INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

The Frequency Range of "The Ling Six Sounds" in Standard Chinese

Aijun Li[1], Hua Zhang[2,3], Wen Sun[2,3]

[1]Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
[2]Clinical Audiology Center, Beijing Tongren Hospital, Beijing, China
[3]Beijing Institute of Otolaryngology, Beijing, China
liaj@cass.org.cn, a-zhang@263.net

Abstract

"The Ling Six Sounds" are a range of speech sounds encompassing the speech frequencies that are widely used clinically to verify the effectiveness of hearing aid fitting in children. This study focused on the spectral features of the six sounds in Standard Chinese. We examined the frequency range of /m, u, a, i, ʂ, s/ as well as three consonants in syllables, i.e., /m(o)/, /ʂ(ʅ)/, and /s(ɿ)/. We presented the frequency distribution of these sounds. Based on this, we further proposed guidelines to improve "the Ling Six-Sound Test" regarding tones in Standard Chinese. We also suggested further studies in other dialects/languages spoken in China with regard to their phonological specifics.

Index Terms: The Ling Six-Sound Test, frequency, Standard Chinese

1. Introduction

According to China National Sample Survey on Disability [1], there are 27.8 million speech-disabled people in China, among which 137,000 are hearing-impaired children (0-6 years old); and each year, there are 23,000 deaf new-borns. This population mainly speaks Chinese (including 10 major dialect families), with a small group speaks minority languages, e.g., Uygur, Mongolian, and Tibetan. To assess the effectiveness of hearing aids fitting in such a large population, reliable hearing tests and standards are thus urgently called for.

"The Ling Six-Sound Test" (hereafter as LST) is a hearing test frequently used in clinical assessment of a child's ability to hear across the frequencies of spoken languages typically with hearing aids [2]. The test sounds include three cardinal vowels /u, a, i/, two sibilants /ʃ, s/, as well as the labial nasal /m/, which are supposed to represent the broad speech spectrum from 250-8000Hz [2]. Table 1 shows the frequencies of the six sounds (from low to high), among which vowels are shown with F1 and F2 values, while consonants with the range of center of gravity values.

Table 1: *Frequencies of the Ling Six Sounds [1].*

Phoneme	F1 (Hz)	F2 (Hz)	Center of gravity (Hz)
/m/			250-500
/u/	350	900	
/a/	700	1300	
/i/	300	2500	
/ʃ/			2000-4000
/s/			3500-7000

As can be seen from Table 1, the center of gravity values of /m/ correspond to the low-frequency band; /m/ is thus usually used to assess the hearing of the low-frequency sounds. On the opposite, /ʃ/ and /s/ mainly test the hearing of high-frequency sounds. The three vowels are mainly used to test the perception of the vowel space.

LST was originally developed for the North American population. Due to the differences in production and spectral content of consonants and vowels in other languages, the "six sounds" are sometimes adapted according to the local phonology, e.g., in Australian English, [3] proposed /ɔ/ instead of /u/ as the high back test vowel, as Australian English does not have the cardinal /u/, but a high mid vowel /ʉ/ instead.

The current study is focusing on the Standard Chinese (hereafter as SC) version of "the Ling Six Sounds". As there is no laminal sibilant /ʃ/ in SC but the postalveolar /ʂ/ instead [4], the six sounds are therefore /m, u, a, i, ʂ, s/. The motivation of this study is three-fold. First, the postalveolar /ʂ/ in SC is different from /ʃ/. Second, the other five sounds in SC also show different phonetic realization and frequency range from those in American English (hereafter as AE) (e.g., [5]). Third, SC is a tone language, which employs f0 changes to signal lexical meanings. The range of f0 changes are typically lower than the low frequency range of the six phonemes. The original LST thus cannot reliably test the frequencies of lexical tones. To improve the effectiveness of the hearing-aid tests for SC speaking children, it is therefore in great need to define the frequency range of "the Ling Six Sounds" for SC.

