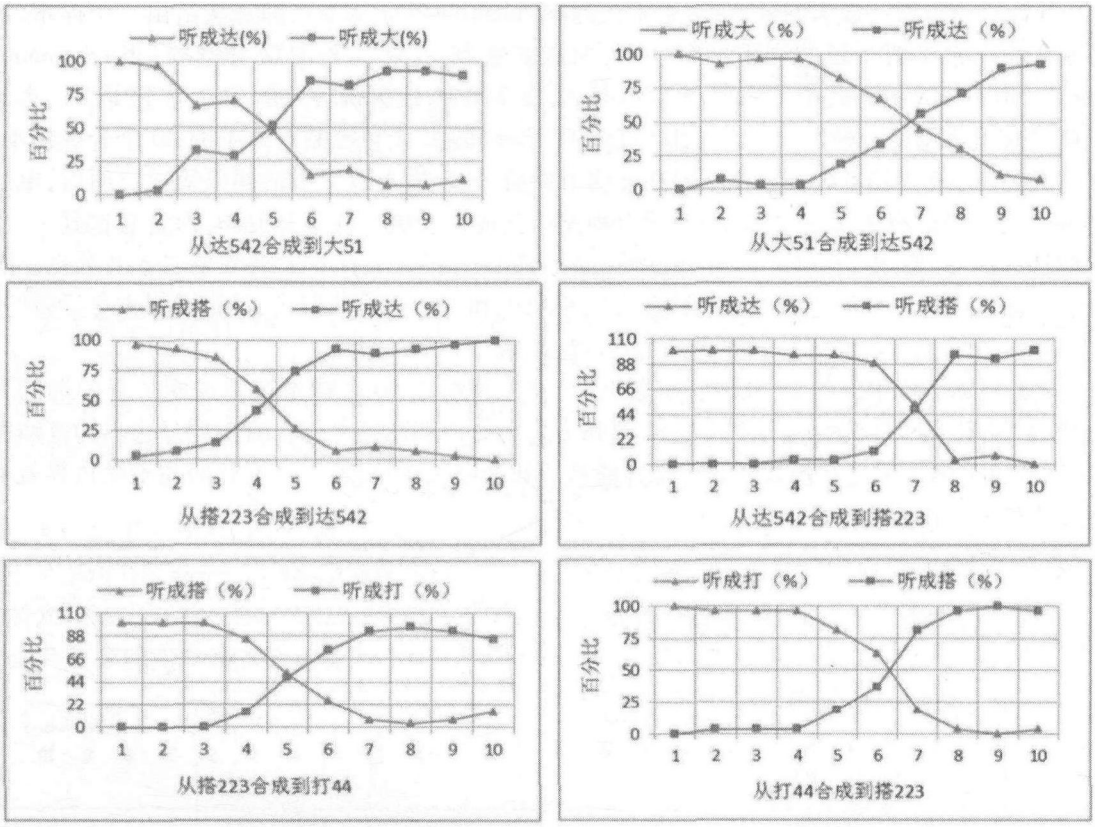


图5　上:阳平对去声;中:阴平对阳平;下:阴平对上声

第二种情况是图5的3组。图5上,在从阳平“达”合成到去声时,感知边界落在左起第五个样本处,然而,在从去声“大”合成到阳平时,感知边界却落在了左起第六个样本后、接近第七个样本处,就是说,两种合成方向的感知边界没有落在同一个点上,而是彼此跨越,有一个重叠段。图5中,从阴平“搭”合成到阳平时,感知边界在左起第四个样本稍后处,从阳平“达”合成到阴平时,感知边界在左起第七个样本处,也有互相跨越的重叠段。图5下,阴平“搭”合成到上声时,感知边界在左起第五个样本处,从上声“打”合成到阴平时,边界在左起第六个样本稍后处,情况相似。不过,这3组还是有不同,阳平-去声组的重叠段比较长,而阴平-阳平组和阴平-上声组的重叠段比较短。参看图6的示意。

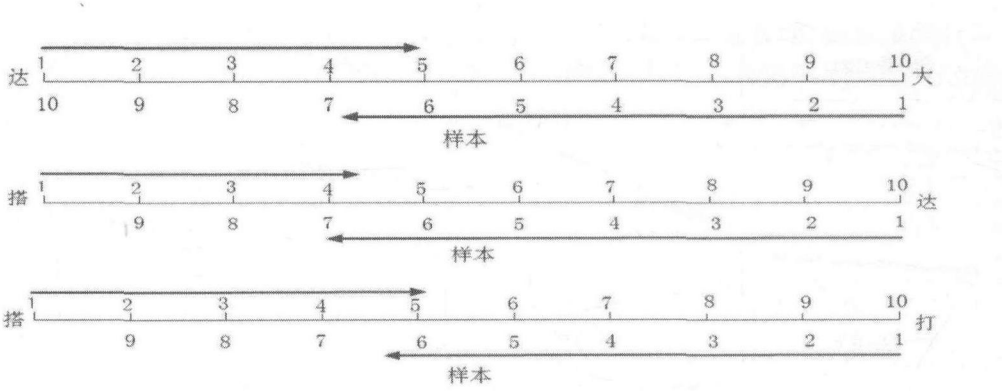


图 6 “达-大”“搭-达”“搭-打”的感知边界跨越示意图

是什么导致了这种感知边界错位而不重合的现象？可以从基频曲线、音强、音长、音色和发声等几个方面去逐一进行分析。首先，用于合成的四个母本是同一个发音人在基本相同的强度上发出的，并且在合成的过程中也没有改变音强，因此可以排除音强的作用。其次，音长的影响也基本可以忽略，因为如上文 2.3 所言，禹州方言的四声在音长上不存在显著性差异，也就是说，长点儿、短点儿对四声的感知并没有多大影响。再次，音色方面，用于听辨的 120 个声音样本都拥有相同的声、韵母结构[tɑ]，都有着基本相同的共振峰模式，因此音色的作用也可以排除。

有一个基频曲线的“拐点”问题需要特别讨论。金健和施其生(2010:544-576)在研究闽语谷饶方言时发现，基频曲线上拐点位置的前后对声调的感知有影响，并认为有必要将拐点前后作为各种调型中都可能出现的区别特征进行研究。从图 2 可知，禹州方言阳平调是先大致持平再下降，中间有拐点，而去声调从最高点直直地降到最低点，中间无拐点。是不是拐点的有无造成了此二调感知边界的跨越不重合现象呢？我们的合成设计可以排除这种可能性。顺着图 3 中的箭头往下看，阳平基频线上的拐点在随着调形的变化而逐渐平滑化，直至拐点消失。逆着箭头往上看，去声调形上本无拐点，随着它慢慢地变成阳平调形，拐点也渐渐出现、最后确定。这种基频模式的均匀缓变肯定会影响到阳平和去声的感知，随着拐点的逐渐模糊消失，听辨人越来越听成去声，反过来，随着拐点的渐渐出现，听辨人越来越听成阳平。但是在没有基频以外的因素起作用时，从阳平合成到去声和从去声合成到阳平两种情况下的感知边界应该是重合的，因为这两种情况下的合成都是采用了同一套基频值，不管有没有拐点，纯粹基频的改变不会造成两种情况下感知边界的交叉跨越。同理，图 5 中的阴平-阳平对及阴平-上声对的感知边界跨越不重合现象也不是拐点造成的。

排除了其他因素，造成感知边界不重合现象的，看来最有可能是发声类型的差异。

3.4 下面对禹州方言四声的发声特点作一些分析。

孔江平(2001:209)曾用基频(F0)、开商(OQ)、速度商(SQ)这三个参数对普通话的声调模式作过研究。他认为基频模式反应的是声源的时域特性，速度商和开商模式反映的是声源的频率域特性，前者可被看作“调时模式”，后两者可被看作“调声模式”，三个参数结合才能更好地描述声调的特征。基于此，我们请发音人男 A 再次进行录音，同步采集了他的声音信号(SP)和喉头仪信号(EGG)，并从他的阴平“搭”、阳平“达”、上声“打”和去声“大”中各选出 4 个双通道语音样本进行了基频、开商和速度商提取，然后按四个调类把各样本的 F0、OQ 和 SQ 值分别作了平均和数学拟合，如图 7 和图 8 所示。

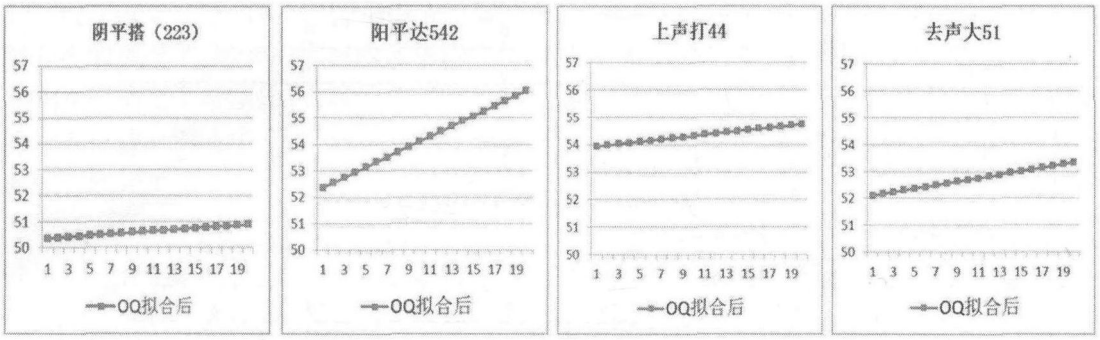


图 7　禹州方言四个声调的开商模式(OQ 拟合后)

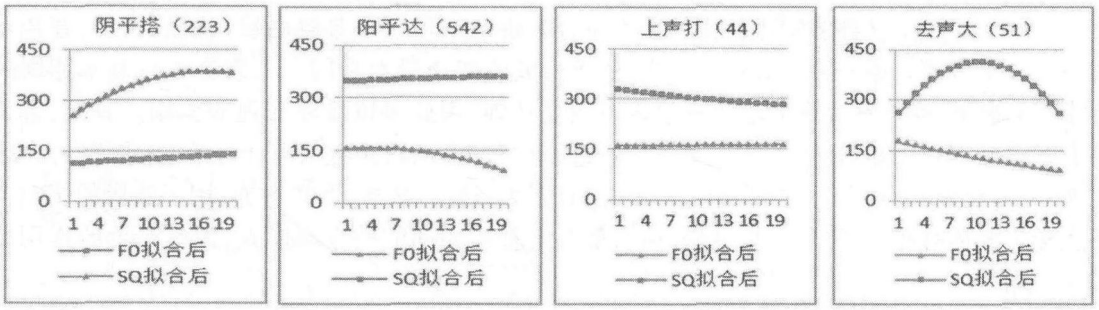


图 8　禹州方言四个声调的基频和速度商模式(每图内上:SQ 拟合后;下:F0 拟合后)

图 7 显示,禹州方言四个声调的开商值各自在不同的范围内变化,阴平从 50.37 微升至 50.90,阳平从 52.36 高升至 56.05,上声从 53.95 微升至 54.75,去声从 52.11 渐升至 53.34。阴平和上声的上升幅度很小,阳平和去声的上升幅度明显较大。若阳平和去声相比较,前者的升幅比后者要大得多,前者各点的开商值都大于后者中各对应点的开商值。

图 8 中四个声调的速度商曲线也显示出了不同的变化趋势。随着阴平调基频的缓缓上升,其速度商也明显上升,只是在末尾才略有降势。阳平调的速度商随基频的弧形下降而略有上升。上声的基频有很轻微的上升,但其速度商却明显呈降势。去声最特别,其基频曲线斜直下降,而其速度商则先陡升再陡降,呈大弧形。总体上看,四个声调速度商的最大值为 476.51,最小值为 210.02,均值为 339.91,标准差为 56.58。

由此可见,发音人男 A 的四个声调除了在基频曲线上表现为不同的模式外,在发声方面也各有特点。

3.5 PSOLA 合成器属于波形拼接的合成方式,它在改变声音样本的基频和音长时,并不改变其发声类型。在一个声调对立组中,把第一个声调的原声一步步合成到第二个声调的基频时,只是改动了基频曲线,原声调的发声特性仍然保留在合成所得的样本中;反之,对第二个声调进行合成时,其发声特性也一直被保留在合成所得的样本中。理论上,当两个声调的发声特性不相同时,在合成样本中滞留下来的发声特性会和基频同时对声调的感知起作用。当基频的逐步变化到达可能的范畴边界时,原声的发声特点会继续起作用,倾向于使听辨者把合成的样本仍然感知为原来的声调,于是就推动范畴边界后移。在来自同一个对立组的前 10 个样本与后 10 个样本中,范畴边界朝着相反的方向被推动,就形成了图 5 和图 6 中所示的、两个不同方向的范畴边界不重合的现象。换言之,发声类型的不同可以影响声调感知。

显然,发声类型在不同的对立组中起作用不同。有三组基本不起作用,起作用的三组,如前文已指出,同样出现感知边界互相跨越的情况,阴平-阳平对立组和阴平-上声对立组在程度上比阳平-去声对立组弱很多。对此需要有一个解释。

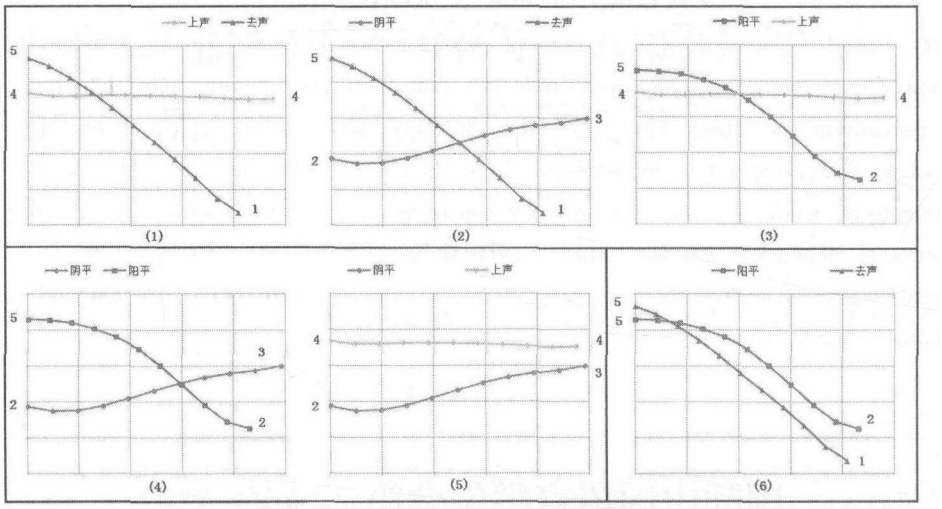


图 9　禹州四声轨迹两两比较

图 9 显示了六个对立组的基频轨迹。可以看出,不同组的声调的终点即调尾音高差有所不同:上-去、阴平-去和阳平-上的调尾音高差都较大(跨 4 度或 3 度);阴平-阳平、阴平-上和阳平-去的调尾音高差都较小(跨两度)。后三组中,阳平-去调头的音高差非常小。总之,直观地看,后三组比前三组调形差别小,而阳平和去声的调形差别最小。

显而易知,调形差别的大小,对应基频模式对声调的区别力的强弱。在禹州方言中,这又恰好对应发声类型对声调感知所起作用的大小:前三组基本上是基频模式在起作用,后三组都有发声类型的作用,而以阳平-去声对立组中的发声类型作用最大。

我们的结论是:在禹州方言的声调感知中,基频模式起主要作用,发声类型起补偿作用;在基频区别力强的情况下,发声对感知的贡献较小,相反,在基频区别力弱的情况下,发声在感知上的作用就会变大。这个认识,能在声调研究上启发我们作更多的思考。

参考文献

贺　巍 2005　中原官话分区(稿),《方言》第 2 期
金　健、施其生 2010　汕头谷饶方言多个降调的声学分析和感知研究,《中国语文》第 6 期
孔江平 2001　《论语言发声》,(北京)中央民族大学出版社
孔江平 1995　藏语(拉萨话)声调感知研究,《民族语文》第 3 期
李淑娟 2004　禹州方言词汇研究,南京师范大学硕士学位论文
屈颜平 2010　禹州方言连词、副词、介词研究,河南大学硕士学位论文
邵文杰等 1995　《河南省志·方言志》,(郑州)河南人民出版社
石　锋 2010　论语音格局,《南开语言学刊》第一期
张启焕、陈天福、程　仪 1993　《河南方言研究》,(郑州)河南人民出版社
朱晓农 2010　《语音学》,(北京)商务印书馆
Edwin M-L. Yiu 2013　International Perspectives on Voice Disorders, Multilingual Matters

Acoustic and Perceptual Studies on the Four Tones of Yuzhou Dialect, Henan Province

ZHANG Ruifeng & KONG Jiangping

Abstract Based on acoustic and perceptual studies, this paper determines the four tone values in Yuzhou Zhongyuan Mandarin, and finds out that both F0 contours and phonation contribute to the perception of tones. The F0 contours play the leading role while the phonation compensates. When the pitch contours are less distinctive, phonation will increase its effect on tone perception, while in cases where F0 patterns are sufficiently distinct from each other, the contribution of phonation will be reduced even to zero.

