An MRI-based Acoustic Study of Mandarin Vowels

Yuguang Wang[1, 3], *Jianwu Dang*[1, 2, 3,*], *Xi Chen*[3],
Jianguo Wei[1, 3], *Hongcui Wang*[1, 3], *Kiyoshi Honda*[1, 3]

[1]School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] School of Information Science, Japan Advanced Institute of Science and Technology, Japan
[3]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, China

heureux@tju.edu.cn,jdang@jaist.ac.jp

Abstract

This study aims at exploring detailed acoustic characteristics of Mandarin sustained vowels based on magnetic resonance imaging (MRI). Cross-sectional area functions of the vocal tract for Mandarin vowels were extracted from volumetric MRI images obtained from eight Chinese speakers (six-male and two-female subjects). Acoustic analysis was performed for the vowels recorded from each speaker. The acoustic analysis includes comparisons between measured and calculated formants, vowel space, simplified models for Mandarin vowels, and correlation between vocal tract length and formant frequencies. Mean absolute errors of calculated formants across 10 Mandarin vowels over eight subjects ranged from 4.6% to 11%. Simple-model also offers good approximations for Mandarin vowels. Since the simple model can give relatively clear relation between acoustic characteristics and vocal tract properties, it is useful for intuitively understanding the mechanism of speech production. The correlation analysis suggests that there exist negative correlations between vocal tract length and first four formants for Mandarin vowels in general, which showed a clearer relation than a recent study by another research group.

Index Terms: mandarin vowels, MRI, acoustic analysis

10.21437/Interspeech.2013-157

1. Introduction

MRI studies on vowels have been reported on several languages, such as American English [1], French [2], and Japanese [3], etc. By our knowledge, there is no systematic MRI study of Mandarin vowels. Zhu and Honda (2002) carried out MRI-based articulatory and phonological study of diphthongized “o” and “e” in Chinese [4]. Our previous study concerned with vocal tract morphological characteristics of Mandarin vowels using MRI [5]. However, none of these studies have looked into detailed acoustic properties of Mandarin vowels using MRI.

The purpose of this study was set at systematic examinations of the acoustic characteristics of sustained vowels in Mandarin Chinese using native Chinese speaker’s MRI data obtained during vowel production. In what follows, we first describe the procedures for acquiring MRI data of the vocal tract and the methods to extract vocal tract area functions. In our study, eight subjects (including six males and two females) have participated in the MRI experiment. The acoustic properties of the vocal tract during production of ten Mandarin vowels are analyzed based on the area functions and their acoustic simulations using the transmission line model.

2. Method

In this section, we introduce the procedures for obtaining MRI data, including the selection of speech materials and subjects, the setup of MRI experiments, and the speech recording. Then, we describe the method to extract vocal-tract area function from volumetric MRI images and formants from speech data.

2.1. Speech material and selection of subject

As for sustain vowel, the present study covers all the ten Mandarin Chinese /a, o, e, i, u, ü, ê, (s)i, (sh)i, er/, corresponding to the IPA symbols [a, ɔ, ɤ,i, u, y, e,ɿ, ʅ, ɚ], respectively [6]. Vowel [ɿ] and [ʅ], adopted by sinologists in China, are two “apical vowels” that appear after apical dental [z, c, x] and retroflex fricatives/affricates [tʂ, tʂ^h, ʂ, r]. The ten vowels were realized by reading aloud Chinese characters of “啊喔屙衣乌淤噎思诗儿”.

In the present study, we employ eight subjects including six male and two female subjects who have no speech pathology and speak fluent Mandarin. The subjects grew up in and around Beijing, having no dialectal deviation. They are trained to ensure the articulatory stability so as to obtain high quality of MRI images. In the training, the subjects were asked to produce the speech materials sustainedly in a supine position while scan noise via earphones to adopt for the MRI experiment.

2.2. MRI data acquisition

MRI data were acquired with the Shimadzu-Marconi ECLIPSE 1.5T Power Drive 250, installed at the ATR Brain Activity Imaging Center (ATR-BAIC). The synchronized sampling method (SSM) with external trigger pulses was applied to acquire static 3D shapes during production of vowels skipping intermittent inhalation periods. Each subject repeated each vowel about 30~36 times, sustaining for 3s, which allows stable and natural articulation. The trigger device presents noise burst trains to the subject through a headset and outputs the scan pulses to the MRI scanner to synchronize the data acquisition to subject’s vowel production. The subject was instructed to listen to the noise burst trains to pace the utterance, while the MRI scanner initiates data acquisition synchronized with the trigger pulses.