So far, most studies on the frequency range of these six sounds are based on data from a limited number of speakers (e.g., [5-10]). [11] used a relatively larger speaker pool, but obtained normalized "V-values" which can hardly be of practical application. [12] used data from 20 speakers of Beijing Mandarin. Although similar with SC, Beijing Mandarin is different in various ways and thus less representative of SC [13]. This study thus aims to fill the gap.

2. Methods

2.1. Participants

60 normal-hearing participants (30 male; 30 female) who speak SC were recruited for the experiment (Mean=22.5, SD=1.8). Their proficiency of SC and hearing ability were examined by professional audiologists (authors of this paper). All were qualified in these two aspects.

2.2. Stimuli

The test sounds included the Ling Six Sounds in SC, i.e., /m, u, a, i, ʂ, s/, as well as three syllables /mo/, /ʂʅ/, and /sɿ/. The three syllables are the names to stand for the three consonants /m/, /ʂ/, and /s/ in oral communication, respectively. They are

http://dx.doi.org/10.21437/Interspeech.2017-329

thus frequently used in real tests instead of the single consonants.

2.3. Recordings

All recordings were made in a sound-proofing recording booth at the Institute of Linguistics in the Chinese Academy of Social Sciences. Participants were instructed to close their mouths when producing /m/ and not to vibrate their vocal cords when producing /ʂ/ and /s/. They were also asked to produce all stimuli in normal speed and consistent volume. Voiced stimuli were produced with a high-level tone. Each stimulus was produced five times in random order. Except for the missing of syllables /mo, ʂʅ, sɿ/ from one male speaker, there were (6 sounds + 3 syllables) * 5 times * 60 participants – 3 syllables * 5 times = 2685 sound files.

2.4. Data Extraction

All sounds were manually segmented in Praat [14]. We extracted values of F1, F2 and F3 of /a, i, u/ by averaging 10 equal-interval points along the formant curves. The spectral properties are well reflected in the first moment calculated by spectral normalization. The extracted information for /m, ʂ, s/ included the center of gravity (with standard deviation), either from isolated consonants or from consonants in syllables /mo, ʂʅ, sɿ/, respectively.

3. Results

3.1. Consonants: Center of gravity

Figure 1 illustrates the spectrum of /m/ vs. /m(o)/, /ʂ/ vs. /ʂ(ʅ)/, and /s/ vs. /s(ɿ)/, based on data produced by a male speaker and a female speaker. The x-axis stands for the frequency, the y-axis for the sound pressure level.

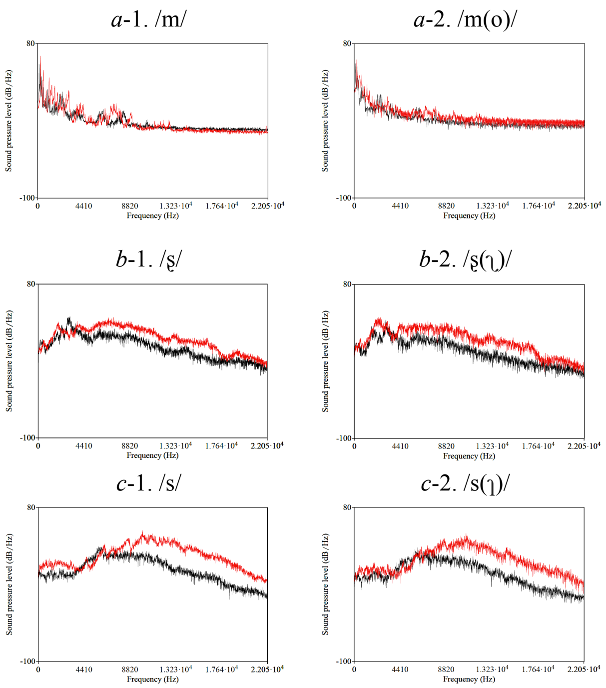

Figure 1: *Representative spectra of /m/ (a), /ʂ/ (b), and /s/ (c). Left column for isolated consonants, right column for consonants in syllables. Light lines indicate females, dark lines for males.*

In Figure 1, the frequency peak (center of gravity) of three consonants generally increases from /m-m(o)/ to /ʂ-ʂ(ʅ)/, and further higher in /s-s(ɿ)/. In addition, the center of gravity values seem to be conditioned by context and gender. The following sections zoom into each consonant pair.