Key words tone, categorical perception, creaky voice, phonation, fundamental frequency, open quotient, speed quotient

---

国际中国语言学学会第22届年会暨第26届北美汉语语言学年会

由国际中国语言学学会主办，马里兰大学语言、文学及文化学院承办的国际中国语言学学会第22届年会(IACL-22)暨第26届北美汉语语言学年会(NACCL-26)于2014年5月2日至4日在美国马里兰大学召开。

会议共收到论文480余篇，经专家匿名评审，最终有来自13个国家和地区、128所大学和研究机构的300余名参会者报告了270余篇论文，内容涉及汉语及中国境内语言的句法、音系、形态、儿童语言习得、二语习得、社会语言学、应用语言学等方面的研究。北京语言大学李宇明教授、台湾新竹清华大学曹逢甫教授、台北中央研究院语言学研究所郑秋豫研究员做了会议主旨演讲。大会另设“汉语全球化研讨圆桌会议”和“语言与方言研讨圆桌会议”。经会议组委会学术委员的严格评选，南开大学博士生盛益民获得本年度“青年学者奖”。

国际中国语言学学会于5月3日下午召开全体会员大会，台北中央研究院语言学研究所郑秋豫研究员就任新一届会长。美国威斯康辛大学张洪明教授在为学会服务了十四年之后卸任执行秘书长一职，由中国社会科学院语言研究所胡建华研究员接任。这是国际中国语言学学会自1991年成立以来，首次由中国内地学者担任执行秘书长，学会秘书处同时也移至中国内地。国际中国语言学学会秘书处负责学会的日常运作，是学会的主要职能部门。

国际中国语言学学会第23届年会将于2015年6月19-21日在韩国首尔汉阳大学举行。

(IACL秘书处)

□韩启超

南昆念白声学实验分析（一）

摘　要：昆曲念白是研究昆曲艺术的重要内容之一，本课题《昆曲念白声学实验分析》在关注音乐本体的基础上，运用言语声学的理论与方法，对昆曲念白进行声学实验分析，以期探索研究昆曲艺术的新途径。因此，本文主要集中在三个方面：昆曲念白的声调、昆曲念白的时长、昆曲念白的音高。通过定性与定量的分析，对昆曲“韵白”声调进行合理拟测，归纳出“韵白”声调特征；在“韵白”与“中州韵”念白时长的比较中归纳出韵白时长与节奏关系;在“韵白”与“中州韵”念白音高的比较中,归纳出“韵白”音高与字声、旋律、曲情的关系等等。

关键词：南昆；念白；声调；韵白；中州韵；构拟

DOI:10.14113/j.cnki.cn11-1316/j.2014.04.016

一、实验材料

戏谚云：“千金话白四两唱”，昆曲界也有“一引二白三曲子”之说，充分说明念白在戏曲的重要地位。正如昆曲大家俞振飞先生所说，曲子有谱，有伴奏，唱者有所依傍；引子有谱，无伴奏，比较难些；白口则既无伴奏，也无谱子，怎样念得准确、好听，全由演员自己掌握，所以难度更大。[①]事实也是如此，念白本身没有旋律，学者无法通过乐谱化的固定旋律来进行字调的比较研究，演员也无法进行依谱念唱，唯一能作为依托的是字声，即字调的四声阴阳规律，所依韵书主要有三：《中原音韵》、《洪武正韵》、《韵学骊珠》。虽然，这三者之间有着密切继承关系，但当前昆曲在传承中到底是遵循何种韵书，学界还有着不同的观点，相对比较统一的观点认为，《韵学骊珠》是其诞生以来昆曲遵循的主要法度。近代以来，部分曲家通过长期的表演实践，对昆曲的念唱规律进行了系统总结，如俞振飞先生的《谈昆曲的唱念做》、《念白要领》、《习曲要解》，宋衡之先生的《昆曲唱念基础知识》等。但，综合历代学者、曲家对昆曲字声规律、念唱规律的相关研究来看，昆曲艺人在念白上虽有“成规”，但并非“定规”。主要是根据字调四声阴阳结合曲情进行“自我处理”，其节奏、声调、长短、强弱均具有一定的弹性和自由性。因此，本文试图从声学角度对昆曲念白进行系统研究，以期探索昆曲念白在声调上的“定规”与“自由度”。

实验材料是选取南昆旦行代表剧目《牡丹亭·惊梦》、《玉簪记·偷诗》、《铁冠图·刺虎》三段念白。发音人是苏昆自“传”字辈开始的第三代演员，师承正宗，音色清纯。（1965 年生，昆曲表演艺术家，国家一级演员。1985 年毕业于江苏省戏剧学校，师从张娴、张继青、胡锦芳，工闺门旦、正旦。）录音样本分表演念白和中州韵念白、普通话念白三部分。表演念白是昆曲演员（生行和旦行）舞台表演中的念白，即中州韵的舞台念白简称“韵白”；中州韵念白是演员用日常说话的方式,按照中州韵（演员师承学习过程中的固定字音）逐字将韵白部分念出，简称为“中州韵”。研究对象则是“韵白”与

作者简介：韩启超（1977～　），男，浙江师范大学音乐学院副教授，北京民族音乐研究与传播基地兼职研究员。

① 俞振飞:《念白要领》,《俞振飞艺术论集》，上海：上海文艺出版社，1985 年，第 363 页。

“中州韵”，全部样本共计 3 段，47 句，280 个字（包括重复）。其中《牡丹亭·惊梦》原有 32 个字，去掉重复，共计 28 个字；《玉簪记·偷诗》原有 106 个字，去掉重复，共计 91 个字；《铁冠图·刺虎》原有 142 个字，去掉重复，共有 101 个字。当然，不同唱段之间的重复不在统计之内。

二、实验方法

录音是在安静的起居室完成。语音信号由索尼电容式麦克风采录，录音时麦克风挂在演员的胸前，距离唇部 15—21 厘米。标准采样频率为 20kHz，分辨率 16 比特。每段录音后都包含了一段用于校准的 1000Hz 声音，以及该声音在相应距离下的声级数值。实验分析中采用的标准采样频率为 11025Hz，分辨率 16 比特。

录音材料综合运用 Labchart7、Audition 进行预处理，所有长于 10 毫秒的空白段都被切除。然后将录音材料以单字和单句为单位分别进行剪切，提取语音分析数据，所用软件是 Wave Surfer。然后用 Praat 软件提取基频、计算时长、音高，并参照 Wave Surfer、Mat Lab 平台自编程序处理结果。用 Excel 进行换算，运用 SSPS 进行相关性分析。

三、实验结果

（一）昆曲念白时长

1. 整句时长对照

时长单位：秒（s），截取样本时遵循四舍五入精确到毫秒（ms）。（见文后所附表 1、2、3）

通过统计发现，三段念白中，韵白和中州韵共有者计 48 句（《铁冠图·刺虎》中，“罢，罢，罢！”三字合在一起，算为一句），其中 42 句韵白时值比中州韵长，占 87.5%。只有 6 句中州韵时值比韵白时值长，占 12.5%，而这 6 句中，最长差额是 1 秒，有 2 句的时值差额还不到 0.01 秒，基本可以忽略。单句最长差额为 15.512 秒，如《牡丹亭·惊梦》“恁般天气好困人也！”之句。可见昆曲整句韵白时值总体比中州韵时值长，说明昆曲生行、旦行等以中州韵为标准的演员在进行念白表演时，其整句的时长普遍比正常说话状态下的整句时长要长，也即昆曲韵白大都是正常语言的拉伸。从这三个唱段时长比较来看，这种语言的延伸拉长往往是根据曲情来确定。曲情舒缓、情绪低回、感情真挚时一般时值较长，乐句呈延伸状态；而曲情激烈、紧张、情感激愤时乐句往往呈压缩状态，即韵白节奏较快，语言急促，相对时值比较短促。

2. 单字时长比较

单位：ms（毫秒），截取样本时遵循四舍五入精确到毫秒。（见表 4）

按照表四所列《牡丹亭·惊梦》韵白、中州韵单字时长数据比较的方法，对《玉簪记·偷诗》、《铁冠图·刺虎》中的韵白、中州韵单字时长进行列表比较（具体数据表不再列出）。

通过统计发现，三段 280 个字中，220 个字韵白时值大于或等于中州韵时值，60 个字韵白时值小于中州韵时值，韵白时值大于或等于中州韵者占 78.6%，韵白时值小于中州韵者占 21.4%，这一数据与整句中韵白、中州韵时长比例近似，充分说明在昆曲表演中，韵白字大都是中州韵字时值的拉伸扩展。

基于这三个样本来看，将中州韵字时值延伸最长为 7988 毫秒，即其延伸的范围为 0—7988 毫秒，平均延伸为 595 毫秒；当然，部分字也根据剧情及字调特征加以紧缩（韵白时值小于中州韵时值），其紧缩的范围为 0—283 毫秒，平均紧缩为 105 毫秒。通过《韵学骊珠》七个声调类别的查证，这 60 个紧缩字中（包括不同段落的重复字，用阿拉伯数字加以区别），上声字最多，为 16 个：我[1]、我[2]、我[3]、我[4]、我[5]、此[1]、此[2]、写、缕、可、想、取、了、把、首、指，占 26.7%；其次是阴平声，有 15 个：心、先、中 1、中 2、思、将、经、心、家、皆、宫、身、君、兄、只，占 25%；阴入声 8 个：一[1]、一[2]、笔、得、国、压、作、得，占 13.3%；阴去声 8 个：寄、自、睡、笑、些、个、在、近，占 13.3%；阳平声 8 个：词、情、唔、云、凡、为[1]、为[2]、时，占 13.3%；阳去声 3 个：自、念、望，占 5%；阳入声 2 个：夺、贼，占 3%。

整体来看，规律并不明显。但排除连读、弱音等现象，基本可以推定昆曲念白中上声字和平声字常常在时值上出现紧缩现象，而入声字则并不明显。还有一个现象比较突出，即“我”字，三段之中共有 8 次重复，紧缩字占 62.5%、；“一”字共有 6 个，紧缩字占 33.3%；、“中”字共有 2 个，紧缩字占 100%；“此”字共 4 个，紧缩字占 50%。这说明昆曲念白中部分常用字本身具有紧缩特性，但具体是哪些字，尚需进一步统计。

统计韵白比中州韵时长差值超过 1000 毫秒的有 28 个，其中上声字最多，9 个，占 32%；其次

是阴平字 5 个，阴去字 5 个，阳去字 5 个，入声字（阴入）仅有 1 个。**由此可知，上声字在所有声调类别中最具弹性，既可以任意压缩时值，又可以极大地延伸拓展时值，从而丰富念白的节奏性、旋律性。其次是阴平字也具有随意紧缩和伸展的特性，去声字的延展性相对比紧缩性显著，而入声字可以紧缩却极少能够延伸拉长。**

从句子结构来看，每句的首字时长与尾字时长有着显著不同，详见文后图 1、2、3、4、5、6。

结合表四以及其他两段的统计数据，可以看出，这三段念白在每句的首字与尾字的时长安排上有着显著的规律。即每一完整句（意思表述完整，往往以句号、问号、叹号为标志）的句尾字韵白时长都数倍于中州韵时长（最长超过 10 倍，如《牡丹亭·惊梦》段中的“也”字），同时也显著长于句首字时长。句首字（包括完整句中的分句句首）平均时长 644.9ms，韵白与中州韵平均差值是 191.5ms。而句尾字（包括分句句尾字）的平均时长是 1634.6ms，韵白与中州韵的平均差值是 1107.4ms。整句句尾字的平均时长达到 2916.5ms，韵白与中州韵的差值达到 2315.1ms。这其中只有一句句尾字例外，即《铁冠图·刺虎》中的“得”字，韵白时长比中州韵时长短 201ms，究其原因，**“得”是入声字。这印证了昆曲界多年唱念经验的总结，即：“入声必短促”之论断，也说明在昆曲创作中，入声字一般是不能作为句尾字，即便在句中也很少重复或者频繁使用。**

分句（意思表述不完整，往往以逗号为标志）句尾字平均时长为 618.5ms，韵白与中州韵平均差值为 123.4ms。将分句句尾字时长与整句句尾字时长比较，可知，整句句尾字时长平均是分句句尾字时长的 4.7 倍。从三段句子结构来看，一个整句含 2 个分句的句式最为普遍，这就形成了一种“短”+“长”的韵白时长模式。**这种一短一长的尾字时长结构造成了一种固定的节奏长短模式，与常规音乐作品的乐句结构有着惊人的吻合。**对此，诸多曲式学论著都提到这一现象，认为旋律和语言一样，并不是永无停歇连贯不断地进行着，而是抑扬顿挫分成许多大大小小的相互联系的部分。一般是两个乐节构成一个乐句，两个乐句构成一个乐段。在每一个乐句结束时往往是以长休止或者长音符作为乐思完整或不完整的标志。[②]问题在于是语言自身的句子节奏规律影响了旋律句子结构特征的形成，还是旋律的曲式结构影响了语言的结构？这很难说清，但从语言发生学的角度来说，显然是人类语言的节奏、结构模式影响了音乐旋律的节奏、结构特征。至少，从昆曲念白的句子特征来看，业已形成了和音乐旋律基本一致的特性。这种特征也影响了昆曲唱腔的旋律结构，使昆曲唱腔旋律中的乐句不仅具有这种“短 + 长”的模式，而且明显成为韵白结构扩展后的再现，即无论是分句句尾还是整句句尾，都是时值相对较长，形成一种稳定、均衡的美学特征，为昆曲在演唱上形成“每度一字，几尽一刻”，[③]“气无烟火，细如游丝”[④]的“水磨腔”提供了可能。

另外，从所有录音样本 280 个字的韵白与中州韵时长曲线图（图 7）来看，昆曲中州韵念白时长比较均匀、舒缓，处于同一水平曲线。而韵白时长则长短不一，极具跳跃性，显示了昆曲韵白的节奏性、律动性比较突出。正如俞振飞先生所说：戏中情节的推进，人物思想情感的变化，经常是通过人物的独白和对白交代给观众的。[⑤]因此，韵白复杂的节奏性、律动性能够增强戏剧矛盾冲突、彰显人物形象，加深观众对剧情的理解，对演唱用典的理解，增强演出的效果。

（二）昆曲念白音高

根据表 5、6、7 完整数据，利用相关软件可以得出这三段 280 个字的韵白与中州韵整句音高走势比较曲线图。以《牡丹亭·惊梦》为例（图中“韵”指“韵白”，“本”指“中州韵”，MAX 代表最大值，MIN 代表最小值，AVE 代表平均值），具体曲线如图 8。

