The relationship between the second subglottal resonance and vowel class, standing height, trunk length, and F0 variation for Mandarin speakers

Jinxi Guo[1], *Angli Liu*[1], *Harish Arsikere*[1], *Abeer Alwan*[1] *and Steven M. Lulich*[2]

[1] Dept. of Electrical Engineering, University of California, Los Angeles, CA 90095, USA
[2]Dept. of Speech and Hearing Sciences, Indiana University, Bloomington

lennyguo@g.ucla.edu, anglil91@ucla.edu, harishan@ucla.edu, alwan@ee.ucla.edu, slulich@indiana.edu

Abstract

The relationship between vowel formants and the second subglottal resonance (Sg2) has previously been explored in English, German, Hungarian and Korean. Results from these studies indicate that vowel space is categorically divided by Sg2 and that Sg2 correlates well with standing height. One of the goals of this work is to verify if the above findings hold true in Mandarin as well. The correlation between Sg2 and sitting height (trunk length) is also studied. Further, since Mandarin is a tonal language (with more pitch variations compared to English), we study the relationship between Sg2 and fundamental frequency (F0). A new corpus of simultaneous recordings of speech and subglottal acoustics was collected from 20 native Mandarin speakers. Results on this corpus indicate that Sg2 divides vowel space in Mandarin as well, and that it is more correlated with sitting height than standing height. Paired t-tests are conducted on the Sg2 measurements from different vowel parts, which represent different F0 regions. Preliminary results show that there is no statistically-significant variation of Sg2 with F0 within a tone.

Index Terms: second subglottal resonance, Mandarin, vowel space, sitting height, tonal language

1. Introduction

The subglottal acoustic system refers to the acoustic system below the glottis, which consists of the trachea, bronchi and lungs [1]. The subglottal resonances (SGRs) are the natural frequencies of the subglottal system, and correspond to the complex conjugate pairs of poles in the subglottal input impedance (measured from the top of the trachea looking down) [2]. Compared to the supraglottal system, the configuration of the subglottal system is relatively fixed and thus SGRs are expected to remain constant for a given speaker. Previous studies [3, 4, 5] have shown that the acoustic contrasts for some phonological distinctive features are dependent on the SGRs. For instance, the second subglottal resonance (Sg2) forms a natural division between front and back vowels [6]. The aforementioned relationship between vowel categories and SGRs has previously been shown in American English [7], German [8], Korean [9] and Hungarian [10], but not in Mandarin.

In previous studies, a uniform tube model of the subglottal system is proposed, which shows an inverse relationship between speaker standing height and SGRs [11]. Such inverse relationship has been demonstrated and applied to automatic height estimation for English speakers [12]. Moreover, it is physiologically hypothesized that, since the effective length of the subglottal system is expected to be correlated with the size of the lungs and the length of the trunk, SGRs are likely to be strongly correlated with trunk length [13]. However, this correlation has not been studied before.

Work supported in part by NSF Grant No. 0905381.

Unlike the languages mentioned above, Mandarin is a tonal language. Its F0 variation is large compared to English. There might also exist correlation and interaction between F0 and SGRs. However, no work on this subject has been done before.

In this paper, we present evidence illustrating the relationship between Sg2 and vowel class, standing height, trunk length, and F0 in Mandarin. The results indicate that: 1) Sg2 forms a boundary between front and back vowels; 2) there exist inverse relationships between standing height and Sg2, and between trunk length and Sg2, with a stronger correlation between trunk length and Sg2; 3) while F0 varies over time within a vowel, there is no statistically-significant variation of Sg2.

2. Database

2.1. The vowel and tonal system of Mandarin and the new corpus

The Mandarin vowel system contains 6 standard vowels, which are [a], [o], [e], [i:], [u], [ü]. A short, retroflex [i] is also studied and distinguished from the long [i:]. Therefore there are 7 vowels in total in our corpus, as illustrated in Table 1. The place-of-articulation feature [+/− back], specifying the tongue position, is also shown in Table 1. [+] indicates that the tongue dorsum bunches and retracts slightly to the back of the mouth, while [−] indicates that the tongue extends slightly forward.