2.3. Extraction of vocal tract area function

MRI has a problem in imaging the teeth or bony structures because these calcified structures lack hydrogen, and thus no resonance signal would be produced. Consequently, the teeth have the same grayscale as that of the airway, where is no

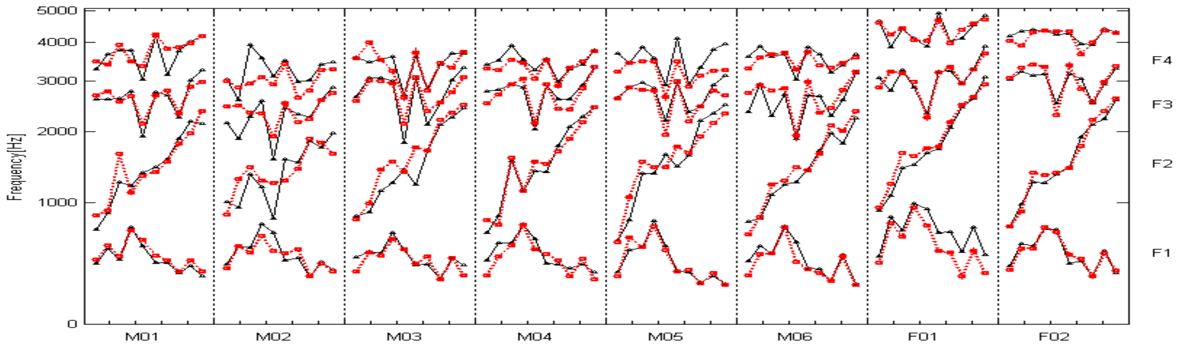


Figure 2: *Comparisons of formant frequencies (F1-F4) between real speech and area function based calculations. The solid lines represent calculated formants, and the dotted lines represent calculated formants. Each column corresponds to one subject. Ten vowels are [u,ɔ, ɤ, a, ɚ, ɿ, ʅ, y, ɛ , i].*

boundary between them. In this study, we adopted a method proposed by Takemoto et al. (2004) to obtain the teeth-air boundary and reconstructed the vocal-tract shape from MRI data [7]. To do so, the "digital jaw casts" were firstly constructed based on a special MRI experiment and then manually superimposed onto the original MRI volumetric data for each vowel.

The vocal-tract airway was extracted from the tissues by threshold segmentation algorithm. After the segmentation of the vocal tract, area function can be extracted from the vocal tract. In the first step, the midline of the vocal tract was calculated on the mid-sagittal plane. In the second step, a set of grid lines was formed. Each grid line is perpendicular to the midline and intersects with the midline at the mid-point of the line segment. In the grid system, the distance between the centers of two adjacent lines is constant. Figure 1 shows the final grid system. In the third step, the 3D vocal tract was re-sliced based on the grid system. The area function of the vocal tract is obtained by measuring the area of the cross-section from the glottis to the lips.

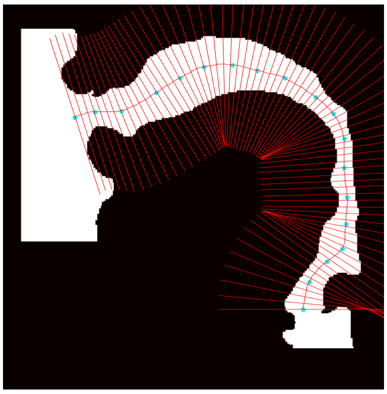


Figure 1: *The center line of the vocal tract on the mid-sagittal plane and grid lines those are perpendicular or nearly perpendicular to the center line. The distance between two centers of the grid lines is 2mm.*

2.4. Speech recording and formants extraction

Since MRI emits noise, speech signals recorded in the MRI room can hardly serve for analysis. Speech sounds were separately recorded from each subject in a soundproof room. Three repetitions of each vowel were uttered sustainedly for more than 3 seconds. The stable segment of the recorded vowels was chosen to extract the lower four formants using the Praat software. Each measured formant is the mean value of all the three repetitions.

The transfer functions were calculated for the MRI-based area functions of the vocal tracts using the transmission line model. The formant frequencies are obtained from the transfer functions, and used for evaluation.

