Individual difference and acoustic effect of female laryngeal cavities

Jing Li[1], *Kiyoshi Honda*[1], *Ju Zhang*[1], *Jianguo Wei*[1, *]

[1] Tianjin Key Laboratory of Cognitive Computation and Application, Tianjin University, Tianjin, China

jingli_tju@yeah.net, khonda@sannet.ne.jp

Abstract

This study examines the acoustic effect of the laryngeal cavity of female speakers on the higher vowel spectra. To do so, MRI data of vowels /a/ and /i/ obtained from three female speakers were analyzed with data from a male speaker as reference. 3D vocal-tract shapes were extracted from the MRI data and printed as solid mechanical models. Transfer functions of the models' vocal tracts were estimated by a transmission line model. Individual variations of the laryngeal cavity were described by the area functions of the cavity. It is observed that female laryngeal cavities tend to have relatively stable cross-sectional areas in comparison to the male's cavity. Acoustic experiments were conducted using the solid vocal-tract models with different glottal conditions simulated by different hole sizes of the glottal coupler. The result suggests that female laryngeal cavities exhibit spectral changes in the wider frequencies at about 3-4.5 kHz in vowels /a/ and /i/ in comparison of data between open and closed conditions of the glottis.

Index Terms: laryngeal cavity, vocal tract, solid models, MRI, vowel

1. Introduction

The laryngeal cavity forms a lower part of the vocal tract above the vocal folds and divides into the laryngeal ventricle and laryngeal vestibule. Acoustic characteristics of the laryngeal cavity and the relationship between cavity resonances and vowel spectra have been investigated by many researchers. Kitamura et al. [1] reported that the hypopharyngeal cavities affect the frequency range of spectra above 2.5 kHz. The laryngeal cavity generates closed tube resonance during the closed period of the glottis, which diminishes when the glottis opens [2]. Takemoto et al. [3] revealed that the laryngeal cavity gives rise to the fourth formant of the vocal tract, with little effects to other formants.

Most of the studies on the acoustic characteristics of the laryngeal cavity were based on male subjects, and female data were few. The previous studies described above indicate that the laryngeal cavity of male subjects produces a regional resonance at about 3 kHz as a Helmholtz resonance. Female laryngeal cavities appear different from males' in size and shape, and their acoustic characteristics may also be different. As Figure 1 shows, the size of the female laryngeal cavity is smaller than that of the male's, suggesting that the resonance frequency may be higher. The shape of the female laryngeal cavity often looks like a simple straight tube rather than a Helmholtz resonator.

Honda et al. [4] conducted acoustic experiment using mechanical vocal-tract models with the same result of male laryngeal cavity resonance at about 3 kHz, and they found that the female laryngeal cavity causes spectral changes in the wider spectral range. In the report, however, the hole size of the glottal coupler simulating the open glottis could be too large for the female glottis, which may have caused the excessively wider frequency effect on vowel spectra. Also, the study on only one female subject does not answer to the questions as to whether the result is common to other female laryngeal cavities or whether the female laryngeal cavity resonance arises at 3-3.5 kHz region.

The purpose of present study is to examine the role of the female laryngeal cavity in individual acoustic characteristics of speech sounds. In this study, we use MRI data obtained from three female and one male subjects during the production of Chinese vowels /a/ and /i/. Then, the mechanical vocal-tract models were built, and the acoustic experiments were conducted to excite the models with white noise from a horn driver unit. We used different hole sizes of glottal couplers to simulate the open and closed glottis conditions to observe the effect of the laryngeal cavity on spectra.

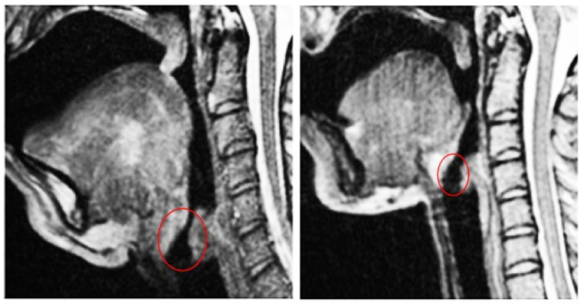

Figure 1: *Shapes of (a) male laryngeal cavity and (b) female laryngeal cavity.*

2. Materials and methods

2.1. Acquisition of MRI data

Three female speakers, CR, LH, SC, and one male speaker WS were served as the subjects of this study. They are all native Mandarin speakers from northern China, and no one reported any history of speech disorders.