3.1.1. /m/ vs. /m(o)/

Figure 2 compares the distribution of center of gravity of /m/ vs. /m(o)/ across genders. Likewise, Figures 3, 4 compare the distribution of center of gravity of /ʂ/-/ʂ(ʅ)/ and /s/-/s(ɿ)/.

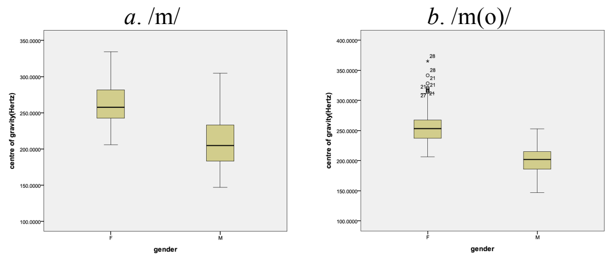

Figure 2: *The center of gravity distribution of /m/ (a) vs. /m(o)/ (b). Left box is for female in each graph, right box for male.*

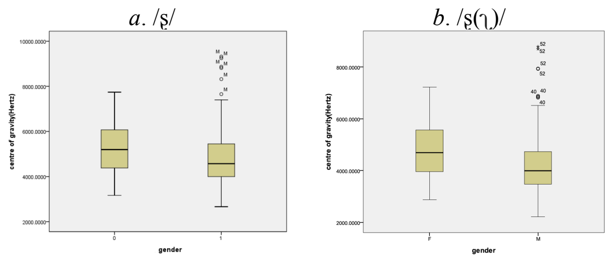

Figure 3: *The center of gravity distribution of /ʂ/ (a) vs. /ʂ(ʅ)/ (b). Left box is for female in each graph, right box for male.*

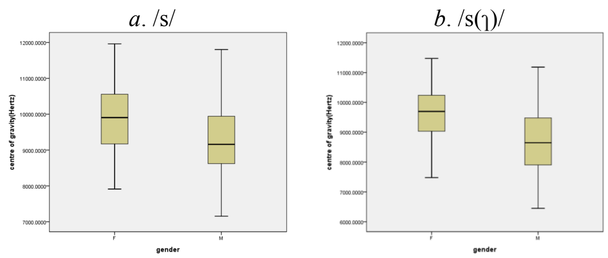

Figure 4: *The center of gravity distribution of /s/ (a) vs. /s(ɿ)/ (b). Left box for female(F) in each graph, right for male (R).*

First, by comparing the two graphs in each figure, we can see that, due to the influence of the following vowels, the consonants in syllables generally show slightly lower center of gravity than those produced in isolation. The ANOVA tests showed that the mean center of gravity of consonants in syllables /m(o)/ (229Hz), /ʂ(ʅ)/ (4493Hz), and /s(ɿ)/ (9200Hz) were significantly lower than that of /m/ (236Hz), /ʂ/ (5000Hz), and /s/ (9577Hz), respectively (/m(o)/-/m/: $F(1,594) = 4.452$, $p<0.05$; /ʂ(ʅ)/-/ʂ/: $F(1,593) = 29.48$, $p<0.001$; /s(ɿ)/-/s/: $F(1,593)=19.734$, $p<0.001$). The differences can be observed in both genders, as shown in Table 2, the only exception was the /m(o)/-/m/ pairs of female speakers, which, however, was still marginally significant ($p=0.052$).