3. Acoustic analysis of Mandarin vowels

The acoustic analysis was conducted for comparisons between measured formants and calculated formants, vowel space, simplified models for Mandarin vowels, and correlation analysis between vocal-tract length and formant frequencies.

3.1. Formant comparison

In order to evaluate the reliability of the area functions extracted from the 3D MRI data, the lower four formants were extracted from the transfer functions using a peak detection algorithm and compared with that of natural speech sounds. Figure 2 shows calculated and measured formants of ten Mandarin vowels of eight subjects. Solid lines indicate the measured formants from natural speech, and dotted lines indicate formants calculated from area functions. The vowels are ordered roughly by F2 in the ascending order.

Mean absolute errors of formants over all vowels for the eight subjects are 7.1%, 11%, 7.4%, 7.0%, 7.7%, 7.4%, 8.5% and 4.6%, respectively. One can see that calculated F1-F4 frequencies almost are consistent with the measured ones for the subjects M01, M03, M04, M05, M06, F01 and F02. Where, some large errors are found in Subject F01: the calculated F1 for vowel [ɚ, ɿ, ʅ, y, ɛ , i] were much lower than the measured one. For Subject M02, a relatively larger discrepancy is seen in the calculated and measured F2 frequencies.

3.2. Vowel space

Figure 3 shows vowel plots with ellipses in a F1-F2 space, which was extracted from uttered Mandarin vowels of eight subjects. Each ellipse is drawn with radii of standard deviations along the two principal axes for each vowel cluster and labeled by its corresponding vowel IPA symbols at the center. In the F1-F2 space, vowels [i], [a] and [u] are located in the vertexes defining the vowel triangle for Mandarin vowels, which is consistent with that in other languages.

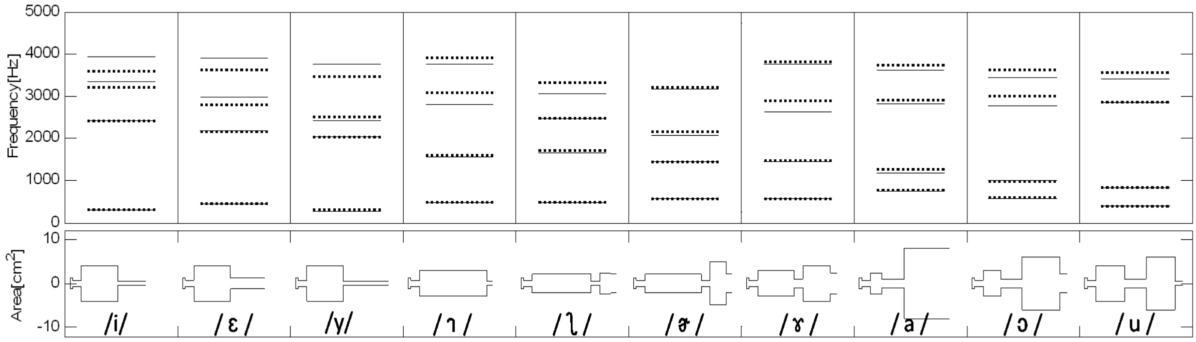


Figure 4: *Simple-model approximations for ten Mandarin vowels. Simple models are shown on the lower panel (four to six segments); the dotted lines in upper plot represent the first four formants calculated from area function, and the solid lines represent formant frequencies of simple models. The vowels are ordered by the second formant frequency of the simple models.*

However, the ellipses for the group of [ɛ, i, y] are distributed near the left high corner, since F1 and F2 are distant for this group of vowels. As for the two apical vowels, their ellipses are in the middle of the plot. Those vowel groups are well organized in the vowel space. As shown in the F2-F1 plot, there are some cases of overlapping vowel ellipses, among [i, y, ɛ], between [ɿ] and [ʅ], and between [ɤ] and [ɚ].