Table 1. *Phonological classification of Mandarin vowels*.

Vowel	i:	i	ü	e	a	u	o
[+/-Back]	-	-	-	central	central	+	+

As a tonal language, there are four tones in Mandarin, which are flat, rising, falling-rising and falling, as illustrated in Table 2. Table 2 also shows the symbols of the four tones, which indicates the corresponding F0 trajectory. The vowels in our database are in a ‘pV’, ‘shV’ or ‘xV’ context. For each context and each tone, we used a corresponding Pinyin (the official phonetic system for transcribing the Mandarin pronunciations of Chinese characters into the Latin alphabet) and a Chinese character, as illustrated in Table 3. Each character was embedded in the carrier phrase, “我把__的说一遍” (“I said __de once” (English version), “wǒ bǎ__de shuō yi biàn” (Pinyin version)).

Table 2. *Four tones in Mandarin.*

Tones	*First*	*Second*	*Third*	*Fourth*
Pitch	Flat	Rising	Falling-rising	Falling
Symbol	—	╱	╲╱	╲

Table 3. *The Mandarin corpus.*

Vowel	i:	i	ü	e	a	u	o
Pinyin	(xī) (xí) (xǐ) (xì)	(shī) (shí) (shǐ) (shì)	(xū) (xú) (xǔ) (xù)	(shē) (shé) (shě) (shè)	(shā) (shá) (shǎ) (shà)	(shū) (shú) (shǔ) (shù)	(pō) (pó) (pǒ) (pò)
Character	吸 席 喜 细	诗 石 史 事	虚 徐 栩 旭	奢 蛇 舍 设	沙 啥 傻 厦	书 熟 鼠 竖	坡 婆 叵 破

2.2. Recordings and height measurements

Acoustic data were collected with simultaneous speech and subglottal recordings for 20 native speakers of Mandarin (10 males and 10 females). Speech data were recorded using a Shure PG27 condenser microphone and subglottal data were obtained using a K&K Sound 'Hot Spot' accelerometer attached to the skin of the neck below the thyroid cartilage. All recordings were sampled at 48 kHz and digitized at 16bits/sample. All speakers, aged between 18 and 24 years, were recorded in one session, and every word (character) was recorded at least 4 times. Speaker height (standing height) and sitting height (trunk length) were measured before recording, with the sitting height being measured from the speakers' hip bone to top of the head. In this corpus, speaker height ranged from 165-195cm for males, and from 150-170cm for females. Speaker sitting height ranged from 73-81cm for males, and from 63-77cm for females.

3. Methods

3.1. Measurements

The first two formants of each vowel were measured from the microphone signals in their steady-state regions, and Sg2 was measured from the corresponding accelerometer signals, using

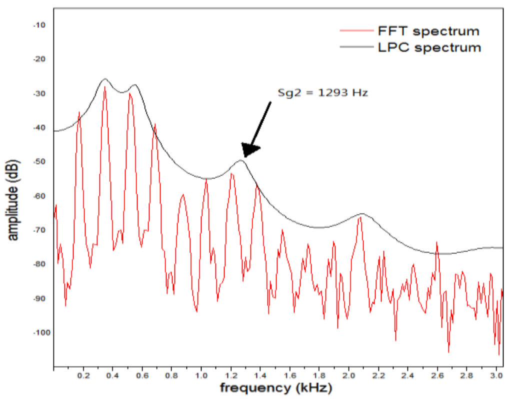

Figure 1: *FFT spectrum and LPC spectrum for a sample accelerometer signal taken from the Mandarin corpus.*

Wavesurfer [14]. Both signals were down-sampled to 8000 Hz before analysis. Before measuring Sg2 from accelerometer signals, each vowel segment was divided into three parts: beginning, middle and end. The reason is that the three different parts representing different F0 regions help us to study the correlation between Sg2 and F0. Sg2 measurements of each part were acquired by visual inspection of the resonance peaks in LPC spectra at the middle point of the corresponding part, as illustrated in Figure 1.