The MRI dataset was obtained by a Siemens Verio 3T MRI scanner at the ATR Brain Activity Imaging Center (ATR-BAIC) in Kyoto, Japan. The subjects lying supine in the MR scanner pronounced sustained Chinese vowels /a/ and /i/. The speech sounds were recorded during the scan and at separate post hoc sessions. The subject's head was stabilized by foam rubber to minimize the movement during scanning. The

978-1-5090-4294-4/16/$31.00 ©2016 IEEE

synchronized MRI scans acquired were conducted to acquire volume data in the sagittal plane with the 2-mm slice thickness, 256×256 mm field of view, and 512×512 pixel image size. Teeth images were also obtained.

2.2. Building solid vocal-tract models

The MRI data were resampled to have voxel resolution of 0.5 ×0.5×1 mm. Then, the boundaries of the vocal tract were manually traced including the upper and lower lips and excluding the nasal cavities. The traced images were interpolated to have 0.5-mm slice thickness by a cubic inter-slice interpolation method. The vocal-tract regions of the interpolated images were converted into a 3D vocal-tract model by a 3D image editor (Materialise MIMICS). The shapes of the 3D model were smoothed by the editor and exported into the STL files. The teeth were superimposed to the vocal-tract models in 3D space [5]. Each model was processed to have an outward wall of 3-mm thickness. Finally, the wall model was trimmed to have the glottal and lip boundaries and then printed by a 3D printer (Formlabs F1+).

In the acoustic experiment, the mechanical models were excited by continuous white noise at the glottal end. The glottal end of the model was fixed on a glottal coupler with a circular hole in the center, which was placed on a wide-range horn driver unit (UNI-PEX P-800N). The white noise was amplified and output the driver unit. A condenser microphone (Behringer ECM8000) was placed at 10 cm from the lip end of the models, and the acoustic responses were recorded with a digital sound recorder (Roland DUO-CAPTURE EX) at a rate of 44.1 kHz sampling.

2.3. Simulation of vocal-tract models

The recorded signals of the models in the acoustic experiment were analyzed by the Imai's cepstral method [6], and the spectral envelops were finally obtained. To evaluate the accuracy of the model, acoustic simulations based on the transmission line model were calculated, and the formants of the subjects' natural vowels were measured using the cepstrum method.

To perform the simulation with the transmission line model, the area functions of the vocal-tract model for all the subjects' vowels /a/ and /i/ were extracted first. The method used for measuring area functions was similar to Takemoto et al. [7]. The mid-sagittal vocal-tract images were obtained from the 3D STL models. Then, the centroid points of the vocal tract from the glottis to the lips were calculated to define the midline. The cross-sectional areas of the vocal tract were measured at 2-mm intervals along the midline.

The transfer functions for all the vocal-tract models were calculated based on the transmission line model [8], where plane wave propagation is assumed in the vocal tract with concatenated short cylindrical tubes. The radiation impedance at the lips was approximated by the method in Causeé et al. [9]. The furthest section that shows the closed circumference was considered as the vocal-tract open end that extends for 0.3 D, where D is the diameter of the end section [10]. The piriform fossa was not considered in the simulation for its effects.

2.4. Experiment on the laryngeal cavity

The laryngeal cavity is supposed to generate its own resonance with the glottis being closed, and the resonance disappears when the glottis opens [2]. To examine the acoustic effect of the laryngeal cavity in this study, different glottal conditions were simulated using four hole sizes of the glottal couplers. As shown in Figure 2, the glottal coupler having a round hole of different diameters in the center was placed between the glottal end of the solid vocal-tract model and the throat of the horn driver unit.

In Honda et al. [4], 1.2-mm and 4.0-mm hole sizes of the glottal couplers were adopted to simulate the closed and open glottis conditions respectively, for their male and female models. Since the 4.0-mm hole size of the glottal coupler may be too large for the female glottis, we employed three sizes of 2.0-, 3.0- and 4.0-mm hole diameters for the open glottis conditions, while using the 1.2-mm diameter for the closed glottis condition. Then, a comparison was attempted to find which hole size is the most suitable. Figure 2 depicts the setting of the experiment.