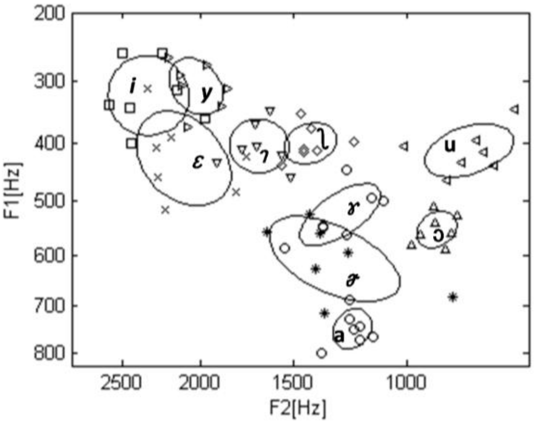


Figure 3: *Vowel ellipses for ten Mandarin vowels in F1/F2 plane.*

3.3. Simple models for Mandarin vowels

It is possible to predict the formant patterns characterizing oral vowels by using a simple model of the vocal tract that consists of four or more tube segments. The first two tube segments, which correspond to the laryngeal ventricle and laryngeal vestibule, were kept constant across the all vowels. For some front vowels such as [i] and [y], the vocal tract was modeled with two segments, except the fix segments. For vowels [u, ɔ, ɤ, ɚ, ʅ], the vocal tract were divided into two cavities and thus modeled with two cavities using four segments.

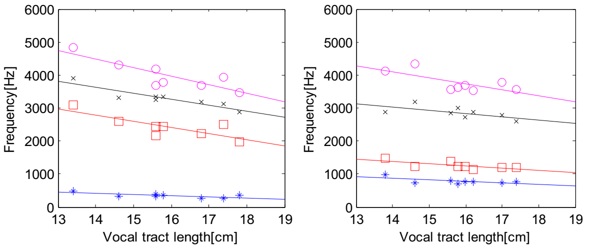
Figure 4 shows the configuration of the simple-model approximations and the first four formant frequencies calculated from the simple models and the transmission line model with the vocal tract area functions. The ten simple model's order is arranged according to F2. Except for a few cases of F3 and F4, most of the formant frequencies for the simple models agree well with that calculated from the transmission line model for the eight sets of the vocal tract area functions. The mean absolute error of formant frequencies between the simple model and the area functions is 3.8%, which indicates that these simple models offer good approximation for Mandarin vowels. These simple models provide clear relation between acoustics and vocal tract for ten Mandarin vowels.

3.4. Correlations between vocal tract length and formant frequencies

A correlation analysis between the vocal tract length measured during speaking and the formant frequencies of the speech sounds was carried out for Mandarin vowels. Figure 5 shows the correlation between the vocal tract length and F1-F4 for vowel [a] and [i]. Two female subjects' data are on the left (shorter than 14cm) and six male subjects' data are on the right part (longer than 15cm). Consequently, the vocal tract length ranges from about 13cm to 18cm. One can see that the formant frequencies are fall around the regression line quite closely, where all the formants decrease less or more as the vocal tract length increases. Table 1 shows detailed correlation analysis between the vocal tract length and formants for Mandarin vowels. For individual vowels, the correlation coefficients are not so high. When putting all data together, however, the general correlation gets large. The correlation for F1-F4 are -0.78, -0.85, -0.9 and -0.91, respectively. The results suggest that there is a negative correlation between formants and the vocal tract length for Mandarin vowels in general.

Table 1. *Correlated coefficient between vocal tract length and first four formant frequencies for ten Mandarin vowels.*

	F1	F2	F3	F4
a	-0.58	-0.75	-0.66	-0.72
ɤ	-0.66	-0.29	-0.84	-0.66
ɔ	-0.76	-0.8	-0.39	-0.49
u	-0.57	-0.35	-0.72	-0.82
i	-0.7	-0.79	-0.91	-0.85
ɛ	-0.81	-0.7	-0.79	-0.88
y	-0.72	-0.37	-0.75	-0.91
ɿ	-0.73	-0.33	-0.83	-0.74
ʅ	-0.76	-0.84	-0.41	-0.7
ɚ	-0.76	-0.46	-0.79	-0.87
All data	-0.78	-0.85	-0.9	-0.91



(a) Vowel [i] (b) Vowel [a]

Figure 5: *Correlation between vocal tract length and formants (F1, F2, F3 and F4) for vowel [i] and [a].*

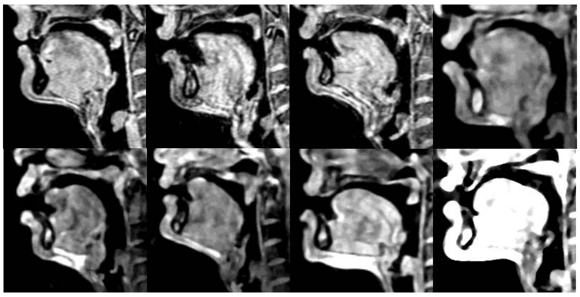


Figure 6: *MRI mid-sagittal view of vocal tract for eight subjects during production of Mandarin vowels [ʅ].*

4. Discussions

Recently, a few studies reported the acoustic and articulatory characteristics of two apical vowels [ɿ, ʅ] [8, 9]. Lee (2005) assumed that the lower F3 for the posterior apical vowel [ʅ] is due to the pharyngealization (the constriction at the pharyngeal cavity). However, our MRI data shows no significant constriction in the pharyngeal cavity (see Figure 6). For vowel [ʅ], the vocal tract was divided into a front cavity (including the sublingual cavity) and a long back cavity by the palate constriction. According to the simple model in Figure 4, the F3 for vowel [ʅ] is actually related the front cavity.