Second, as can be seen from the comparison between female vs. male data, the mean center of gravity of female data is significantly higher than that of the male data in all sounds. Table 3 lists the mean center of gravity values and ANOVA tests of all consonants, which confirms our observations from the figures.

Table 2: *ANOVA results of the mean center of gravity differences between consonants in isolation vs. in syllables by gender. F stands for female, M for male.*

/m(o)/-/m/	F	F(1,298)=6.29, p<0.05
	M	F(1,293)=3.818, p=0.052
ʂ(ʅ)/-/ʂ/	F	F(1,298)=12.48, p<0.001
	M	F(1,293)=18.036, p<0.001
/s(ɿ)/-/s/	F	F(1,298)=46.618, p<0.001
	M	F(1,293)=19.612, p<0.001

Table 3: *Mean center of gravity of each consonant produced by difference genders.*

	Female	Male	ANOVA Results
/m/	263Hz	208Hz	F(1,298)=229, p<0.001
/m(o)/	255Hz	201Hz	F(1,293)=351, p<0.001
/ʂ/	5196Hz	4821Hz	F(1,298)=7.919, p<0.01
/ʂ(ʅ)/	4764Hz	4212Hz	F(1,293)=17.99, p<0.001
/s/	9891Hz	9263Hz	F(1,298)=33.597, p<0.001
/s(ɿ)/	9667Hz	8720Hz	F(1,293)=17.99, p<0.001

3.2. Vowels: Formants

Tables 4-6 list the mean F1-F3 of vowels /a/, /i/, and /u/. F stands for female, M for male, All for the grand mean.

In Tables 4-6, the female data show clearly higher F1, F2 and F3 values than that of males in all vowels. Except in the F2 of /u/, the differences between two genders are significant in all the other cases, as shown in Table 7.

Table 3: *Mean values of each formant of vowel /a/(Hz).*

		N	Mean	95% Conf. Int. Lower	Upper	Min	Max
F1	F	150	1028	1016	1042	826	1124
	M	150	765	754	776	589	952
	All	300	897	880	914	589	1224
F2	F	150	1552	1535	1568	1274	1763
	M	150	1224	1211	1237	1033	1369
	All	300	1388	1366	1409	1033	1763
F3	F	150	2920	2897	2944	2495	3313
	M	150	2588	2563	2614	2158	2924
	All	300	2754	2729	2780	2158	3313

Table 5: *Mean values of each formant of vowel /i/(Hz).*

		N	Mean	95% Conf. Int. Lower	Upper	Min	Max
F1	F	150	322	315	330	240	459
	M	150	278	274	282	216	361
	All	300	300	295	305	216	459
F2	F	150	2724	2701	2746	2267	3026
	M	150	2182	2160	2204	1799	2412
	All	300	2453	2418	2487	1799	3026
F3	F	150	3652	3627	3677	3185	3996
	M	150	3130	3105	3155	2705	3444
	All	300	3391	3356	3425	2705	3996

Table 6: *Mean values of each formant of vowel /u/(Hz).*

		N	Mean	95% Conf. Int. Lower	Upper	Min	Max
F1	F	150	379	370	387	285	518
	M	150	341	338	345	298	390
	All	300	360	355	365	285	518
F2	F	150	749	731	766	506	1160
	M	150	730	719	741	576	924
	All	300	739	729	750	506	1160
F3	F	150	2826	2788	2865	1557	3369
	M	150	2384	2361	2408	1767	2817
	All	300	2605	2572	2639	1557	3369

Table 7: *ANOVA results of* the *mean F1-F3 values across genders.*

	Formants	Females vs. Males
/a/	F1	F(1,298)=885.589, p<0.001
	F2	F(1,298)=960.373, p<0.001
	F3	F(1,298)=355.462, p<0.001
/i/	F1	F(1,298)=108.577, p<0.001
	F2	F(1,298)=1167.884, p<0.001
	F3	F(1,298)=851.11, p<0.001
/u/	F1	F(1,298)=62.6, p<0.001
	F2	F(1,298)=3.21, p=0.074
	F3	F(1,293)=17.99, p<0.001

3.3. Summary

To establish the frequency standard of “the Ling Six Sounds” in SC, Table 8 summarizes the frequencies of the test phonemes in this study, including /m, u, a, i, ʂ, s/ produced in isolation, as well as /m, ʂ, s/ produced in syllables.