3.2. The relationship between Sg2 and vowel class

To investigate whether Sg2 divides the vowel space in Mandarin, means (Mean) and standard deviations (Std) of Sg2 measurements were calculated for each speaker. A frequency range from Mean-Std to Mean+Std was chosen to represent the Sg2 interval for each speaker as in [15], since we hypothesized that Sg2 measurements might exhibit increased variance during the large pitch excursions typical of Mandarin tones. For each vowel and each tone, the average values of the first two formants measured from the four respective repetitions were used. The Sg2 interval was then compared with the second formant (F2) for all seven vowels and four tones in order to test whether Sg2 defines the boundary of front and back vowels.

3.3. The relationship between Sg2 and standing height and sitting height (trunk length)

Mean Sg2 was used to investigate correlations between Sg2 and both standing and sitting height measurements. Mean Sg2 values along with standing and sitting height and gender is illustrated in Table 4. IDM is the abbreviation for speaker ID (male), IDF for speaker ID (female), STH (cm) for standing height, SIH (cm) for sitting height and SG2 (Hz) for Sg2. The data in Table 4 are in line with what we found before for English [2].

Table 4. *Mean Sg2 values along with standing and sitting height and gender for each speakers. IDM (speaker ID (male)), IDF (Speaker ID (female)), SG2 (Sg2 in Hz), STH (standing height in cm) and SIH (sitting height in cm)*

IDM	SG2	STH	SIH	IDF	SG2	STH	SIH
1	1292	177	77	2	1458	162	70
11	1240	171	77	3	1348	170	67
9	1193	195	81	18	1595	150	65
7	1342	165	75	12	1489	164	65
17	1355	170	73	16	1515	153	68
4	1332	178	78	6	1494	153	63
13	1354	174	74	8	1371	157	77
19	1268	175	76	10	1470	159	69
5	1268	175	76	14	1473	151	68
15	1336	171	76	20	1460	163	73

3.4. The relationship between Sg2 and F0

In the study of possible F0-Sg2 interaction, we chose the 3 corner vowels [u], [i:] and [a]. For every vowel, Sg2 measurements from 10 speakers (5 males and 5 females which are randomly chosen but cover the height range in the database), 4 tones and 4 utterances were selected. For each utterance, Sg2 measurement of 3 parts (beginning, middle and end of the vowel) were used for study.

Hypothesis testing [16] was applied to detect possible F0 influence on Sg2 using paired t-tests. The null hypothesis was that the mean difference between Sg2 in two parts of a vowel (beginning and middle, middle and end, beginning and end) is zero.

Since the paired t-test requires the difference between pairs to be normally distributed, it was necessary to verify this property of our data [17]. Figure 2 shows the 8 class histogram of Sg2 difference value between the middle and end of the vowel [u] with the first tone. The vertical line denotes the mean. Similar to the case in Figure 2, most of the data fit a normal distribution.

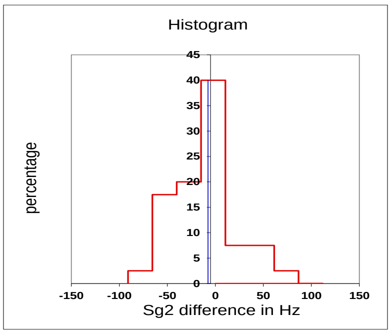

Figure 2: *Histogram of Sg2 difference for all 10 selected speakers between beginning and middle of vowel*[u]*, first tone.*

4. Results and Discussion

4.1. Sg2 and distinctive feature [+back] and [-back]

Two speakers (denoted as Female3 and Male11) were selected to show the division of vowel space by Sg2, as representative examples. Vowel plots for each speaker are shown in Figure 3. The plots are the sample for a given vowel regardless of the tones. The mean Sg2 is indicated by the solid horizontal lines and the upper and lower dashed lines indicate the Sg2 interval bounded by the values Mean+Std and Mean-Std.

Table 5 gives the percentage of the front and back vowels on the correct side of Sg2 (denoted as Pfront and Pback) for each speaker.