从表七、表八、表九以及图 8 所代表的三段念白整句音高走势比较曲线图来看：

韵白音高跳跃性较大，其整体音域范围为

---

② 李重光：《基本乐理》，北京：高等教育出版社，1992 年，第 104–105 页。

③［明］袁宏道：《虎丘记》，《袁中郎全集》，上海：中央书店，1935 年，第 1 页。

④［明］沈宠绥：《度曲须知·曲运隆衰》，《中国古典戏曲论著集成》（五），北京：中国戏剧出版社，1959 年，第 198 页。

⑤ 俞振飞：《戏曲表演艺术的基础》，《俞振飞艺术论集》，上海：上海文艺出版社，1985 年，第 272 页。

137Hz——913Hz；中州韵比较平稳，音域范围为55Hz—372Hz。韵白的平均音高是504Hz，中州韵平均音高是228Hz，二者相差276Hz。换算成固定音高值，韵白整体音域范围是♯C—♭b²，跨越3个八度，中州韵整体音域范围是A₁—♯f。日常男声说话的频率范围是：95Hz—142Hz，女声频率范围是：272Hz—653Hz；声乐演唱中，一般能够有效利用的音域范围是80Hz—1300Hz，女中音约为170Hz—683Hz，女高音约为246Hz—1024Hz。这表明三段昆曲韵白女声样本的音域比正常女声说话频率范围宽，在同一人发声的情况下，韵白整体音域比中州韵高两个八度还多，这应该是一种比较特殊的现象，或者说演员在念白时是处于一种真假声结合的状态，不仅仅是在时长上对语言进行拉伸或紧缩，还在音高上进行“转喉押调”，整体移高，以形成“疾徐、高下、清浊”⑥、“婉协、毕匀”之声⑦。

另外，韵白单字音域跨度也比较大，在所有280个韵白字中，有134个单字音域超过200Hz，单字音域最大跨度的是《铁冠图·刺虎》“为君父报仇”句中的“报”字，音域范围是234Hz——913Hz（接近♭b——♭b²），落差为679Hz，单字音高曲线跨越2个八度，这在昆曲演唱中几乎是不可能的，足见念白难度之大。如图9。

单字音域超过400Hz（音高在♯g¹之上）有14个字，其中13个字集中在《铁冠图·刺虎》中（费、宫、流、篡、报₁、主、报₂、他、为、罢、待、有、理），另外一个在《牡丹亭·惊梦》中（谴）。结合平均音高值来看，韵白中的大部分字音高还是集中在中音区，以真声为主。从三段韵白的曲情来看，《铁冠图·刺虎》是全剧的高潮，情绪起伏较大，字里行间充满激愤、仇恨之意。如此曲情造就了演员在念白时，频繁将字调音域范围拉大，以突出剧情的戏剧化。反过来，正因为是大量的宽音域字的出现，频繁的音域跳动进一步彰显了曲情、人物的激愤之意。结合调类来看，这些跨度大的字在调类归属上并没有明显的规律，这进一步说明**昆曲念白的音高域值往往是根据曲情来处理的。**昆曲念白在音高上虽有“起调”一说，并没有严格的“定规”，而是有着一定的“自由度”。魏良辅所谓“疾徐、高下、清浊之数一依本宫”⑧，显然并非针对昆曲念白实践。

从句子（包括整句和分句）音高走向来看，中州韵音高相对平稳，基本在228Hz（接近♭b）上下波动，说明声音非常低沉。韵白的音高走势基本呈现出三种类型：

（1）“低开——中高——低收”；

（2）“低开——中高——高收”；

（3）“高开——中高——低收”；

这三种类型的共性是“中高”（句子中间的部分音域高）。从三段样本24整句（包括48个分句）情况来看，除了2句是“低开——中高——高收”，1句是“高开——中高——低收”外，其他21个句子都是“低开——中高——低收”，大量的分句也都符合“低——高——低”这一类型。**这足以说明昆曲念白中的句子在音高处理上普遍是中间部分的音域偏高，而“低开——中高——低收”则构成了昆曲念白中句子音高走势的基本形态。**当然，“中高”并不是绝对的，这要视句子的长短，如果句子长的话，“中高”往往会呈现出波动，从而形成“中高——中低——中高”的现象。

将这三段样本所有分句的首字、中字（以均分为主，如果不能均分，大体取中间部分音高均值最高的字）、尾字平均音高进行统计，其结果也印证了这一说法，见图10。

从统计情况来看，最高音域的字一般在句中，“首低”并不意味着首字是最低音，“尾低”也不意味着尾字是最低，仅是相对而言。一般情况下，句子的倒数第2或3字音最低，如“黄—稳”句、“我—得”句、“为—端”句、“指—仇”句等。

昆曲旦行念白句子在“首——中——尾”部分所呈现出的“低开——中高——低收”的特征与昆曲在唱腔时针对念字所形成的“橄榄腔”非常吻合。所谓“橄榄腔”，清代叶怀庭就曾提到过，俞振飞先生界定曰：“每支曲子的每个字都有它一定发挥的地方。所谓抑扬的唱腔，最适宜用在俗称‘宕三眼’的地方，或者是散板中的卖腔，唱时由轻而响，再

⑥［清］余怀：《寄畅园闻歌记》，见［清］张山来《虞初新志》第二册卷四，上海进步书局影印本，民国年间，第5—6页。

⑦［明］沈宠绥：《度曲须知》，《中国古典戏曲论著集成》（五），北京：中国戏剧出版社，1959年，第198页。

⑧［清］余怀：《寄畅园闻歌记》，见［清］张山来《虞初新志》第二册卷四，上海进步书局影印本，民国年间，第5—6页。

由响而收，像橄榄一样，首尾细，中间大。这种唱法，名为橄榄腔。”⑨顾聆森先生进一步指出，作为昆曲腔格之一，“橄榄腔”在演唱时是“凡一音延伸数拍，唱时先控制音量，然后慢慢放足，过半后又渐渐收细。音量两头轻细，中间宏大。”⑩

不言而喻，昆曲韵白中所呈现的句子音高走向特征也正是这种“唱时由轻而响，再由响而收，”“音量两头轻细，中间宏大”的橄榄状，只是韵白是以句子的形式呈现出音高上的橄榄状，而非唱腔单字在音量上的橄榄状。正是因袭了单字腔格“橄榄腔”的韵味，才形成了昆曲从单字唱腔到整句韵白都呈现出典型的抑扬顿挫之独特韵味。

总之，韵白相对于中州韵在音高上所显示的大幅跳跃性，形成了韵白蜿蜒曲折、跌宕起伏的旋律特色，构成了昆曲念白特殊的润腔特色，增强了韵白的旋律性、音乐性。但这种字与字之间快速的音高跳动、单字内的大幅音高转换也增强了韵白乐谱化、固定旋律化的难度，加剧了演员学习和表演的难度，使得昆曲在传承、普及上受到局限。

（三）昆曲念白调值

1. 昆曲念白字调归类

根据《韵学骊珠》统计，综合韵书的注释和曲情内容，将三段念白 280 个字的声调划分为七个韵类，以《牡丹亭·惊梦》为例，列如表 8。

《玉簪记·偷诗》、《铁冠图·刺虎》所属调类、韵部统计表，与表十格式相同，不再一一列出。仅就调类归属来看，《玉簪记·偷诗》原文有 105 个字，去掉重复者，共计 91 个字。具体如表 9。

《铁冠图·刺虎》原文有 142 个字，去掉重复者，共有 101 个字，调类归属如表 10。

归纳起来，三段 280 个字中，去掉所有重复，阴平声 44 个字、阳平声 39 个字、上声 38 个字、阴去声 28 个、阳去声 30 个、阴入声 12 个、阳入声 11 个。

2. 昆曲念白中州韵调值拟测

利用 Praat 软件将所有 280 个中州韵字的声调基频曲线提取出来，提取时设置 pitch 值为 50Hz——500Hz，去掉弯头降尾，确定声调长度之后，通过固定程序设置获取 20 个等分基频数据。在归一化处理的基础上，计算出七个调类的均值，具体如图 11（具体操作时同时参考 Wave Surfe 软件所提取数据）：

将上述基频平均数据根据五度值公式（公式 1）求出声调五度值格局图（图 12）。

公式 1：

$T = (\lg x - \lg b) / (\lg a - \lg b) \times 5$。

公式中 a 为调域频率最大值，b 为调域频率最小值，x 为测量点频率值，取常用对数目的是为了让五度值与听感相吻合。

从图 11、图 12 中可以明确看出，**这三段念白中州韵所反映七个调类的平均调值及其曲线总体呈下降趋势。**具体来说，阴平声略降，调值为 54；阳平声为曲线，时长均值最长，调值接近为 313；上声为降调，调值为 41；阴去声为降调，调值为 51；阳去声与阴去声基本相同，为降调，调值为 51；阴入声最为短促，为降调，调值为 52；阳入声较为短促，曲线，调值为 323。

当然，三段念白中，有个别字的声调曲线与上文所拟测的均值有着显著不同，据统计，在七个调类中，差异最多的是阳平声，有 10 个，如“么”、“聊”、“奴”、“谁”、“为”、“于”、“宜”等；其次是阴平声，有 5 个，如“和”、“精”、“幽”等；上声有 3 个，如“也”、“纸”；阳去声 2 个，如“沉”、“缠”；阳去声以及入声均没有。个别字的实际声调曲线如图 13、14、15、16（篇幅限制，每个调类列举一例）。

由此可以说明**演员在以正常言语状态念中州韵过程中，阳平声的字在调值上相对比较自由，或者说昆曲演员在阳平声调的处理上自由度相对于其他调类来说比较大，而阴去声、入声基本是固定化的，在声调上有着严格的限制。**

当然，上述对中州韵调值所进行的构拟结果以及部分字声调与五度均值产生差异化的具体原因，主要有三点：

第一，这些字受连读、重音等影响；

第二，受曲情影响；

第三，受普通话与地方方言的影响，演员在学念过程中出现字调的偏离。

3. 昆曲韵白调值构拟及其与中州韵调值比较

按照计算中州韵字五度值的相同方法，得出

⑨ 俞振飞：《振飞曲谱·习曲要解》，上海：上海文艺书出版社，1982 年，第 22 页。

⑩ 顾聆森：《昆曲腔格汇释》，《戏曲艺术》1991 年第 1 期，第 101 页。

280 个韵白字的声调五度值格局图，见图 17（当然，由于是表演韵白，在提取基频曲线时特别注重与听感配合，去掉一些与声调无关的参数）。

从图 17 中可以明确看出，**这三段韵白字所反映的七个调类的平均调值及其曲线总体呈下降趋势。**具体来说，阴平声为降，调值为 53；阳平声为曲线，调值为 341；上声为降调，时长均值最长，调值为 41；阴去声为降调，调值为 51；阳去声与阴去声基本相同，为降调，调值为 41；阴入声极为短促，为降调，调值为 53；阳入声最为短促，曲线，调值为 545。

将韵白字五度拟测值与中州韵字五度拟测值进行比较，可以发现两者调值基本相同，七个调类的调值都呈现下降趋势。其中上声、阴去声调值完全相同，均为 41、51；阴平声调值基本一样，中州韵为 54，韵白为 53；阳去声也基本一致，中州韵为 51，韵白为 41；阴入声基本一致，中州韵为 52，韵白为 53；阳入声基本一致，曲线，只是高度不同，中州韵为 323，韵白为 545.

二者的差异主要集中在三个方面：

其一，阳平声虽都是曲线，但调值走向截然不同，中州韵是 313，韵白是 341，也就是说二者曲线呈相反方向；

其二，在时长均值上，中州韵是阳平声最长，阴入声最短，韵白是上声最长，阳入声最短；

其三，韵白字七个调类除了阳去声之外，调首都呈现出上升趋势，在上升一定高度之后才开始呈下降趋势，尤其是阴去声和阳平声，在调尾又呈现出上升期趋势，整体呈斜 S 形。

这充分说明昆曲演员虽然都是以《韵学骊珠》为法度，但舞台念白（韵白）与自然说话（中州韵）在调值上还是有着相对明显的差异性。也即舞台表演的念白在声调上具有一定的“活法”。

四、结语与讨论

俞振飞先生在《念白要领》以及《习曲要解》中也多次总结了念白的声调特征，归纳起来有几点：

第一，韵白的高低音次序是：

上声——高

阴平——次高

阳平——中

入声——低

去声——次低[11]

第二，念白的调门（即笛色）。昆曲小锣的调门是 F 调“3”音，也就是标准的念白调门。[12]

第三，昆曲念白上声字的调值符号应为“∖”，去声字的声调符号应为“∨”。无论南、北曲，都是如此。

第四，入声字出口即住。[13]

但，从目前录音样本的实验数据及其分析结果来看，这三段昆曲旦行的念白显然与俞先生所说有着一定的差异。首先，韵白的高低音次序（从起音及整体音值来看）则是阴平——阴入——阴去——阳入——上声——阳去——阳平（由高到低排序）；其次，韵白的平均音高是 504Hz，首字均值是 500Hz，尾字均值是 450Hz（参见图 10），这说明韵白起音略高于 b^1，收音略高于 a^1，即只有收音符合俞先生所说念白的标准调门——F 调的“3”音；再者，上声调值为 41，与俞先生一致，但去声为降调（调值为 41 或 51），其调形符号与俞先生所说“∨”差距甚远。当然，由于实验样本的局限性，并不能轻易判定二者孰是孰非，俞先生对念白声调所做的特征归纳，以及目前样本的实验结论都有待于进一步论证。

综上，仅从这三段样本实验结果来看，韵白调值与中州韵调值基本吻合，证明了昆曲韵白并没有过于乐曲化、旋律化、自由化，而是深受字调影响，是在字调基本规定的基础上进行不同程度的拉伸或紧缩，也可以说是正常音调的一种夸张化结果。简言之，这种“定规”与“自由度”体现了昆曲“有规律的自由行动”的程式化精神。[14]

深入探究二者之间的这种“总体相同，细部差异”的原因，可以从韵白的实际调形走势上窥见一二。

通过对三段样本 280 个韵白字、中州韵字调形数据及其走势分析，并结合听感，发现韵白字调形

---

⑪ 俞振飞：《振飞曲谱·念白要领》，上海：上海文艺书出版社，1982 年，第 26 页。

⑫ 俞振飞：《振飞曲谱·念白要领》，上海：上海文艺书出版社，1982 年，第 30 页。

⑬ 俞振飞：《振飞曲谱·念白要领》，上海：上海文艺书出版社，1982 年，第 3 页。

⑭ 俞振飞：《程式与表演》，《俞振飞艺术论集》，上海：上海文艺出版社，1985 年，第 313 页。

基本有四种类型：

第一，正常调型。与正常说话状态下的中州韵调形完全一致，是上文所构拟的声调五度值的基本形态，共 168 个字，是主要类型。

第二，弯头提尾型。属正常调形的夸大化，与正常调形的“弯头降尾”不同，是指其调形呈现“低开——高升——下降——提尾 / 收尾”形状，其主体在中段下降部分。共有 54 个字，其中阴平字 12 个，阳平字 18 个，上声字 10 个，阴去字 6 个，阳去字 8 个，没有入声字。它又表现出两种基本形式：“弯头直尾型”和“弯头提尾形”，如图 18、图 19。

第三，上折线型。这类调形都呈现出低开，直线或弧线上升，然后呈直线或弧线下降。即整个调形是上折线，中间高，两端低。共 45 个字，其中阴平字 7 个，阳平字 10 个，上声字 15 个，阴去字 6 个，阳去字 5 个，阴入字 1 个，阳入字 2 个，如图 20、图 21。

第四，波浪型。整个声调曲线呈波浪状，类似颤音，是韵白固定旋律化的结果。共 13 个字，其中阴平字 3 个，阳平字 1 个，上声字 5 个，阳去字 2 个，阴入字 1 个，阳入字 1 个，没有阴去字。波线型还有一种变体，即通过演员的嗓音控制，形成一种非连续性的波浪状（这类调形在计算五度值的时候，是根据听感提取部分能区别意义的调值区间参数），如图 22、图 23。