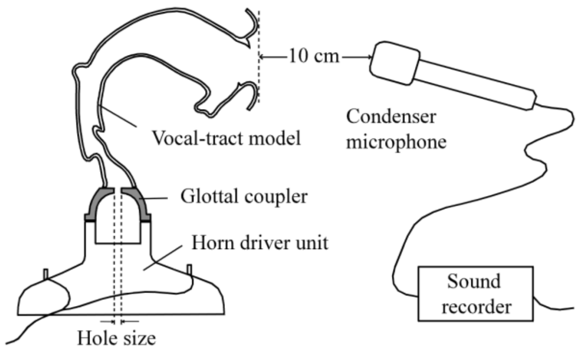

Figure 2: *Setup for the acoustic experiment on the laryngeal cavity with different hole sizes of the glottal coupler.*

3. Results

3.1. Evaluation of models' accuracy

The area functions of the extracted 3D vocal-tract models for three female and one male subjects are shown in Figure 3. Vocal-tract transfer functions were calculated to evaluate the accuracy of the models. Table 1 shows the lower four formants acquired from the natural vowels and the transfer functions, which are labeled with the prefixes "n" (natural) and "t" (transfer function), respectively. The percent errors in Weber fraction between the two measures are marked by "e". The mean absolute percent error is 4.3% (within the difference limen of formants in Flanagan [11]), which indicates that the vocal-tract shapes were successfully extracted, and the 3D vocal-tract models were accurate.

3.2. Morphological analysis of laryngeal cavity

The gray portion in each panel of Figure 3 corresponds to the laryngeal cavity in the vocal tract. Figure 4 shows the area functions of the laryngeal cavities for vowels /a/ and /i/ measured with 1-mm intervals. For the male subject WS, the lower six sections represent the ventricle region, and the seventh to the twenty-fifth sections correspond to the vestibule region. For the female subjects, the first to the fifth sections represent the ventricle region, and the vestibular regions are shown by the rest sections. The mean and standard deviation of the ventricle and vestibule areas (cm^2) for each subject are calculated in Table 2. The area functions and the numerical values indicate that the ventricle and vestibule of female

Table 1. *First to fourth formants from the natural vowels and the transfer functions of the 3D vocal-tract models at vowels /a/ and /i/. "n" and "t" are natural and calculated formants, respectively. "e" is percent error between the two.*

	CRa	LHa	SCa	WSa	CRi	LHi	SCi	WSi
nF1	835	900	964	565	390	403	357	339
nF2	1489	1441	1580	1102	3147	2634	2930	2104
nF3	3518	3149	3264	2772	3687	3431	3997	2697
nF4	4024	3855	3994	3391	4845	4410	5056	3416
tF1	801	946	1036	611	346	441	321	311
tF2	1441	1431	1656	1026	3171	2666	2926	2101
tF3	3461	3026	3266	2606	3966	3416	4056	2691
tF4	—	3716	3981	3686	4781	4186	4951	3136
eF1	-4.1	5.1	7.5	8.1	-11.3	9.4	-10.1	-8.3
eF2	-3.2	-0.7	4.8	-6.9	0.8	1.2	-0.1	-0.1
eF3	-1.6	-3.9	0.1	-6.0	7.6	-0.4	1.5	-0.2
eF4	—	-3.6	-0.3	8.7	-1.3	-5.1	-2.1	-8.2