In general, the vowel space is divided by Sg2, with front vowels on one side of Sg2 and back vowels on the other side. This result is consistent with previous studies [7, 8, 9, 10].

Based on our observation of average Sg2 of 10 male and 10 female Mandarin speakers, we should note that the location of the central vowels which are close to Sg2 in vowel space such as [a] and [e], varies considerably for different speakers. For some speakers the central vowels show up above the upper boundary of the Sg2, while in some other cases they are located below the lower boundary of Sg2.

4.2. Correlation between Sg2 and speaker's standing height and sitting height

The scatter plot of height and sitting height versus Sg2 (mean Sg2 for given speaker) is shown in Figure 4 (all 20 speakers).

The correlation coefficients for Sg2 (mean Sg2 for a given speaker) with standing height and sitting height were also calculated for 10 male speakers, 10 female speakers and all 20 speakers respectively, as illustrated in Table 6.

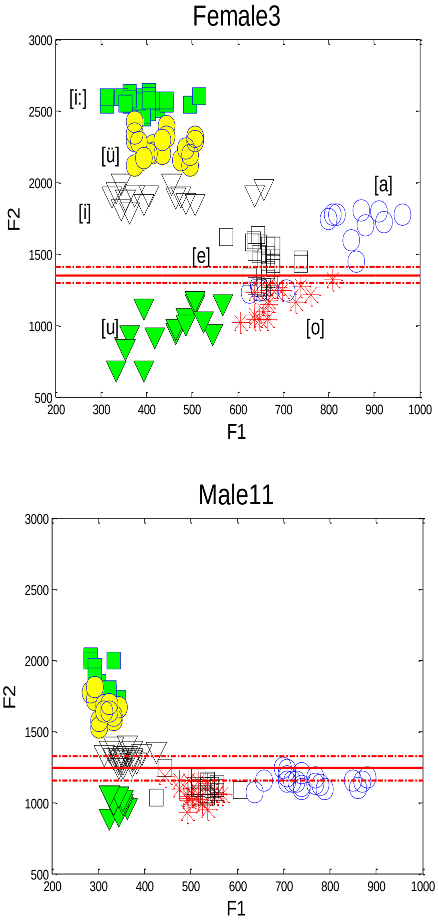

Figure 3: *Vowel plots of the two speakers (Female3 and Male11). Horizontal dashed lines indicate Sg2 interval (Mean ±Std). Different symbols represent different vowels. The vowel identities are labeled in the vowel plot for speaker Female3.*

Table 5. *Percentage of the front and back vowels on the correct side of Sg2 for each speaker. IDM (speaker ID (male)), IDF (Speaker ID (female)), Pfront (the percentage of the front vowels on the correct side of Sg2) and Pback (the percentage of the back vowels on the correct side of Sg2).*

IDM	Pfront (%)	Pback (%)	IDF	Pfront (%)	Pback (%)
11	90	85	3	100	97
9	100	100	18	100	100
7	100	98	12	100	100
17	100	100	16	100	100
1	100	98	2	100	100
4	100	100	6	100	100
15	100	100	8	100	100
13	100	95	10	100	100
19	100	100	14	100	95

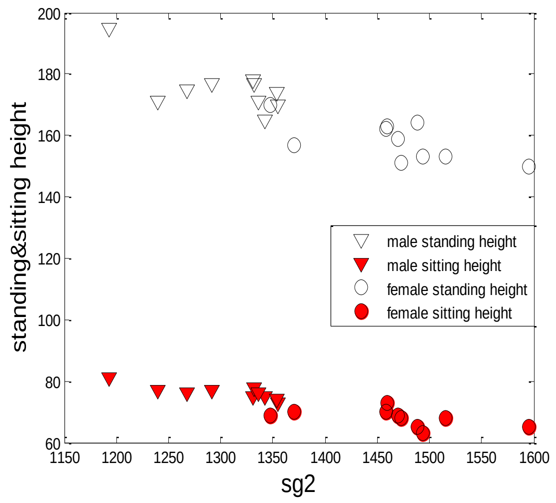

Figure 4: *Plots of the Sg2 versus standing and sitting height for male and female speakers*