从字数统计来看，“弯头提尾型”阳平字最多，占 33%，其次是阴平字 22%，上声字 18%；“上折线型”上声字最多，占 33%，其次是阳平字 22%，阴平字 15%；“波浪型”也是上声字最多，占 38%，其次是阴平字 20%。这虽然不能充分说明某一调类与某种调型有着直接的对应关系，但至少能进一步印证前文有关音高和时长的统计结果，即**昆曲念白中上声字、阴平字、阳平字的声调弹性比较大，在具体念唱中的自由度远远高于其他调类。相反，入声字由于有着严格的定规（入声必短促），基本没有弹性。**

因此，虽然在提取时已经尽量排除一些干扰因素，但这些不同类型的韵白单字实际调形曲线决定了所拟构的韵白五度调值与中州韵五度调值有着一定的差异。当然，正是这种差异才决定了韵白的特殊性，形成了韵白的润腔特色。即基于中州韵调值基本规则的基础上，演员通过各种发声方法、咬字吐字方式、润腔手段，形成了不同的调形特点，构造了昆曲念白的抑扬顿挫特色、丰富的节奏形式和近似旋律化的特征。俞振飞先生将其总结为“音乐性”和“语气化”，这种特征的形成也是昆曲自身发展的必然，因为昆曲念白在多数情况下没有乐器伴奏衬托，音量、音高、速度、节奏全靠演员自己掌握，要实现与伴奏乐队、唱腔之间的过渡自然，必然在字调节奏、音高、音量，甚至音色上体现出与唱腔和打击乐队近似的节奏性和律动性。

当然，目前为止，本研究还处于描述与构拟阶段，更多复杂问题，如多样化条件下样本的客观性、古代声调韵律实验的科学性、相关听感实验的体系性、与昆曲名家对昆曲念白声调经验总结的差异性等等，还需要进一步研究。

附言：本文是国家自然科学基金“戏曲嗓音发声类型声学分析及建模研究”（批号 11204275）、中国博士后特别资助项目（批号 2013T60013）的阶段性研究成果。同时，本课题系列实验研究也获得北京大学语音实验室、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研究”（批号 10ZD&125）的支持。

表 1 :《牡丹亭·惊梦》韵白、中州韵整句时长比较表

念词	蓦地游春转，	小试宜春面。	春吓春！	得和你两留连，
韵白	7.640（s）	10.981	12.754	12.218
中州韵	3.026	3.704	1.768	4.234
差值	4.614	7.276	10.986	7.984
念词	春云如何谴？	恁般天气好困人也！		
韵白	11.292	22.050		
中州韵	3.975	6.538		
差值	7.317	15.512		

表 2:《玉簪记·偷诗》韵白、中州韵整句时长比较表（部分数据）

念词	我，妙常，	自见潘郎之后，	不觉精神恍惚，	情思飘荡。
韵白	3.086	7.299	5.510	6.977
中州韵	1.973	3.186	3.307	2.500
差值	1.113	4.113	2.203	4.477
念词	对此凄凉时序	好伤感人也。	※想我在此出家，原非本性，	※只为身借宿于此，哪只弄假成真，罢。
韵白	4.295	8.417	3.062；2.949	5.223；4.791；0.922
中州韵	4.001	3.166		
差值	0.294	5.251		

注：统计时表格较为庞大，限于篇幅，仅列出局部数据，下同。

※：原剧中有，但录音时韵白中有，中州韵中无。

表 3 :《铁冠图·刺虎》韵白、中州韵整句时长比较表（部分数据）

念词	奴家费氏，	小字贞娥，	自幼选入宫闱，
韵白	3.204	4.359	6.531
中州韵	2.042	2.544	3.475
差值	1.162	1.815	3.056
念词	蒙国母娘娘命我服侍公主。	谁想（可恨）流贼篡夺我国，	杀死君父，
韵白	12.007	5.155	2.385
中州韵	6.507	5.222	1.653
差值	5.500	−0.067	0.732

表 4 :《牡丹亭・惊梦》韵白、中州韵单字时长比较表

念词	蓦	地	游	春	转，	小	试	宜	春	面。	春
韵白	911	1324	1535	2047	1281	1159	1329	1352	2293	3939	1864
中州韵	485	573	521	667	685	505	859	536	755	712	597
差值	426	751	1014	1380	596	654	470	816	1538	3227	1267
念词	吓	春！	得	和	你	两	留	连。	春	去	如
韵白	2024	7915	715	1360	935	3482	2425	1184	1380	3324	1754
中州韵	457	714	498	442	706	594	661	634	652	747	837
差值	1567	7201	217	918	229	2888	1764	550	728	2577	917
念词	何	谴？	恁	般	天	气	好	困	人	也！	
韵白	2311	1550	862	1088	1434	1255	1598	2105	1762	8843	
中州韵	694	764	590	597	576	688	689	578	854	855	
差值	1617	786	272	491	858	567	909	1527	908	7988	

表 5 :《牡丹亭・惊梦》韵白与中州韵音高数据表（部分数据）

念词	蓦	地	游	春	转，	小	试	宜	春	面。	春
韵 MAX	510	458	472	508	555	632	573	526	585	581	662
韵 MIN	382	275	241	428	351	447	314	286	507	396	503
韵 AVE	446	366.5	356.5	468	453	539.5	443.5	406	546	488.5	582.5
中 MAX	319	280	233	298	302	372	285	219	287	270	311
中 MIN	173	157	180	251	163	179	157	133	267	154	275
中 AVE	246	218.5	206.5	274.5	232.5	275.5	221	176	277	212	293

注：MAX 代表最大值，MIN 代表最小值，AVE 代表平均值，“韵”代表韵白，“中”指“中州韵”，单位：Hz，下同。

表 6 :《玉簪记・偷诗》韵白与中州韵音高数据表（部分数据）（单位：HZ）

念词	我，	妙	常，	自	见	潘	郎	之	后。	不	觉
韵 MAX	675	581	471	577	566	766	602	609	608	652	585
韵 MIN	378	274	322	437	467	600	430	559	422	523	441
韵 AVE	526.5	427.5	396.5	507	516.5	683	516	584	515	587.5	513
中 MAX	358	308	254	321	285	336	306	297	299	340	300
中 MIN	174	160	169	210	171	310	172	256	174	226	163
中 AVE	266	234	211.5	265.5	228	323	239	276.5	236.5	283	231.5

表 7 :《铁冠图 · 刺虎》韵白与中州韵音高数据表（部分数据）(单位 : HZ)

念词	奴	家	费	氏，	小	字	贞	娥。	自	幼	选
韵 MAX	667	624	667	536	603	626	772	482	484	604	619
韵 MIN	500	336	256	324	506	430	390	296	258	302	319
韵 AVE	583.5	479.5	461.5	430	554.5	528	581	389	370.5	453	469
中 MAX	342	283	339	256	333	320	309	226	287	233	313
中 MIN	212	151	167	172	282	160	260	82	191	159	175
中 AVE	277	217	253	214	307.5	240	284.5	154	239	196	244

表 8 :《牡丹亭 · 惊梦》念白所属调类统计表

平　声		上　声	去　声		入　声	
阴平声	阳平声	上声	阴去声	阳去声	阴入	阳入
春，真文韵	游，鸠侯韵	转，天田韵	试，支时韵	地，机微韵	得，拍陌韵	蓦，拍陌韵
和，歌羅韵	宜，机微韵	小，萧豪韵	吓，家麻韵	面，天田韵		
般，欢恒韵	留，鸠侯韵	你，机微韵	去，居鱼韵	恁，侵寻韵		
天，天田韵	连，天田韵	两，江阳韵	气，机微韵			
	如，居鱼韵	谴，天田韵	困，真文韵			
	何，歌羅韵	好，萧豪韵				
	人，真文韵	也，车蛇韵				

注 : 原文有 32 个字，去掉重复，共计 28 个字。

表 9

阴平声	潘、之、不、精、飘、思、凄、伤、心、幽、消、哎、呀、松、青、灯、钟、昏、孤、衾、先、中、丝、身、将、经
阳平声	常、郎、神、情、凉、时、人、词、聊、怀、么、唔、云、堂、黄、愁、强、凡
上声	我、恍、此、好、感、也、有、纸、把、写、遣、了、闪、鼓、展、稳、缕
阴去声	妙、后、对、寄、舍、卷、在
阳去声	自、荡、序、免、事、闷、沉、睡、念、静、动、万、缠
阴入声	见、觉、惚、笔、作、一、压
阳入声	什、独、欲

表 10

阴平声	家、贞、官、公、君、都、非、忠、之、皆、中、身、端、装、他、兄、只
阳平声	奴、娥、闹、蒙、娘、谁、流、于、臣、为、仇、男、藏、模、来、时
上声	小、选、母、我、主、想、死、可、子、有、耻、此、取、了、把、匕、首、假、指、与、虎、理
阴去声	费、幼、篆、笑、些、竟、个、报、女、做、在、近、闯、寇、刺、弟、配
阳去声	氏、自、字、命、侍、父、那、义、事、又、样、望、巨、待、到、罢
阴入声	国、杀、一、骨、雪、得
阳入声	入、服、贼、夺、肉、没

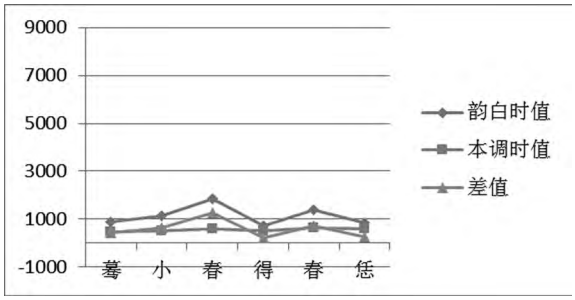


图 1 :《牡丹亭・惊梦》韵白、中州韵句首字时长比较图（单位：ms；图中“本调”指“中州韵”，下同）

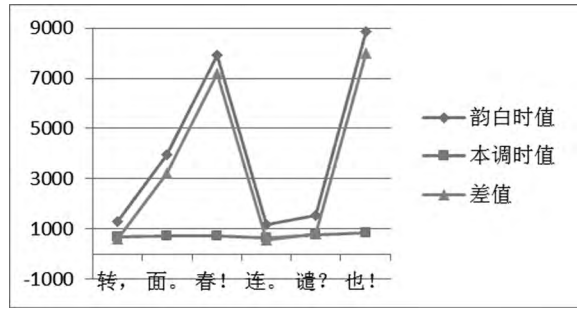


图 2 :《牡丹亭・惊梦》韵白、中州韵句尾字时长比较图（单位：ms）

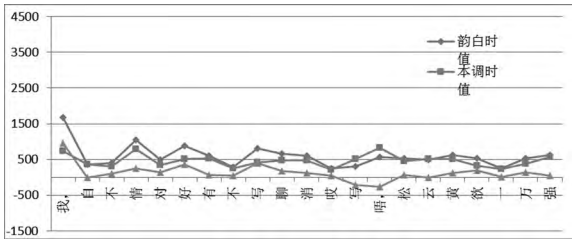


图 3 :《玉簪记・偷诗》韵白、中州韵句首字时长比较图（单位：ms）

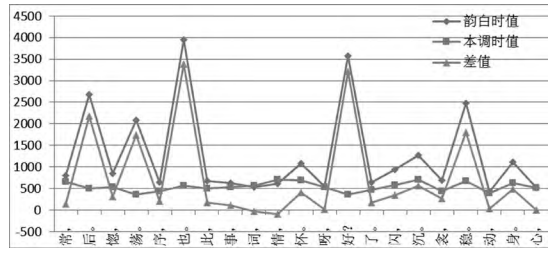


图 4 :《玉簪记・偷诗》韵白、中州韵句尾字时长比较图（单位：ms）

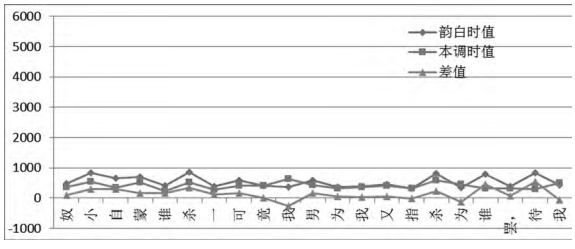


图 5 :《铁冠图・刺虎》韵白、中州韵句首字时长比较图（单位：ms）

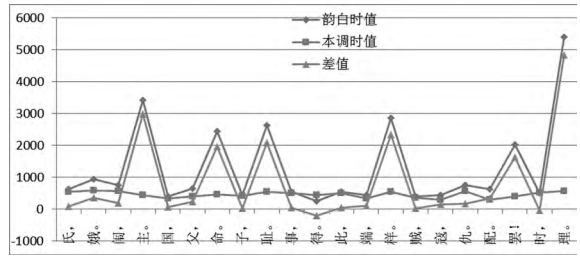


图 6 :《铁冠图・刺虎》韵白、中州韵句尾字时长比较图（单位：ms）

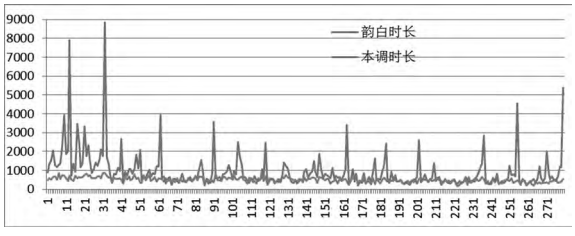
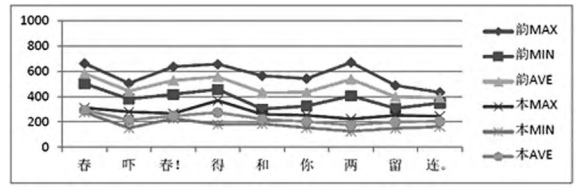
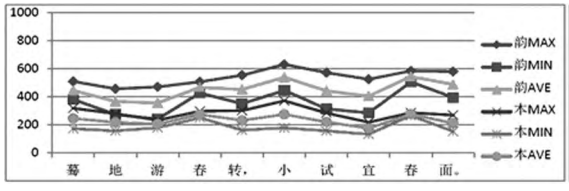