Table 8: *A summary of frequency of test sounds (Hz).*

	F1 Mean(SD); Min-Max	F2 Mean(SD); Min-Max	Center of Gravity Mean(SD); Min-Max
/m/			236(41); 147-335
/m(o)/			229(36); 147-365
/u/	360(45); 285-518	739(91); 506-1160	
/a/	897(153); 589-1224	1388(188); 1033-1763	
/i/	300(43); 216-459	2453(304); 1799-3026	
/ʂ/			5009(1169); 2665-9318
/ʂ(ʅ)/			4493(1148); 2217-8789
/s/			9577(990); 7155-11961
/s(ɿ)/			9201(1074); 6449-11472

4. Discussions

4.1. American English vs. Standard Chinese

By comparing Tables 1 vs. 8, we can see that the six sounds in AE and SC version show greatly different frequency distributions. Among them, the three vowels show less differences between the two versions, as can be seen from Figure 5 which plots the three vowel clusters in the two versions all together (in bark). The data of AE was obtained from [15-16]. We can see from Figure 5 that, despite subtle deviations, e.g., F2 of /u/ and /F1/ of /a/, the three vowels in SC are basically in similar ranges with those in AE for both genders. This is comparable with the production data in [17].

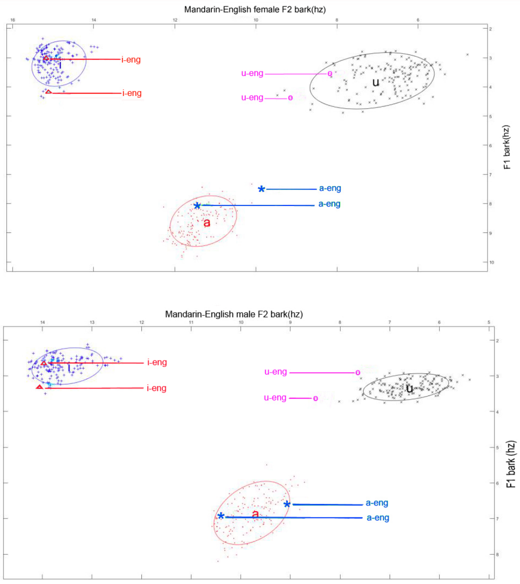

Figure 5: *Clusters of /a, i, u/ in AE and SC, upper panel for female and lower panel for male*

Comparatively, the consonants show larger differences between the two language versions. To be specific, /m/ in SC is slightly lower and shows narrower range than that in AE in terms of the center of gravity. The two sibilants /ʂ, s/, however, have much higher center of gravity values than /ʃ, s/ in AE, about 2000Hz higher for /ʂ/ (4000-6000Hz) and 5000Hz higher for /s/ (8000-10000Hz). Such a shift might be resulted from the phonological differences between the two languages. In SC, in addition to /ʂ, s/, there is a palatal fricative /ɕ/ whose frequency ranges in between /ʂ/ and /s/, relatively corresponding to /s/ in AE. The frequency range of /s/ in SC is thus generally raised to strengthen the perceptual differences among the three sibilants. To test the high-frequency range covered by /s/ in AE, we thus suggest adding /ɕ/ into the test sounds in SC.