Table 6: *Correlation between Sg2 and standing and sitting height for male, female and all speakers. (All correlations are significant)*

Gender	*Males*	*Females*	*Combined*
Correlation (height, Sg2)	-0.6776	-0.6602	-0.8699
Correlation(sitting height, Sg2)	-0.8063	-0.5488	-0.8733

For male speakers, the inverse correlation of Sg2 with sitting height is stronger than the correlation between Sg2 and standing height. A possible reason for this is that SGR frequencies are actually determined primarily by the 'acoustic length' (effective length of the subglottal system) [11]. Physiologically, the 'acoustic length' may be expected to be correlated with the size of the lungs and the length of the trunk, Sg2 is likely to be more strongly correlated with sitting height (trunk length). According to physiological data reported in [18], trunk length itself appears to be moderately correlated with overall body height. Such a relationship between trunk length and standing height seems to be partly responsible for the relatively weaker correlations with standing height [13]. In essence, the result that Sg2 has stronger correlation with sitting height but relatively weaker correlation with standing height for male speakers is reasonable.

As for the female speakers, the opposite trend is observed: correlation with sitting height is weaker compared to correlation with standing height. This can be explained by two interesting cases. Speaker 3 with a 1348Hz average Sg2 has a 170cm standing height but a 67cm sitting height, while Speaker 8 with a 1371Hz Sg2 has a 157cm standing height and a 77cm sitting height, which indicates that one is tall but has a short trunk length, while the other speaker is short but has a relatively long trunk length.

Overall, the correlation coefficients calculated by using all data from the 20 speakers are high for both standing height and sitting height cases. The correlation in the sitting height case is somewhat higher than standing height.

4.3. Relationships between F0 and Sg2

Table 7 shows the t value of Sg2 difference and their upper boundary. If the t value is less than the upper boundary, we accept the null hypothesis, i.e. the difference of Sg2 is zero. As shown in the table, under a significance level of 0.01, there is no statistically significant Sg2 difference between different parts of the vowel for all three vowels and all four tones.

Under the assumption that three measurements sufficiently capture change of Sg2 over time, we find no statistically significant Sg2 variation in any tone of any of vowel. Even if we loosen the significance level to 0.05 and upper bound of t reduces to 2.022, there are only 2 outliers, while in most cases we still accept the null hypothesis. A higher t value denotes a larger difference between mean Sg2 of different parts. And we notice that the fourth tone has the widest F0 range, whereas its t values are not significantly higher than the other tones. Therefore it can be inferred that despite the large range of F0 variation within a vowel, Sg2 tends to stay at about the same value.

Table 7. *The results of t-test of Sg2 comparisons between different vowel parts. (significance level of 0.01)*

Tone	Parts of Vowel	t(upper bound)	t(/u:/)	t(/i:/)	t(/a:/)
First	Begin-Mid	2.708	1.282	1.900	1.079
	Mid-End		0.263	0.882	0.103
	Begin-End		1.218	0.907	0.103
Second	Begin-Mid		0.592	0.533	1.435
	Mid-End		2.059	1.045	0.953
	Begin-End		1.678	1.333	1.731
Third	Begin-Mid		1.252	0.603	1.194
	Mid-End		1.849	0.742	1.527
	Begin-End		1.308	1.285	2.066
Fourth	Begin-Mid		0.848	0.674	0.288
	Mid-End		0.233	0.882	0.641
	Begin-End		1.107	1.287	0.156

5. Conclusions

Our results illustrate that, as in other languages studied before, in Mandarin, Sg2 is the boundary between front and back vowels. It is also important to note that the neutral vowels in Mandarin are not always located in the Sg2 interval and they could be located outside the Sg2's upper and lower boundary.

We also showed that, Sg2 is more inversely correlated with sitting height (trunk length) than standing height for our database. This result is more concrete for male speakers, since there were 2 female speakers who have an unexpected ratio of trunk length-to-standing height.