图 7:280 个字的韵白与中州韵时长曲线图（纵轴单位：ms，横轴：字数）





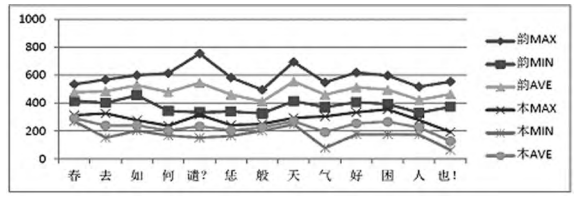


图 8 :《牡丹亭・惊梦》韵白、中州韵音高曲线图

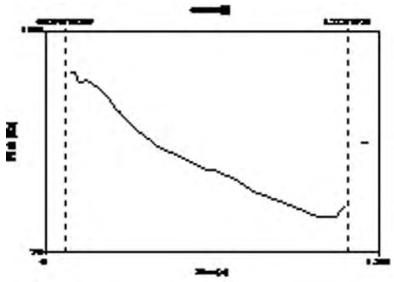


图 9：“报”字音高曲线图

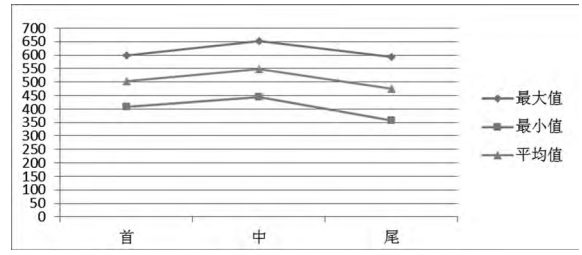


图 10：韵白首字、中字、尾字音高比较图（单位 Hz）

注：此图统计的是包含有 3 个字以上的短句 46 个，包含 2 字及其以内的短句忽略不计。

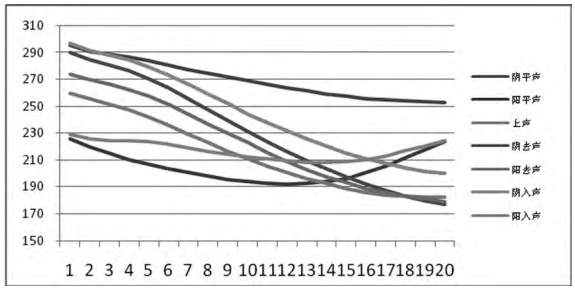


图 11：七个调类均值曲线

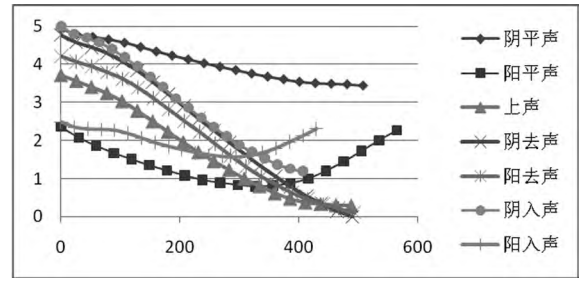


图 12：昆曲念白中州韵五度值格局图

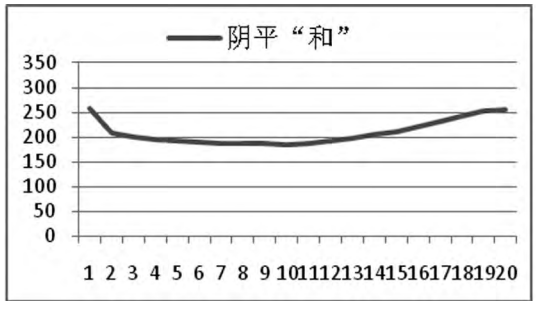


图 13：阴平声“和”声调曲线

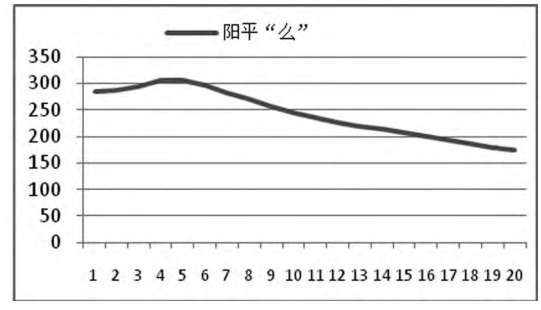


图 14：阳平声“么”声调曲线

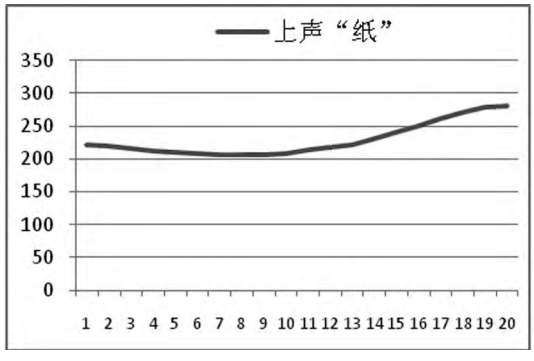


图 15：上声“纸”声调曲线

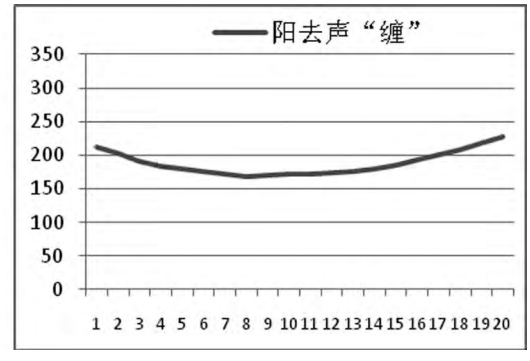


图 16：阳去声“缠”声调曲线

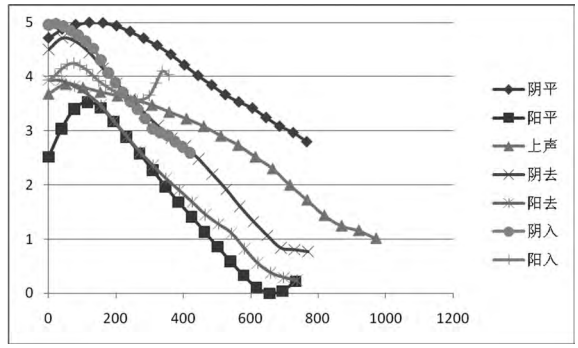


图 17：韵白字声调五度值格局图

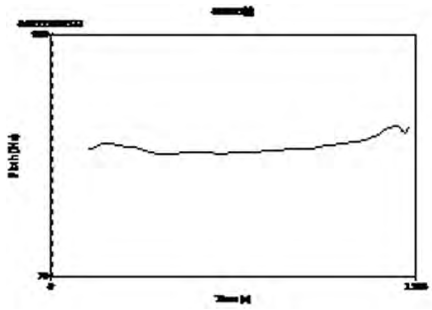


图 18：“弯头提尾型”（一）

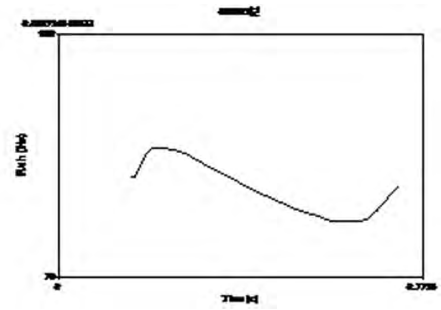


图 19：“弯头提尾型”（二）

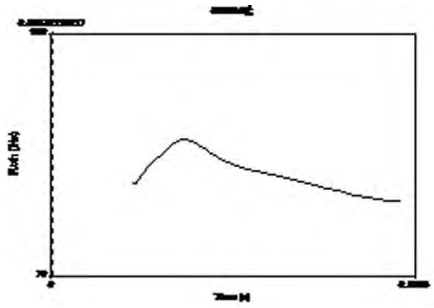


图 20：“上折线型”（一）

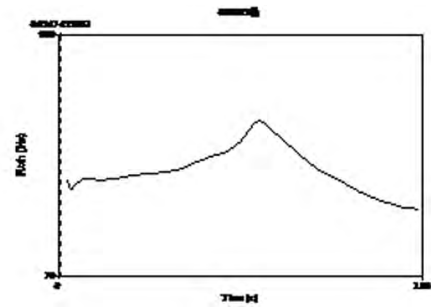


图 21：“上折线型”（二）

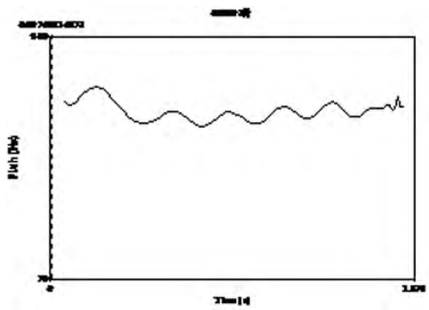


图 22：“波浪型”（一）

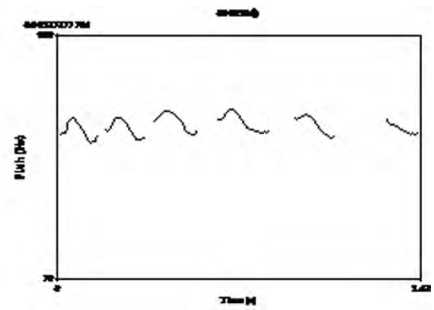


图 23：“波浪型”（二）

DOI:10.14084/j.cnki.cn62-1188/n.2014.03.007

第35卷总第95期 西北民族大学学报(自然科学版) Vol. 35,No. 3
2014年9月 Journal of Northwest University for Nationalities(Natural Science) Sep,2014

《格萨尔》说唱的声学分析

索朗德吉,孙 婷,达哇彭措

(西北民族大学 中国民族语言文字信息技术重点实验室,甘肃 兰州 730030)

[摘 要] 文章选取了《格萨尔》说唱中<珠鸠>曲调的语音信号,运用实验语音声学的分析方法,对语音进行了信号标注和参数提取,通过基频、能量等声学参数来分析《格萨尔》说唱当中,<珠鸠>曲调的基本发音原理和说唱技巧,以数字化方式更为深入和直观地研究《格萨尔》说唱的风格特色.

[关键词] 《格萨尔》说唱;珠姆;声学分析

[中图分类号] J607.215 [文献标识码] A [文章编号] 1009-2102(2014)03-0029-06

藏族诗史《格萨尔》是世界上最长的史诗,它以口耳相传的方式流传至今. 同时,《格萨尔》也被世人看作是古代藏民族生活的百科全书,其中包括了藏民族历朝历代的征战、社会的变革、区域版图的变迁、民族间的交往、意识形态、宗教信仰、道德观念、风俗习惯、文化艺术、天文历算等丰富的信息内涵. 因此开展关于《格萨尔》诗史的整理和研究工作,不但有利于我们更好地继承和发扬民族文化艺术,也是我们进一步了解、探究不同历史时期藏民族社会发展变迁的重要途径.

作为一门独立的新兴学科,当前格萨尔学研究已从文学、史学的一般性理论探讨、故事阐释开始深入到了音乐、民俗、宗教、经济、伦理学等多个方面,并取得了一定的研究成果. 当前现有的研究成果主要集中在两个方面:在文学历史领域如《格萨尔王全传》(降边嘉措、吴伟1997)、《<格萨尔>中的古代藏族社会及其文化》(王兴先1992)、《<格萨尔>生命学思想论》(韩伟、庞泽华2008);在音乐文化领域如《<格萨尔>音乐多元结构》(扎西达杰1996);《<格萨尔>之<辛丹内让>音乐初析》(乔迁2004);《试论<格萨尔>的唱腔特点及结构特色》(李晓玲2008). 目前,涉及到语言学领域的《格萨尔》相关研究还比较少,运用现代语音设备,结合现代技术条件,从声学角度对《格萨尔》演唱的发声原理和演唱风格的数字研究尚处于空白状态.

神将(擦向·丹玛强察)和王妃(森贾母·珠姆)是《格萨尔》故事中的重要人物,本文选取了《格萨尔》曲调中专门用来歌唱这两位人物的<珠鸠>曲调(<塔拉珠鸠>和<九狮珠鸠>),希望采用现代语音研究设备和研究技术方法,通过对提取的相关声学参数进行分析,发现《格萨尔》<珠鸠>曲调中的发声特色,从而由声学角度对《格萨尔》<珠鸠>曲调的发声原理和演唱风格进行数字化的研究保存工作.

1 实验研究方法

1.1 说唱背景

《格萨尔》是说唱艺术,即有说又有唱,其中所有人物均以歌唱的形式呈现. 艺人们根据不同人物与

[收稿日期] 2014-08-20

[基金项目] 国家自然科学基金—基于依存关系的藏文语义角色标注研究(61363057);国家自然基金地区基金基于动态腭位的藏语发音生理模型研究(61262052)和社科基金重大项目(10&ZD125).

[作者简介] 索朗德吉(1985—),女(藏族),西藏日喀则人,硕士研究生,主要从事藏文信息处理及语言学方面的研究.

曲牌的关系，配以不同的说唱曲调．

民间认为《格萨尔》说唱中约包含 80 至 100 多种不同的曲调，在《玉树藏族民间歌舞音乐大全之四》(代尕，2014) 中，关于曲调种类的调研数字为 139 种．＜珠鸠＞是藏文(འབྲུག་འགྱུར།)的音译，汉语翻译意思就是六变调．＜珠鸠＞调在《格萨尔》说唱中运用较多，且流传地域也广，主要用来叙述人物．在整个《格萨尔》说唱艺术当中，运用＜珠鸠＞调说唱的人物有十几个，这当中分别采用的＜珠鸠＞调式也各有不同，其中＜塔拉珠鸠＞(ཐ་ལ་འབྲུག་འགྱུར)主要是用来描述神将擦向·丹玛强察(ཚ་ཞང་འདན་མ་བྱང་ཁྲ ts^{h} ɕaŋ ^{n}dan ma ɕ^{w}aŋ ʈʂ^{h}a)的一种唱调，＜九狮珠鸠＞(དགུ་སེང་འབྲུག་འགྱུར)主要是用来描述格萨尔王的王妃森贾母·珠姆(སེང་ལྕམ་འབྲུག་མོ sаŋ tɕam ^{n}ʐuk mo)的一种唱调．本文选取了＜塔拉珠鸠＞和＜九狮珠鸠＞曲调进行比较分析．

གླུ་ཐ་ལ་ལ་མོ་ཐ་ལ་ལེན།
^{ɦ}lə a la la mo a la len
གླུ་ཨ་ལ་ལ་མོ་ཨ་ལ་ལེན།།
^{ɦ}lə t^{h}a la la mo t^{h}a la len
གླུ་ཐ་ལའི་ལེངས་གསུམ་ལེན་ནས་ཡོང་།།
^{ɦ}lə t^{h}a li laŋ ^{ɦ}səm len ne joŋ
མི་ང་དང་ང་ངོ་མ་ཤེས་ན།།
mə ŋa ^{n}dʑa ŋa ŋo ma xi na
ང་ཚ་ཞང་འདན་མ་བྱང་ཁྲ་ཡིན།།
ŋa ʦ^{b}a ɕaŋ ^{n}dan ma ɕ^{w}aŋ ʈʂ^{h}a jən
ང་སྐྱིན་པོའི་མདའ་ལ་གྲུབ་པ་ཐོབ།།
ŋa ^{h}tɕen pi ^{n}da la ʈʂəp pa t^{h}op
གླུ་གོན་ན་བའི་བདུད་རྩི་རྩོས།།
^{ɦ}lə ko na ^{ɦ}na wi d^{w}ət tsə ɕ^{w}i
ཡང་མ་གོ་གླུ་ལ་འབྲེལ་བ་མེད།།
Jaŋ ma ko ^{ɦ}lə la ndʐel wa me

图 1 ＜塔拉珠鸠＞的藏文歌词及国际音标

ཨོ་ན་ཡང་།
o na jaŋ
གླུ་ཨ་ལ་ལ་མོ་ཨ་ལ་ལེན།།
^{ɦ}lə a la la mo a la len
གླུ་ཐ་ལ་ལ་མོ་ཐ་ལ་ལེན།།
ɦlə t^{h}a la la mo t^{h}a la len
ཤར་གཡུ་ལོ་བཀོད་པའི་ཞིང་ཁམས་ནས།།
xar ^{ɦ}jə lo kot pi ɕaŋ k^{h}am ne
མ་ཚེ་འཛིན་ལྷ་མོས་གླུ་སྣ་སྔོངས།།
ma lse ^{n}dən ɬa mi ^{ɦ}lə n̥ a ʈʂoŋ
ང་སེང་ལྕམ་འབྲུག་མོ་ཧོར་ལ་འཆེར།།
ŋa saŋ tɕam ^{n}dʐuk mo ho la tɕ^{h}er
ང་རྒྱལ་པོར་མ་ཐུག་ལོ་དགུ་འགོར།།
ŋa ^{ɦ}dʑa wo ma t^{h}ək lo ^{ɦ}gə ^{n}gor
ང་སེང་ལྕམ་འབྲུག་མོར་ཞེན་ཆགས་མང་།།
ŋa saŋ tɕam ^{n}dʐuk mo ɕan tɕhak maŋ

图 2 ＜九狮珠鸠＞的藏文歌词及国际音标

1.2 信号采集

实验以甘肃省玛曲格萨尔口头传统研究基地为调研点，该地区人口流动和迁移相当缓慢，保留着最初的生产、生活方式，并且仍然保存着古老而原始的格萨尔演唱形式.