In general, the frequency range that the SC version can test is about 200-12000Hz, which is broader than the AE version, i.e., 250-7000Hz. Consonants in syllables, i.e., /m(o)/, /ʂ(ʅ)/, /s(ɿ)/, realized with slightly lower center of gravity than those produced in isolation, still show broader frequency range than that in AE. /m(o)/, /ʂ(ʅ)/, /s(ɿ)/, therefore, can also be used as test sounds.

4.2. Lexical Tones and Suggestions on LST in SC

In a tonal language as SC, tones are employed and thus have heavy function loads in distinguishing lexical meanings [18]. However, lexical tones are not considered in the LST currently, among which the tonal range is decided by two tones, i.e., the Low-Dipping tone (Tone 3, as T3), and the High-Level tone (Tone 1, hereafter as T1).

Previous studies (e.g., [9, 19]) show that the mean frequency range of four lexical tones in SC is 125-270Hz for female speakers, and 87-167Hz for male speakers. Among all tones, T3 usually lies at the bottom of the pitch range and shows the narrowest pitch range (87-122Hz for males and 125-203Hz for females); T1, however, usually lies at the upper end of a speaker's pitch range (about 270Hz for females and 167Hz for males). As the current SC version of LST has only considered frequencies above 200Hz, it only covers the higher part of the female lexical tones (e.g., T1). The lexical tones of males, however, are completely not covered.

We thus suggest to add the lexical tone tests in the SC version. T3-T1 minimal pairs could be used to test the contrast of tone height (T3 for Low; T1 for High). In the tone tests, we could ask younger children to point to pictures they have heard, or ask older children to repeat what they have just heard. If extra segments are not considered, we could use /i/ and /u/ in the T3-T1 tone tests, as these two vowels make meaningful words with T3 or T1 that could be visualized with pictures, for example, /i^3/ *'chair'*- /i^1/ *'cloth'*, /u^3/ *'dance'*- /u^1/ *'house'*.

If more segments are considered in the tests, we could also consider other T3-T1 minimal pairs, such as /ma^3/ *'horse'* - /ma^1/ *'mother'*, /ʂu^3/ *'rat'* - /ʂu^1/ *'book'*, /ian^3/ *'eye'* - /ian^1/ *'smog'*, /piŋ^3/ *'pie'* - /piŋ^1/ *'ice'*, /tɕiŋ^3/ 'well' - /tɕiŋ^1/ *'whale'*, /ua^3/ *'tile'* - /ua^1/ *'frog'*. This might thus result in "the Ling Eight-Sound Test" instead of "Six-Sound", by either adding T3-T1 pairs in /i/ or /u/, or in other syllables. Given the space limitation of this paper, we will report the detailed experimental results in the future.

4.3. LST in Chinese Dialects

Despite the fact that SC has been the *lingua franca* in China, there are 10 major dialectal areas in China, each of which exhibits great variation in terms of the acoustic realization of the six sounds and lexical tones. Moreover, several other languages are spoken in China in addition to Chinese, such as Uygur, Mongolian, and Tibetan. These dialects/languages are phonologically different from SC to various extent.

For example, as illuminated from our 41-Chinese-dialect monosyllabic corpus, a quarter of the dialects have only one sibilant /s/ (e.g., Hokkien and Cantonese); others have /ɕ/ instead of /ʂ/ (e.g., Southwestern Mandarin family). Vowels sometimes show extra contrasts in different dialects, such as the nasal vs. non-nasal contrast in the Hokkien family, as well as glottal vs. un-glottal contrast in dialects with check tones (e.g., Wu Chinese). These phenomena not only call for further studies on the frequency distribution of sounds in different dialects/languages, but also raise the question as to what extent LST should cover necessary hearing abilities of different phonetic cues in spoken languages. Future studies are thus in great need to develop different versions of LST based on the linguistic features of each dialect/languages.

5. Acknowledgments

This work is supported by the National Natural Science-Foundation of China (No. 61175016, 61304250), Key Fund Projects (No. 61233009), and the financial support from CASS Innovation Project "Articulatory Model for Pronunciation Training" to the first author, as well as the Capital Health Research and Development Projects (No. 2014-3-1091) to the second author.