The result of the correlation between Sg2 and F0 is also illustrated. Paired t-test was used to test the Sg2 of different regions of the vowel, which represent different F0 values. We illustrated that there is no statistically-significant difference between Sg2 with different F0 values within a vowel.

For future work, further interaction and correlation between SGRs, formants and F0 will be studied in Mandarin, such as the coupling effect between subglottal acoustic system, source model and vocal tract. The results in the paper will also have application in automatic speech recognition and height estimation for Mandarin speakers.

6. References

[1] Shizhen Wang, "Rapid Speaker Normalization and Adaptation with Applications to Automatic Evaluation of Children's Language Learning Skills", Ph.D. dissertation, University of California, Los Angeles, 2010.

[2] Lulich, S. M., Morton, J. R., Arsikere, H., Sommers, M. S., Leung, G. K., and Alwan, A., “Subglottal resonances of adult male and female native speakers of American English”, The Journal of the Acoustical Society of America, 132(4): 2592-2602, 2012.

[3] S.M. Lulich, “The role of lower airway resonances in defining vowel feature contrasts”, PhD Dissertation, MIT, Cambrige, MA, 2006.

[4] S.M. Lulich, A. Bachrach and N. Malyska, “A role for the second subglottal resonance in lexical access”, J. Acoust. Soc. Am., col. 122(4), pp. 2320-2327, 2007.

[5] M. Sonderegger, “Sublottal coupling and vowel space: An investigation in quantal theory”, Undergraduate thesis, MIT, Cambrige, MA, 2004.

[6] Stevens, K. N., Acoustic Phonetics, Cambridge, MA: MIT Press, 1998.

[7] Lulich, S. M., “Subglottal resonances and distinctive features”, Journal of Phonetics, doi:10.1016/j.wocn.2008.10.006, 2009.

[8] Madsack, A., Lulich, S. M., Wokurek, W., Dogil, G., “Subglottal resonances and vowel formant variability: A case study of high German monophthongs and Swabian diphthongs”, in Proc.LabPhon, 11:91–92, 2008.

[9] Jung, Y., “Subglottal effects on the vowels across language: Preliminary study on Korean”, JASA 125:2638, 2009.

[10] Csapó, T. G., Bárkányi, Z., Gráczi, T. E., Bohm, T., and Lulich, S. M., "Relation of formants and subglottal resonances in Hungarian vowels", Interspeech, 2009.

[11] S. M. Lulich, A. Alwan, H. Arsikere, J. R. Morton, and M. S. Sommers, “Resonances and wave propagation velocity in the subglottal airways”, The Journal of the Acoustical Society of America, 130(4): 2108-2115, 2011.

[12] Arsikere, H., Leung, G. K., Lulich, S. M., and Alwan, A., "Automatic height estimation using the second subglottal resonance", ICASSP, 2012 IEEE International Conference on, 2012.

[13] Arsikere, H., Leung, G. K. F., Lulich, S. M., and Alwan, A. “Automatic estimation of the first three subglottal resonances from adults’ speech signals with application to speaker height estimation”, Speech Commun, http://dx.doi.org/10.1016/j. specom.2012.0.004, 2012

[14] Sjölander, K., Wavesurfer. KTH, Stockholm, Sweden. Online: https://www.speech.kth.se/wavesurfer/, 2012.

[15] Gráczi, T. E., Lulich, S. M., Csapó, T. G., and Beke, A., “Context and speaker dependency in the relation of vowel formants and subglottal resonances - evidence from Hungarian”, Proceedings of Interspeech 1901–1904, 2010.

[16] Zhang, T.D. and Wang, W.K., “Probability theory and Mathematical Statistics Fourth Edition”, Zhejiang University. Higher Education press, Beijing, China, 2008.

[17] McDonald, J.H., “Handbook of biological statistics”, Department of Biological Sciences, University of Delaware, Delaware, DE, United States, Online: http://udel.edu/~mcdonald/ statpaired.html, 2014.

[18] Hrdlička, A., The Old Americans, The Williams and Wilkins Company, Baltimore, MD, 1925.