实验选取1名格萨尔专业硕士研究生作为发音人：喇嘛扎西，男，27岁. 发音人有熟练的演唱技巧，能够运用各种不同的音位，熟练地演唱几十首不同曲调的《格萨尔》曲目.

录音设备主要由喉头仪(EGG)、外置声卡、高质量的领夹式麦克风、笔记本电脑、调音台组成. 采集信号时，用Audition 3.0软件采集语音信号，信号采样频率设定为48kHz，分辨率为16 bit，录音格式设定为无压缩PCM编码的“.wav”文件格式. 声学分析采用Praat软件，根据不同的音段特征，对录音语音进行切分、打标记和声学分析，提取基频、能量、时长等方面的声学参数来分析《格萨尔》说唱中的发音特点和发音方式.

图3~4是<塔拉珠鸠>和<九狮珠鸠>中的典型唱调波形图时长分别是9s和12s.

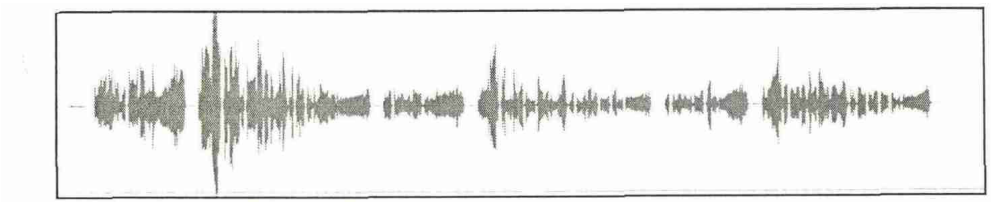


图3 <塔拉珠鸠>的波形图

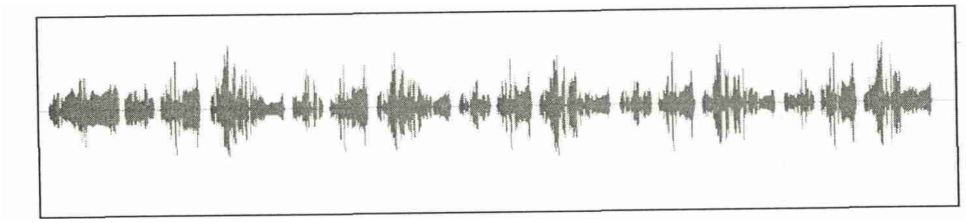


图4 <九狮珠鸠>的波形图

1.3 参数设置

在本研究中，我们主要基于以下的声学参数来探究《格萨尔》<珠鸠>曲调的演唱声学特点.

1.3.1 基频

基频表示单位时间内声带振动的次数，单位是赫兹(Hz). 一般情况下，通过声带的拉紧变薄，可以提高音高，反之，声带松弛会降低音高. 人耳对声音调子高低的主观感觉称之为音高，音高主要取决于频率的高低. 频率低的调子给人以低沉、厚实、粗犷的感觉；频率高的调子给人以高亢、明快、尖刻的感觉.

1.3.2 能量

能量是声音的另一个重要特性，单位是分贝(dB). 人们可以用它来衡量声音强度的大小. 人耳对声音音高的另一主要感知途径即始于声音的强弱. 声波的能量越大，声音的强度就越大，越容易为人耳所感知，能量大的调子给人以热烈、兴奋、积极的感觉；能量低的调子给人以低沉、温婉、安逸的感觉.

1.3.3 二维频谱

声音信号是周期信号，任何周期信号，都是由若干个不同频率成分(一般用正余弦信号表示)的分音组成. 每个频率成分均有一定的振幅和相位. 如果将它们的振幅或相位按次序加以排列，纵坐标表示谐波的振幅，横坐标表示频率，就得到二维频谱图. 二维频谱图将我们对声音信号的研究从时域的范畴扩展到了频域的范畴.

1.3.4 三维语谱图

语谱图是对一段语音在频域各个频段的能量的一种显示方式. 在语谱图纵向代表频率，单位是千赫兹(KHz)，横向代表时间，单位是毫秒(ms)，颜色深浅代表信号的幅度大小. 依据三维语谱图我们可以来同时观测声音信号的频率、幅度和时间等物理参量这三者之间的动态关系，从而了解声音信号的本质.

2 声学分析

史诗《格萨尔》的唱腔是通过一定的音乐调式反复演唱来展开故事，推动情节发展的，因此，在本文中我们分别节选了两首 < 珠鸠 > 曲调的代表片段来进行声学研究 .

2.1 基频分析

基频的高低变化代表声带振动的快慢和演唱者在演唱过程中对气流速率的控制程度 . 通过对提取的频率参数进行分析(见图 5 ~ 6) .

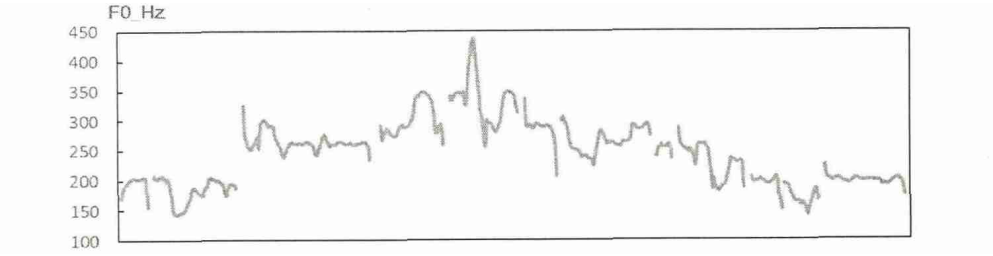


图 5 < 塔拉珠鸠 > 的基频图

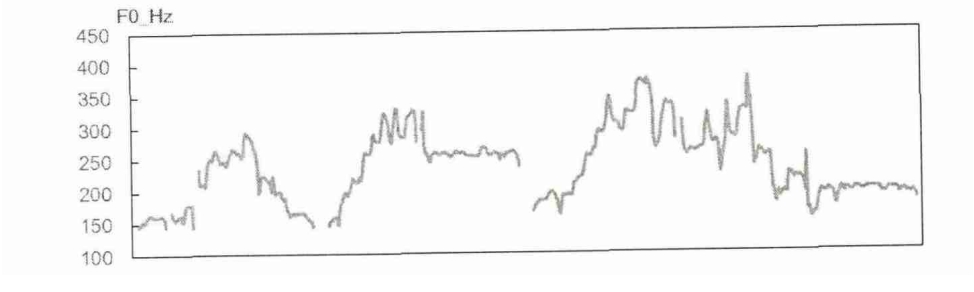


图 6 < 九狮珠鸠 > 的基频图

从图 5 ~ 6 中我们可以看到:

1) 用来歌唱男性人物的曲调 < 塔拉珠鸠 > ，其基频变化范围在 107 ~ 439 Hz 之间，而用来歌唱女性人物的曲调 < 九狮珠鸠 > ，其基频变化范围则在 143 ~ 377 Hz 之间 .

2) 观察两首曲调的基频变化特点，< 塔拉珠鸠 > 曲调的基频高低起伏变化较大，而 < 九狮珠鸠 > 曲调的基频变化较为舒缓，因此，我们可以从声学角度看到，在运用 < 珠鸠 > 曲调的演唱过程中，歌唱者巧妙地运用演唱技巧，灵活地控制了声音的基频变化范围，用丰富 < 珠鸠 > 曲调演唱形式，生动地展现不同人物形象 . < 塔拉珠鸠 > 曲调表现的是神将擦向 · 丹玛强察在阵前叫威时唱起来的，基频的高低变化差距较大，使听者很好地感受到了神将威风凛凛，势不可挡的气势 . < 九狮珠鸠 > 的曲调表现的是格萨尔的王妃森贾母 · 珠姆在敌国被扣押作人质时的情景，基频变化规律舒缓，使听者感受到了一个痴情女子的忧伤 .

2.2 能量分析

能量的高低变化代表声音振动的强弱和演唱者在演唱过程中对气流量多少的控制 . 通过对提取的能量参数进行分析(见图 7 ~ 8) .

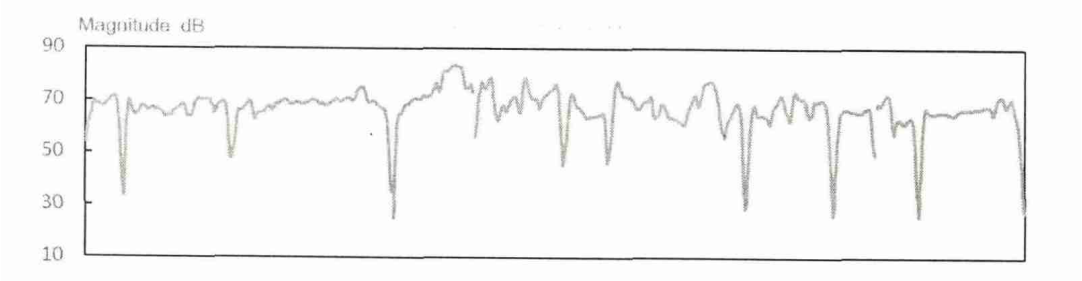


图 7 < 塔拉珠鸠 > 的能量图

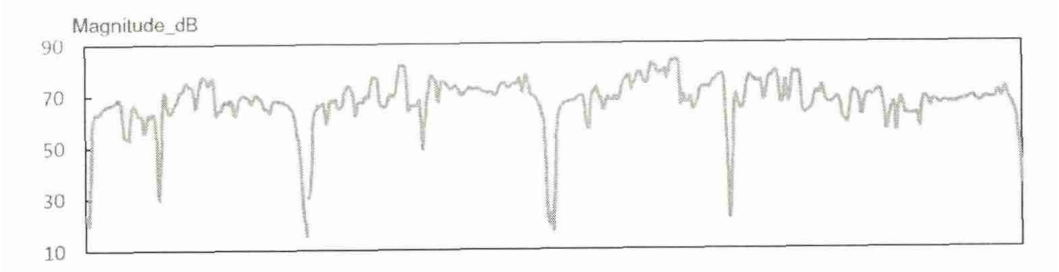


图 8　＜九狮珠鸠＞的能量图

从图 7 ~ 8 中我们可以看到:

1) ＜塔拉珠鸠＞和＜九狮珠鸠＞曲调的能量变化范围基本一致,只是后者的能量下限相比更低一些,其中曲调＜塔拉珠鸠＞的能量变化范围在 25 dB ~ 83dB 之间,曲调＜九狮珠鸠＞的能量变化范围在 15dB—83dB 之间,由此我们可以发现,尽管＜珠鸠＞曲调变化方式较多,但其声音的能量范围均处固定的范围之内.

2) 对比两首曲调的平均能量,差异较为显著,其中＜塔拉珠鸠＞的平均能量为 41 dB,＜九狮珠鸠＞的平均能量为 29 dB.《曲艺音乐概论》中认为史诗《格萨尔》的演唱中十分注重唱腔的合理性、准确性,所创唱腔尽可能正确地为故事内容的叙述服务,以便让人们在似说似唱的表演中清楚地了解故事内容. 通过对能量参数数值的分析我们可以发现,演唱艺人在充分理解并掌握语音音调对旋律的影响后,演唱时,即使是同一曲调的反复演唱,也能借助对声音强弱的良好控制能力来生动叙述故事情节,刻画人物形象.

2.3　频谱分析

频谱图反映声音信号中各个谐波的能量强弱程度,图 9 - 10 为两首曲调的三维语谱图,其中横轴表示时间,纵轴表示频率,图形色彩的灰度表示某一时刻某一频率分量的振幅大小. 色彩(或灰度)浓,就表示能量集中,振幅达,声音强;反之则相反.

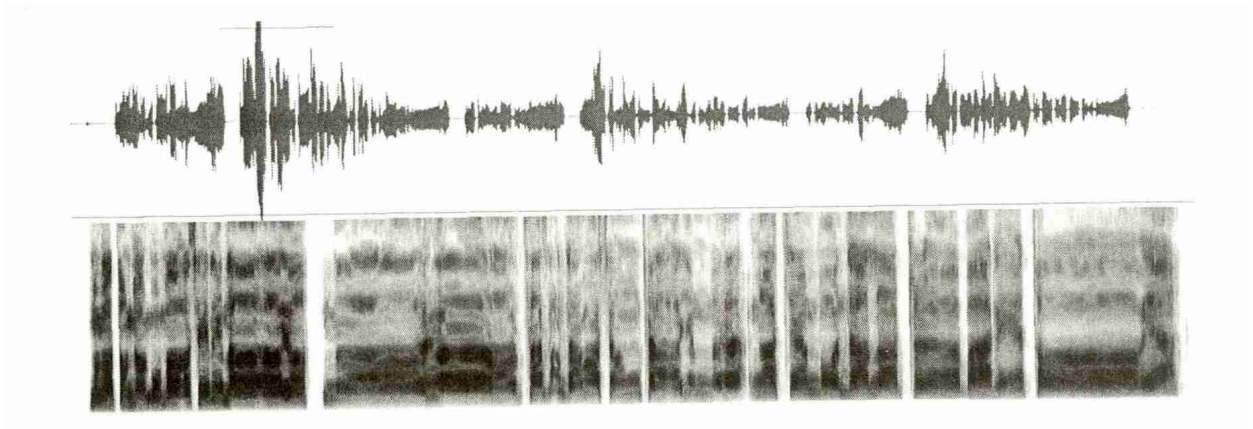


图 9　＜塔拉珠鸠＞的音频图和频谱图

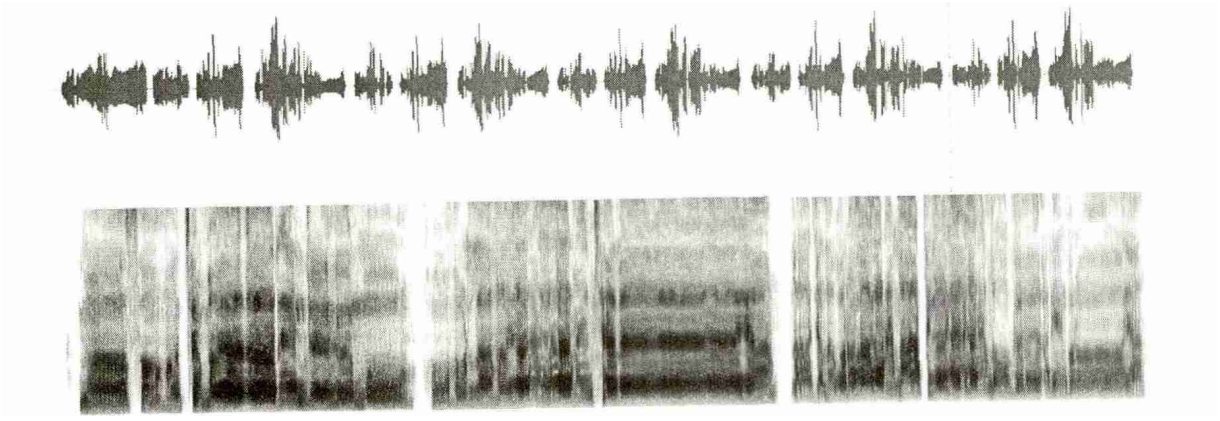


图 10　＜九狮珠鸠＞的音频图和频谱图

图 9 ~ 10 为两首曲调选段的频谱图，从图中我们可以看到:

1) <塔拉珠鸠>的曲调随声音波形的变化，整体声音能量较强，F1、F2、F3、F4 在语图上较为显著，低频共振峰 F1 为 591 Hz，F2 为 1 577 Hz; 高频共振峰 F3 为 263 4Hz，F4 为 3 530 Hz.

2) <九狮珠鸠>的曲调随着声音波形的变化，整体声音能量较弱，只有 F1、F2 在语图上较为显著，其中 F1 为 580 Hz，F2 为 1 390 Hz.