6. References

[1] X. Sun, L. Yu, C. Qu, et al., "Zhongguo tingli canji goucheng tedian ji kangfu duice [An epidemiological study on the hearing-impaired population identified in China and proposed intervention strategies]," *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, vol. 27, no. 2, pp. 21-24, 2008.

[2] D. Ling, *Speech and the hearing-impaired child: Theory and practice.* Washington, DC: Alexander Graham Bell Association for the Deaf, 1976.

[3] K. B. Agung, S. C. Purdy, and C. Kitamura, "The Ling sound test revisited", *The Australian and New Zealand Journal of Audiology*, vol. 27, no. 1, pp. 33-34, 2005.

[4] W.-S. Lee and E. Zee, "Standard Chinese (Beijing)", *Journal of the International Phonetic Association*, vol. 33, no. 1, pp. 109-112, 2003.

[5] S. Lee, "Spectral analysis of Mandarin Chinese sibilant fricatives", *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 1178–1181, 2011.

[6] Z. Wu, "Putonghua yuanyin he fuyin de pinlü fenxi ji gongzhenfeng de cesuan [The spectral analysis of vowels and consonants and the measurement of formants in Standard Chinese]", *Acta Acustica*, vol. 1, 1964.

[7] Z. Wu (ed.), *Hanyu putonghua danyinjie yutuce [The Spectrogram of Monosyllables in Standard Chinese]*. Beijing: Chinese Social Science Press, 1986.

[8] H. Bao, "Putonghua danyuanyin fenlei de shengli jieshi [Physiological accounts of monophthongs in Standard Chinese]", *Studies of the Chinese Language*, vol. 2, 1984.

[9] Z. Wu and M. Lin, *Shiyan yuyinxue gaiyao [An Introduction to Experimental Phonetics]*. Beijing: Higher Education Press, 1989.

[10] J. Yang, *"Putonghua seyin yu cayinpu de shengxue texing yanjiu [An acoustic study on the spectral properties of stops and fricatives in Standard Chinese]"*. Thesis. Beijing: Chinese Academy of Social Sciences, Institute of Linguistics, 2006.

[11] P. Wang and F. Shi, "Hanyu putonghua jichu yuanyin de texing fenxi [A statistical analysis of basic vowels in Standard Chinese]", *Nankai Linguistics*, vol. 2, 2014.

[12] E. Zee, "The phonetic value of the vowels, diphthongs, and triphthongs in Beijing Mandarin", *Modern Phonetics in the New Century – Proceedings of the 5th Conference of Modern Phonetics*, pp. 54-60, 2001.

[13] E. Chirkova and Y. Chen, "Beijing Mandarin, the language of Beijing", https://hal.archives-ouvertes.fr/hal-00724219, 2011.

[14] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer", http://www.fon.hum.uva.nl/praat/, 2017.

[15] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels", *Journal of the Acoustical Society of America,* vol. 24, pp.175-184,1952.

[16] J. Hillenbrand, L. A. Getty, M. J. Clark, et al., "Acoustic characteristics of American English vowels", *Journal of the Acoustical Society of America*, vol. 97, no. 5, 1995.

[17] B. Wen, "Meiguo xuesheng hanyu yuanyin xide de shiyan yanjiu [An experimental study on the acquisition of Chinese vowels by American learners]", *Chinese Language Learning*, vol. 3, pp.95-104, 2010.

[18] Y.-R. Chao, *Yuyan wenti [The Problem of the Chinese Language]*. Beijing: Shangwu Press, 1980.

[19] Y. Zhang and F. Shi, "Putonghua danziyin shengdiao de tongji fenxi [A statistical analysis on lexical tones in Standard Chinese]", *Chinese Journal of Phonetics*, vol. 6, pp. 38-45, 2016.