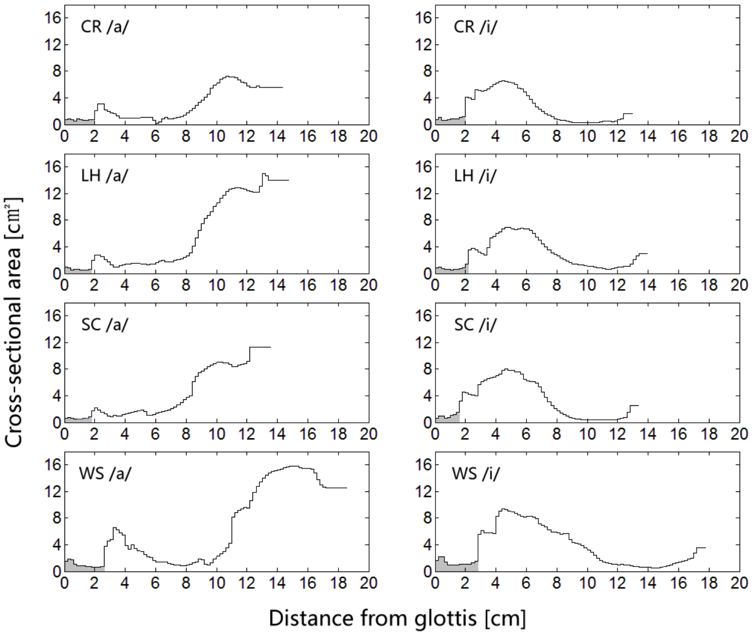

Figure 3: *Equal interval (0.2-mm) vocal tract area functions of vowels /a/ and /i/ for all the subjects. Gray regions indicate the laryngeal cavities.*

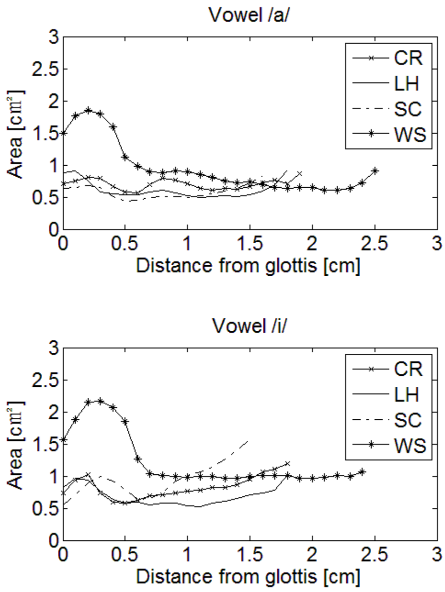

Figure 4: *Equal interval (0.1-mm) laryngeal cavity area functions for vowels /a/ and /i/.*

laryngeal cavities possess similar cross-sectional areas, having the smaller variation in shape compared with those for the male subject.

Table 2. *Mean and standard deviation of the ventricle and vestibule areas (cm²) for all the subjects.*

	CR	LH	SC	WS
$AVE_{ventricle}$	0.77	0.77	0.71	1.77
$STD_{ventricle}$	0.13	0.15	0.16	0.30
$AVE_{vestibule}$	0.75	0.61	0.78	0.88
$STD_{vestibule}$	0.16	0.12	0.32	0.16

3.3. Results of acoustic experiment

Figure 5 shows the spectra of the vocal-tract models' output signals in different glottal conditions. The black, blue, green and red lines correspond to the 1.2-, 2.0-, 3.0- and 4.0-mm hole sizes of the glottal couplers, respectively. For the male subject (bottom two panels in Figure 5), the primary differences between the spectra are the spectral peaks at about 2.5-3 kHz both in vowels /a/ and /i/. The peaks in the closed glottis condition (1.2-mm) decrease in frequency in those with the 3.0- and 4.0-mm hole sizes of the glottal couplers, while the peaks are similar between the conditions with 1.0-mm and 2.0-mm hole sizes.

For the female subjects, the spectral differences are found in the wider frequency region. In vowel /a/, CR's and LH's spectra show attenuation at about 3-4.5 kHz in the 2.0-, 3.0- and 4.0-mm conditions, while they demonstrate amplification at about 5.5-6.5 kHz in the 4.0-mm condition. The spectral regions in SC's data are widely reduced at about 3-5.5 kHz in the 2.0-, 3.0- and 4.0-mm conditions, while the data shows an increase at about 5.5-6.5 kHz in the 3.0- and 4.0-mm