《格萨尔》歌唱艺术吸取了很多民族歌曲的演唱特点，形成了自己独特的艺术表演风格，常常体现在歌唱共振峰等声学参数上．共鸣是歌唱艺术中的重要表现手段，它能使人的嗓音变得或清脆明亮、或低沉温婉．<塔拉珠鸠>和<九狮珠鸠>同属<珠鸠>调式，但由于它们所描绘的内容不同，在声学上声音的共振峰特点也有所区别．由此，我们可以发现在《格萨尔》演唱中，歌者能够依靠高技巧的气息控制力，呈现出不同共鸣效果，并以此来突出史诗中的故事情节和人物个性．

3 结语

本文选取了《格萨尔》说唱中<珠鸠>曲调的两首典型歌唱曲目，基于声学角度对两首曲调的演唱风格进行了研究分析，发现在<珠鸠>曲调的演唱中，声音的音高变化会因所要呈现的不同人物或舒缓或激烈; 同时，我们发现，不同<珠鸠>曲调的声音能量变化范围较为一致(10 ~ 90 dB) ，但由于所要塑造的不同人物形象，艺人也能借助对声音强弱的良好控制能力，在同一曲调的反复演唱中，通过完全不同的声音强弱变化形式来生动叙述故事情节、刻画人物形象．

随着社会的快速发展，老一辈说唱艺人在日益减少，青年艺人的培养又存在诸多局限，古老的《格萨尔》说唱艺术在传承方面面临着很多前所未有的挑战．目前，学者们搜集整理出的《格萨尔》说唱曲调已有一百多种，数量庞大，但它们各具特色的演唱方式都是《格萨尔》说唱艺术不可缺失的组成部分．在计算机技术高速发展的时代，传统、单一的“口传心教”民族文化传承方式已不能跟上现代社会发展的脚步，如何借助现代技术，全面采用数字化的方式完整有效地记录《格萨尔》说唱艺术，是当前和未来我们在保护和继承民族文化工作中亟待解决的问题．

参考文献:

[1] འཛོར་པ་དོན་གྲུབ། ཧོར་གླིང་གཡུལ་འགྱེད། ཀན་སུའུ་མི་རིགས་དཔེ་སྐྲུན་ཁང་། 2014.3

[2] བྱང་ཆུབ་སེམས་དཔའ་ཚོས་ཀྱི་བློ་གྲོས། [Z]. འཇང་གླིང་གཡུལ་འགྱེད། བོད་ལྗོངས་མི་དམངས་དཔེ་སྐྲུན་ཁང་། 1983.

[3] ནང་སོ་བློ་བརྟན། ཇ་རོང་འབྲུ་རྫོང་། [Z]. བོད་ལྗོངས་མི་དམངས་དཔེ་སྐྲུན་ཁང་། 1986.

[4] དཀོན་མཆོག་ཚེ་བརྟན། འཛམ་གླིང་སྤྱི་བསང་། [Z] ཀན་སུའུ་མི་རིགས་དཔེ་སྐྲུན་ཁང་། 1980.

[5] 降边嘉措，吴伟编．格萨尔王全传(上下) [M]．北京: 作家出版社，1997. 8.

[6] 代尕．玉树藏族民间歌舞音乐大全之四 [M]．北京: 中央民族大学出版社，2014. 7.

[7] 方华萍，李永宏．蒙古长调《圣》韵律特征声学研究 [J]．西北民族大学学报(自然科学版) ，2012. 6.

[8] 扎西达杰．《格萨尔》音乐多元结构 [M]．拉萨: 西藏艺术研究期刊社，19964: 36 – 39.

[9] 卢国文．史诗《格萨尔》说唱音乐艺术性与社会功能 [J]．中央民族大学学报，1994，3: 73 – 78.

[10] 柯林．格萨尔》说唱音乐结构特征 [J]．中央民族大学学报，1998，(4) : 69 – 72.

[11] 乔迁．格萨尔》之《辛丹内讧》音乐初析 [D]．上海音乐学院，2004. 4.

[12] 李晓玲．试论《格萨尔》的唱腔特点及结构特色 [J]．西北民族大学学报(哲学社会科学版) ，2008，2(20) : 99 – 101.

[13] 马海龙，高璐，于洪志．几种不同基频提取算法的比较研究 [J]．西北民族大学学报(自然科学版) ，2014，12: 59 – 63.

[14] 徐慧，胡阿旭，于洪志．浅析蒙古族短调的声学特性 [J]．西北民族大学学报(自然科学版) ，2014，9: 44 – 48.

[15] 吴宗济，林茂灿．实验语音学概要 [M]．北京: 高等教育出版社，1989.

[16] http: //baike. so. com/doc/2539543. html.

[17] http: //www. docin. com/p – 100962715. html.

Advanced Materials Research Vols. 926-930 (2014) pp 1791-1794
© (2014) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMR.926-930.1791

An Acoustics Study on Prepositional Consonant [h] in Xiahe Tibetan

Ting Sun[a], Hongzhi Yu[b], Yasheng Jin[c]

Key Lab of China's National Linguistic Information Technology, Northwest University for Nationalities, LanZhou, China

[a]sukkyang_fly@sina.com, [b]yuhongzhi@hotmail.com , [c]xbgsj@xbmu.edu.cn

Keywords: Tibetan, consonant clusters, acoustics analysis

Abstract. The voiceless consonant [h] can be collocated with the voiceless plosives, voiceless affricates, voiceless fricatives, nasals, laterals, etc. This paper makes a study on different collocations with duration and spectrum, with the result , it is can be seen that, the differences are greater among the acoustic characteristics of [h]as initial consonant and prepositional consonant, and that present the different ways of pronunciation.

Introduction

In Tibetan the Initial Consonant Clusters is a unit in a syllable, which is constituted by initial consonant and basic consonant. Initial consonant clusters always related with the tone's classification, the new pronunciation's happening, the devocalization in sonant, and the monophthong tuning into diphthong etc [1]. There are two types of the studies on consonant clusters in Tibetan, based on the research method and aim. The one is according to the different Tibetan dialects, analyzing the consonant clusters' characters, structures, and types; making a description of the consonant clusters in different Tibetan dialects, and doing comparison between the Tibetan dialects and fthe Written Tibetan, in which the changes of consonant clusters can be found [2, 3]. The other is to study the consonant clusters' acoustics characters which based on the method of experimental phonetics [4, 5].

The typical language of Anduo pastoral area is Xiahe Tibetan, and which contains too many initial con-sonant clusters. In general, the combination between prepositional consonant and basic consonant will be affected themselves. It is like that, when both the prepo-sitional consonant is voiceless, the basic consonant also is voiceless; when the prepositional consonant is voiced, the basic consonant also is voiced. The consonant [h] always can be the prepositional consonant and combine with different basic consonants.

Study method

Types. In Xiahe Tibetan, it is very often that the consonant clusters' prepositional consonant be as [h]. The prepositional consonant [h] come from the pro-posed word like 'ག', 'བ', 'ད', and the head word like 'ར', 'ལ', 'ས' in Tibetan, with such kind combination the original forms of consonant clusters can be more simple. The voiceless consonant [h] can be collocated with the voiceless plosives, voiceless affricates, voiceless fricatives, nasals, laterals and semivowels.

Sample collection. The experimental equipment, for the hardware, it contains neck-style microphone, mixing console, external sound card and electronic larynx diagnostic apparatus; for the software, it contains Audition1.5, Praat, and some written programs. The speaker is a Tibetan teacher. The samples' frequency is 22050 Hz, precision is 16bits, and the file format is.wav.

Analysis of consonant [h]

[h]comes from the Tibetan initial consonant 'ཧ', in pronouncing, the air flows over the glottis, and a little skin friction drag is created. Meanwhile, the shape of the mouth changed as the vowel which is after the initial consonant. /h/ is voiceless consonant, the vocal cords do not vibrate. Table 1 is about the example words with [h] of Xiahe Tibetan. The duration is the initial consonant [h]'s duration.

Take 'ཧྭ' as an example, to describe the initial consonant [h]'s acoustic characteristics. In pronouncing, first, the vocal cords is opening, and the opening of the mouth is large; then from [h] to [a], the vocal cords is closing up slowly.

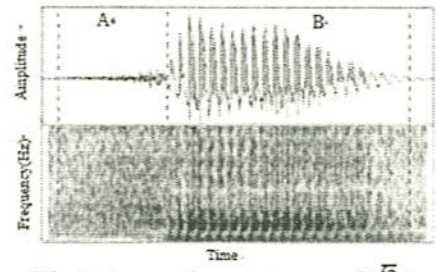


Fig.1 Acoustic spectrum of 'ཧྭ'

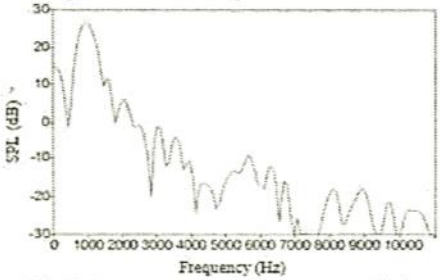


Fig2 Spectrum of consonant [h]

Table 1.The example words with initial consonant [h] of Xiahe Tibetan

Tibetan	Latin transliteration	Chinese	International Phonetic Alphabet	Duration (ms)
ཧྭ	hwa	衣服 领口	ha	100
ཧིན	hin	哼	hən	90
ཧུར	hur	凝视	hər	70
ཧེའུ	heu	侯爵	hi	80

The figure 1 is an acoustic spectrum of 'ཧྭ', the upper half of it is the speech waveform; the lower of it is 3-D spectrogram, and its frequency is 1Hz-5500Hz. In the figure, the initial consonant [h]'s duration is marked, and part A is for initial consonant [h], part B is for final [a]. Figure 2 is the spectrum, which is of the middle point of the consonant [h]'s duration. From Figure 1 and Figure 2, it is can be got as the following:(1) The initial consonant [h]'s duration is 100ms, and the final [a]'s is 200ms. (2) As [h] is the voiceless glottal stops, its speech signal is irregular and no periodicity, and shows obvious noise in the sonogram.(3) The spectrum energy of [h] is concentrated at the lower frequency region, especially the SPL is 27dB with the frequency is 1000Hz. As the frequency's growing, the energy attenuation is quick, from 1000Hz to 5000Hz, the energy attenuation is about 40dB.

Analysis of consonant [h] + plosives

Prepositional consonant [h] can combine with the voiceless plosives ([t], [k]) or voiceless affricates ([ts], [tɕ], [tʂ]) to be a consonant cluster, as table 2. The Du-ration 1 is of the prepositional consonant [h], and the Duration 2 is of the mute segment.

From Figure 3, 4, 5, it is can be got as the following:(1) In pronouncing, the vocal cords is opening, the shape of the mouth is small, and the airflow is blocked in the closing position of the basic consonants, so it is can be seen that the power has grown at the end of the prepositional consonant [h]'s duration, meanwhile, the spike shows in the sonogram. (2) The prepositional consonant [h]'s duration is 80ms. The mute segment's duration is 120ms, and it is also the on-glide duration of the basic consonants, so it is easy to be heard of the mute segment between the prepositional consonant and basic consonants. The part C's duration is 18ms. (3) The spectrum energy of prepositional consonant [h] is concentrated 1500Hz, 3500Hz, 5000Hz, 7000Hz and other regions. SPL is 0-10dB.

Table 2. The example words of basic Consonant as plosives and affricates

Basic consonant	Tibetan	Latin transliteration	Chinese	International Phonetic Alphabet	Duration1 (ms)	Duration2 (ms)
t	ལྟེ	lte	肚脐	hte	70	110
k	རྐ	rka	水渠	hka	60	130
ts	རྩྭ	rtswa	草	htsa	70	125

tɕ	ལྕེ	lce	舌头	htce	85	120
tʂ	སྤྲ	spra	猿	htʂa	80	120

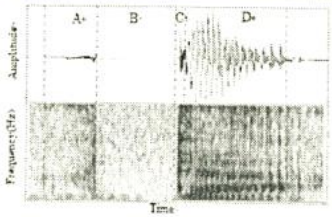


Fig.3 Acoustic spectrum of 'སྤྲ'

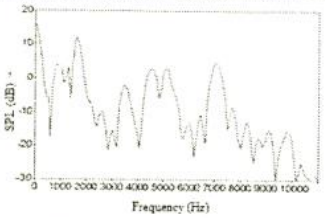


Fig.4Spectrum of prepositional consonant [h]

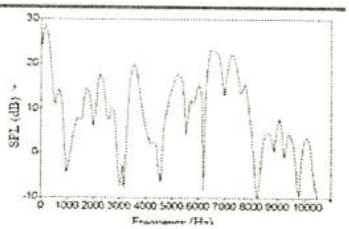


Fig.5 Spectrum of basic consonant affricates [tʂ]

There is not an obvious energy attenuation from low frequency to high frequency.(4) Comparing Figure 4 and Figure 5, the spectrum energy of 'སྤྲ''s affricates is distribution uniformity, under 7000Hz, the power is from 10 dB to 20 dB, which is more than the prepositional consonant [h]'s.

Analysis of consonant [h] + nasals

Prepositional consonant [h] can combine with the nasals, [m], [n], [ɳ], [ŋ] as table 3.

Table3. The example words of basic Consonant as nasals

Basic consonant	Tibetan	Latin transliteration	Chinese	International Phonetic Alphabet	Duration (ms)
ŋ	ལྔ	lnga	五	hŋa	100
n	སྣོ	rno	钢	hno	120
m	སྨོ	rmo	犁地	hmo	115
ɳ	རྙ	rnya	借	hɳa	130

Take 'རྔ' as an example, to describe the prepositional consonant [h]'s acoustic characteristics. Figure 6 is an acoustic spectrum of 'རྔ', part A is for prepositional consonant [h]. Part B is for the basic consonant [ɳ]'s duration. Part C is for final [a]. Figure 7 is the spectrum, which is of the middle point of the prepositional consonant [h]'s duration.

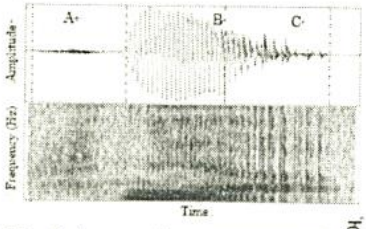


Fig.6 Acoustic spectrum of 'རྔ'

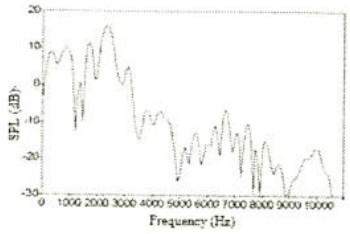


Fig.7 Spectrum of prepositional consonant segment[h]

From Figure6 and 7, it is can be got as the following :(1) In pronouncing, the airflow in the mouth has disappeared gradually at the end of part a, but the airflow in the nose has grown from zero, so it can be seen that there is a transition before the basic consonants, and which is different from the mute segment of plosives. Sometimes such a mute segment after part A is not obvious as the speech rate is faster.(2) The prepositional consonant [h]'s duration is shorter, it is 130ms, but part B's duration is 300ms and part C's duration is 250ms.(3) The spectrum energy of prepositional consonant [h] is stronger under 3000Hz, and above 3000Hz, it is weaker.

6Analysis of consonant [h] + fricatives

Prepositional consonant [h] can combine with the fricatives [ɕ], [s]; but for the combination with [s], [h] is always to be the loss of sound. Table 4 is about the ex-ample words with consonant [h] + fricatives.