Hatano et al. conducted a correlation analysis for Japanese vowels using fifteen adult male subjects [10]. Their study showed that negative correlations between the vocal tract length and the formants of F1, F2 and F4 were found only for Japanese vowel [e]. While, our study showed that there exist relatively high negative correlations between vocal tract length and first formants for Mandarin vowels. To clarify the difference between their result and ours, we pick up the six male subjects and analyze the correlation. The correlation between F1-F4 and vocal tract length are 0.18, -0.30,-0.78 and -0.46 for all vowels, where the correlations become very weak for adult male subjects. The cause of the difference may be that variance of the vocal tract length for male subject alone is not large enough for such a calculation. Actually, the vocal tract length range in Hatano's study was from about 16cm to 18cm, which is much smaller than the range from 13.4cm to 19.4 cm in this study. For a larger range of the vocal tract length, the correlation is consistent with the common knowledge that the vocal tract becomes longer, formants frequencies decrease in general.

5. Conclusions

MRI was used to image the vocal tract shape for eight subjects during the production of ten sustained steady vowels of Mandarin Chinese. The acoustic analysis was conducted for comparisons between measured formants and calculated formants, vowel space, simplified models for Mandarin vowels, and correlation analysis on Mandarin vowels. The mean absolute errors of formants across all the vowels for each subject ranged from the minimum 4.6% to the maximum 11%.The absolute error was 3.8% between the simple models and the transmission line model. This indicates that the simple models offer good approximations for vocal-tract area functions. Since the simple model can give relatively clear relation between acoustic characteristics and vocal tract properties, it is useful for intuitively under the mechanism of speech production. And the correlation analysis suggests that there exist negative correlations between vocal tract length and first formants for Mandarin vowels.

6. Acknowledgements

This work was supported in part by a grant from the Major State Basic Research Development Program of China (973 Program: No. 2013CB329301), and in part by grants from the National Natural Science Foundation of China (Key Program No. 61233009, and General Program No. 61175016).

7. References

[1] Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. 100,537–554.

[2] Badin P., Bailly G., et al., "A three-dimensional linear articulatory model based on MRI data", Proceedings of the Third ESCA/COCOSDA International Workshop on Speech Synthesis: 249-254, 1998.

[3] Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., and Honda, K., Acoustic roles of the laryngeal cavity in vocal tract resonance," J. Acoust. Soc. Am., 120, 2228–2238, 2006.

[4] Zhu C, Honda K. "An MRI-based articulatory and phonological study of diphthongized "o" and "e" in Chinese," Linguistics and Phonetics, 2002.

[5] Wang, Y., Wang, H., Wei, J.., Dang, J., "Detailed Morphological Analysis of Mandarin Sustained Steady Vowels", ISCSLP, Hong Kong, China, 2012.

[6] Wu Z., Zhao J., Zhu Z., L Y., " An introduction to Modern Mandarin Chinese Speech", Sinolingua Press, 1992, pp. 30.(In Chinese)

[7] Takemoto, H., Kitamura, T., Nishimoto, H., and Honda, K. (2004). "A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions," Acoust. Sci. & Tech. 28, 33–38.

[8] Lee, W. S. (2005). "The articulatory and acoustical characteristics of the apical vowels in Beijing Mandarin." Acoustical Society of America Journal 118: 2027-2027.

[9] Lee, S. (2011). "An articulatory and acoustic investigation of Mandarin apical vowels." The Journal of the Acoustical Society of America 130(4): 2550-2550.

[10] Hatano, H., Kitamura, T., Takemoto, H., Mokhtari, P., Honda, K., Masaki, S., (2012). "Correlation between vocal tract length, body height, formant frequencies, and pitch frequency for the five Japanese vowels uttered by fifteen male speakers", Interspeech 2012, Portland USA.