Table4. The example words of basic Consonant as fricatives

Basic consonant	Tibetan	Latin transliteration	Chinese	International Phonetic Alphabet	Duration (ms)
c	གཤགས	gshags	诉讼	hɕaχ	40
s	གསག	gsag	隐秘	hsaχ	45

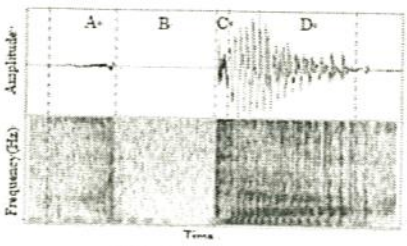


Fig8. Acoustic spectrum of 'གཤགས,

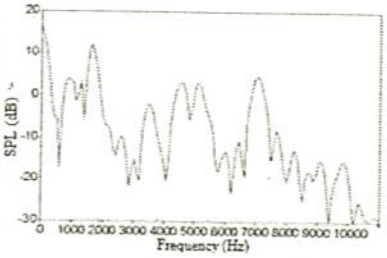


Fig.9 Spectrum of prepositional consonant [h]

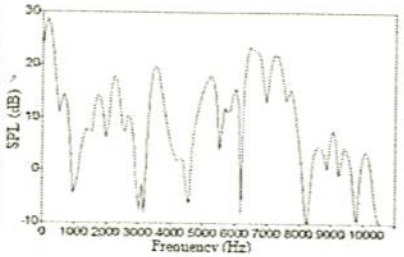


Fig.10 Spectrum of basic consonant fricatives [ɕ]

From Figure8, 9 and 10, it is can be got as the following:(1) As both of the prepositional consonant [h] and basic consonant are fricatives, the transition between them is continuous and natural.(2) The prepositional consonant [h]'s duration is shorter, it is 40ms, but part B's duration is 170ms and part C's duration is 110ms.(3) The spectrum energy of prepositional consonant [h] is stronger as a whole, and the main focus in 4000Hz-8000Hz.

Conclusion

This paper focus on the prepositional consonant [h]'s acoustic characteristics of Xiahe Tibetan, the results are as the following:(1) There is an obvious mute segment between the prepositional consonant and basic consonant in Consonant [h] + plosives, which is 100-130ms; by contrast, the duration be-tween the prepositional consonant and basic consonant is closer. (2) The collocation of [h] + fric-atives is the shortest, about 40ms; collocation of [h] + plosives is 60-80ms; as initial consonant, [h]'s duration is 70-100ms; col-location of [h] + nasals is the longest, which is 100-130ms.(3) Viewed from spectrum energy, [h]> [h] + nasals > [h] + plosives > [h] + fricatives. With the above results, it is can be seen that, the differences are greater among the acoustic characteristics of [h]as initial consonant and prepositional consonant, and that present the different ways of pronunciation.

Acknowledgements

The success of this project was subsidized by the Major Programs of Social Science Foundation of China (10&ZD125) and the Key Programs of State Ethnic Affairs Commission State Ethnic Affairs Commission 2012 (The Multi-mode and Digital Research of Tibetan Phonetic).

References

[1] Aitang Qu: Fifty years of Tibetan language study in China. China Tibetology. Vol. 65 (2004) No. 1, p. 84-96. (In Chinese)

[2] Kan Hua: the developing of the consonant clusters and the consonant endings of Anduo Tibetan. Journal of Northwest Minorities University. (2013) No.1, p. 26-34. (In Chinese)

[3] Kan Hua: the consonant clusters comparison between Xiahe Tibetan and Maxu Tibetan in Gannan. Journal of Northwest Minorities University. (2013) No.4, p. 35-42. (In Chinese)

[4] Jiangping Kong: An acoustics study of the initial fricative consonants of Daofu Tibetan. Minority Languages. Vol. 3(1991), p. 59-64. (In Chinese)

[5] Jiangping Kong: An acoustics study of the initial plosive consonants of Daofu Tibetan. Minority Languages. Vol. 2(1991), p. 122-133. (In Chinese)

Advanced Materials Research Vols. 926-930 (2014) pp 1846-1849
© (2014) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMR.926-930.1846

The acoustic analysis of the male's F0 in Mongolia Folk Long Song's Vibrato

Ting Sun [a], Hongzhi Yu [b, *], Yasheng Jin [c]

Key Lab of China's National Linguistic Information Technology, Northwest University for Nationalities, LanZhou, China

[a]sukkyang_fly@sina.com, [b]yuhongzhi@hotmail.com , [c]xbgsj@xbmu.edu.cn

Keywords: F0, acoustic analysis, vibrato, Mongolia folk long song.

Abstract. This paper focuses on the F0 in Mongolia folk long song's vibrato, with the basic voice parameters, it is found that the three styles of the vibrato: symmetrical style, right-side style, left-side style. On the one hand, this paper explains the acoustic characters of the Mongolia folk long song's vibrato, and on the other hand, the results of this paper can be the references for establishing the standards of the oral cultures' digital protections.

Introduction

The Mongolia folk long song has more than 2000 year's history, it is not the representation of the Mongolia culture, but it is also regarded as a world class art. In 2005, Mongolia folk long song has been selected into the world non-material cultural heritage list. The melody of long song is slow and leisure, and contains many different typical singings, among of them, "Nugula" is the most representative singing [1, 2]. "Nugula" is the most important expression skill in the long song's singing, and which can produce a kind artificial vibrato.in the long song's singing, vibrato is the most widespread expression skill, which best represent the Mongolia music, most of the characters in long song, cannot be lack of "Nugula" [3-5].

Now, the existing research about the vibrato in Mongolia folk long song are not deep enough, most of them just have divided the vibrato's different singing styles, which are based on the physiological angle.Though all of the existing research results are significant to the teaching practice and the art theories study, but if the study methods of modern speech acoustics could be used in the study of vibrato, for one hand, it is could have come true that, take an acoustic study on the Mongolia Folk Long Song's with the quantitative analysis of computer; for other hand, researching results could have been the foundation for establishing the standards of the oral cultures' digital protections. This paper using the methods of experimental phonetics, expecting do some research about the F0 in Mongolia folk long song's vibrato, and the results could support some foundational data for the teaching practice and protection in Mongolia folk long song.

The Research Procedures

Using the study methods of experimental phonetics, the research procedures contains collecting signals, defining parameters, and analyzing parameters.

Experimental Equipment. The experimental equipment contains neck-style microphone, mixing console, external sound card and a laptop. All the samples signals have been collected in a professional recording studio, and which contains speech and voice. The samples' frequency is 40 kHz, and the file format is.wav.

Speech Sample. The speaker is a male professional singer for Mongolia folk long song in Alashang Zuoqi. All the study samples come from 10 collected songs which are different in styles but all are very popular. From the fig.1, we can see the vibrato's speech signals and energy signals are both waveforms.

Defining Parameters. Pitch is the subjective sensation of the human auditory. Generally, pitch depends on the frequency's range and the loudness of sound' range. Low frequency can bring a kind of deep, thick and wild feeling; and the high frequency can bring a kind of shining, bright and shrieking feeling.

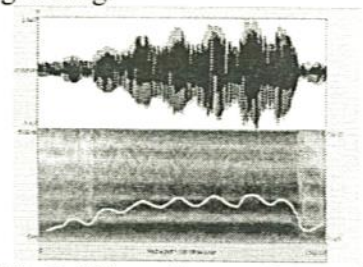


Fig.1 Vibrato's speech waveform and spectrum

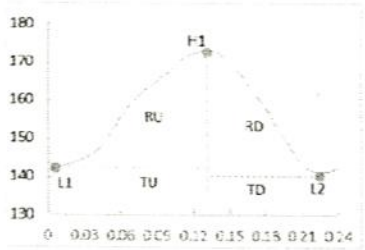


Fig.2 Defining parameters

The fundamental frequency is the number of vibrations in per unit time, the unit is Hz. In general, when the vocal cord is tensed, the sound is high, and it is loosened, the sound is low, that is the human's feeling about the changes of melody. In general, female's vocal cord is thinner than male's, so their fundamental frequency is higher than male's.

Vibrato is also called trill. When the breathing current passes the vocal cord, and the singer make it vibrating conscious, Vibrato will be produced. The vibrato characters are the numbers of frequency of vocal cord in physiology, but are the changes of fundamental frequency in acoustics.

In this paper, after collecting the fundamental frequency data from the vibrato samples, the parameters have to been defined. All the parameters are as the following: the vibration period (VP), the up period (TU) , the down period (TD); the sound range(R), the up sound range (RU) , the down sound period (RD), the average sound range (RA); The up numbers of vibration (VU), the down numbers of vibration (VD), and the average numbers of vibration (VA). With fig. 2, it is can be shown:

$$VC\ (ms) = L2\_time - L1\_time = TU + TD \quad (1)$$

$$TU\ (ms) = H1\_time - L1\_time \quad (2)$$

$$TD\ (ms) = L2\_time - H1\_time \quad (3)$$

$$RU\ (Hz) = H1\_f0 - L1\_f0 \quad (4)$$

$$RD\ (Hz) = H1\_f0 - L2\_f0 \quad (5)$$

$$RA\ (Hz) = (R1 + R2)/2 \quad (6)$$

$$VU\ (times) = (L1\_f0 * TU + H1\_f0 * TU)/2 \quad (7)$$

$$VD\ (times) = (H1\_f0 * TD + L2\_f0 * TD)/2 \quad (8)$$

The Analysis of Vibrato's F0

F0's vibration period. There are 50 vibrato samples have been collected from 10 songs. All of their vibration period as the table.1.With the table.1, it is can be seen that the range of VP is120-220ms, and when the VC is 120-150ms, the vibrato waveform is left-side style; when the VC is longer than before, the vibrato waveform is symmetrical style.

TU is the time when the vibrato's fundamental frequency from the lowest point to the highest point in one cycle. TD is the time when the vibrato's fundamental frequency from the highest point to the lowest point in one cycle.

From the fig.3, it is can be seen that in one cycle TU>TD and TU<TD both exist in the vibrato's fundamental frequency, with it, it is can be assumed that the diversity of the changing for vibrato's fundamental frequency has the close relationship with the Mongolia folk long song's artistic features.

Table1 VC's range

No.	Cycle (ms)	No.	Cycle (ms)	No.	Cycle (ms)	No.	Cycle (ms)
1	200	14	180	27	180	40	150
2	200	15	220	28	190	41	180
3	200	16	180	29	200	42	140

4	170	17	190	30	190	43	120
5	200	18	190	31	150	44	190
6	220	19	170	32	140	45	180
7	190	20	180	33	200	46	160
8	220	21	180	34	190	47	150
9	190	22	170	35	190	48	210
10	200	23	170	36	180	49	190
11	190	24	180	37	160	50	170
12	170	25	200	38	170	51	210
13	190	26	180	39	170		

F0's ranges. As the physiological difference between male and female's in vocal organ, the F0's ranges are different be male and female. In general, for the male, it is 50-250 Hz; for the female, it is 120-500 Hz; [4]. In this paper, the male's upper limit is 385Hz, and the lower limit is 130Hz.

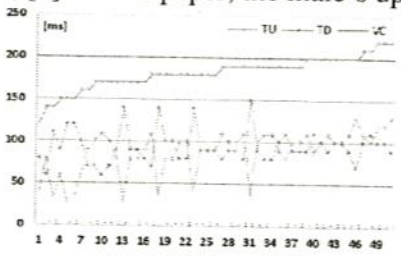


Fig.3 TU and TD

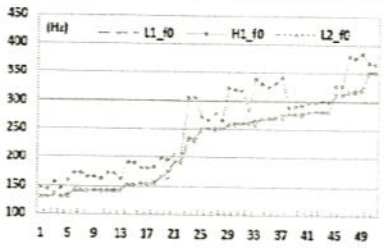


Fig.4 R's range

Fig.5 RU and RD

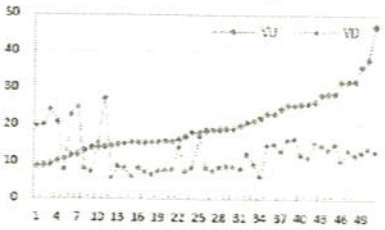


Fig.6 VU and V

R is the difference between the highest F0 and the lowest F0. From fig.4, it is can be seen that with the comparable lower R, the vibrato can exist in both high range frequency (100-200Hz) and the low range frequency (250-350Hz). According the further observation, it is can be seen that when the vibrato's F0 is under200Hz, R is 15-40Hz, and it is the symmetrical style; when the vibrato's F0 is in a high range, R is larger than before, the largest is 70Hz, and it is the left-side style.

RU is the difference between the highest F0 and the lowest F0 in TU. RD is the difference between the highest F0 and the lowest F0 in TD. From fig.5, it is can be seen that there are not too much differences between RU and RD, and in general, difference is under 3Hz.

F0's vibration rate. The vibration rate is the numbers of vibration in one cycle. In this paper, we use VU, VD and other parameters in one vibrato's cycle.

With the fig.6, it is can be seen that the largest number of vibration is 47, and the smallest number of vibration is 5; the VU's range is 9-38 while the VD's range is 9-27, in most situation it is VU>VD, but when situation like VD>VU, it is the left-side style.

The Styles of the Vibrato

With the study, it is can be defined that, for the male singer, there are three vibrato styles of the Mongolia folk long songs: symmetrical style (fig.7), right-side style (fig.8), left-side style (fig.9).

(1) It is can be seen that the vibrato of symmetrical style exist both in high range frequency and low range frequency, and the numbers of vibration is 30-20.

(2) It is can be seen that the vibrato of right-side style exist both in high range frequency and low range frequency, and the largest number of vibration appears in high range frequency, and the numbers of vibration in the low range frequency is less than 20.

(3) It is can be seen that the range of the vibration frequency of left-side style is often higher than 260Hz, and it is very often as TU>TD. The numbers of vibration is 30-10, and VU<VD.

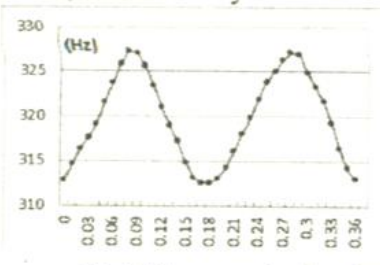


Fig.7 Symmetrical style

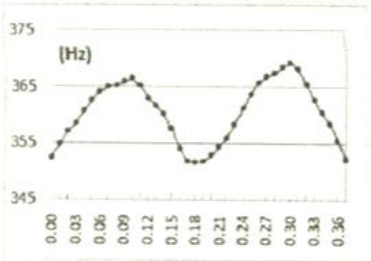


Fig.8 Right-side style

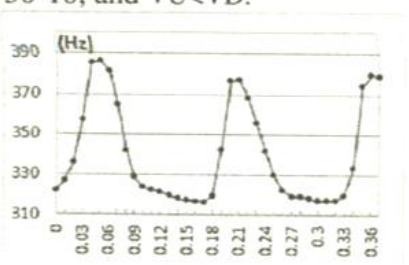


Fig.9 Left-side style

Conments

Vibrato is the most stylish and the most typical sound production phenomenon in Mongolia folk long songs. In this paper, there are 40 vibrato samples which have been collected from 10 popular Mongolia folk long songs. Base on the acoustic study, we get three styles of the vibrato: symmetrical style, right-side style, left-side style. On the one hand, this paper explains the acoustic characters of the Mongolia folk long song's vibrato, and on the other hand, the results of this paper can be the references for establishing the standards of the oral cultures' digital protections.

Acknowledgements

The success of this project was subsidized by the Major Programs of Social Science Foundation of China (10&ZD125).

References

[1] Baodaerhan, Wuyuntaoli: Mongolia Folk Long Song (Zhejiang People's Publishing House, China 2007), (In Chinese)

[2] Guoxin Shi, Xu Cao: Basic Features and Applications of Trills in Singing. Journal of Nantong University. Vol. 29 (2013) No. 4, p. 77-81. (In Chinese)

[3] Rong Rao: The Art Features and the Singing methods of Mongolia Folk Long Song. Journal of Music time and Space. (2007), p. 62-63 (In Chinese)

[4] Jie Zhang, Ziye Long: A Summarize of Pitch Detection Algorithmicin Speech Signals Processing. Journal of University of Electronic Science and Technology of China. Vol. 39(2010), p. 99-102. (In Chinese)

[5] Huaping Fang, Yonghong Li: Prosody Acoustic Feature Study of Mongolia Folk Song-Holy Lord Two Horses. Journal of Northwest University for Nationalities (Natural Science). Vol. 33 (2012) No. 2, p. 66-70 (In Chinese)