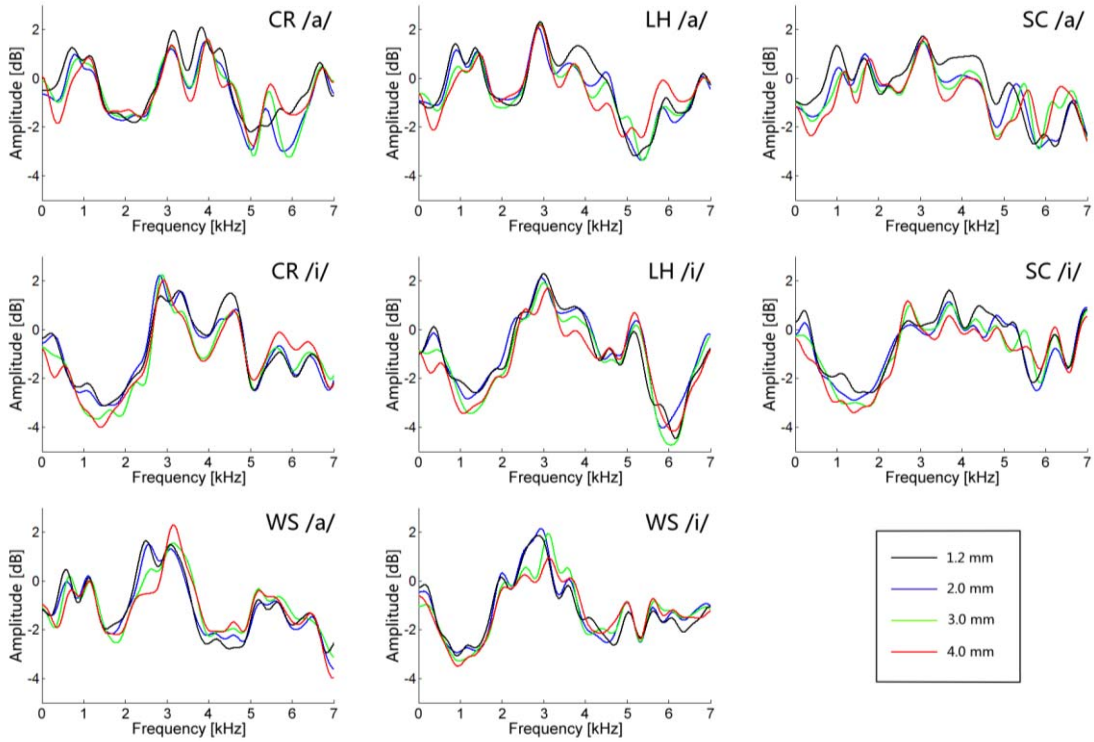

Figure 5: *Spectral envelopes of vowels /a/ and /i/ for female (CR, LH, SC) and male (WS) vocal tracts. The black, blue, green and red lines indicate the 1.2-mm, 2.0-mm, 3.0-mm and 4.0-mm hole sizes of the glottal couplers, respectively.*

conditions. In vowel /i/, the spectral attenuation at about 3-4.5 kHz is also shown in the 3.0- and 4.0-mm conditions, but the data for CR shows an increase at about 5.5-6.5 kHz in the 4.0-mm condition. It appears that the spectral amplification seen in the high frequency are caused by the large hole size of the glottal coupler such as 4.0-mm. For all the female spectra, the first formant gradually reduces as the glottis takes the closed to open conditions, while a little change appears at the second formant.

4. Discussions and Conclusion

In this study, acoustic effects of female laryngeal cavities were investigated using solid vocal-tract models for vowels /a/ and /i/ obtained from three female speakers and one male speaker as a reference. Laryngeal-cavity resonance in female tend to show a broader peak at higher frequencies than that in male. Despite the difference, our data support the previous notion that laryngeal-cavity resonance plays a critical role in signaling individual acoustic characteristics at the frequency region that is sensitive to the human ears [1].

The effectiveness of acoustic experiment with MRI-based 3D vocal-tract models was shown in the data (Table 1) comparing the spectra between the transfer functions and natural vowels, and the models' accuracy was also confirmed. Morphological analysis of female laryngeal cavities based on the area functions shows that the ventricle and vestibule have uniform areas compared with the male subject's large ventricle. One possible reason for the female uniform laryngeal cavity is that the ventricle is too small to be visualized on MRI volume with the 2-mm slice thickness.

The frequency region of laryngeal cavity resonance in our male subject was 2.5-3 kHz both in model simulations and recorded sounds, which is lower than that in previous studies (3-3.5 kHz). The difference appears due to the large size of the subject's ventricle showing the cross-sectional area of almost 2 cm^2. Since male laryngeal cavity forms a Helmholtz resonator, the larger the ventricle, the lower the frequency of resonance with a sharp peak. In our female subjects, the resonance frequency was in the region of 3-4.5 kHz, and the peak was broader. The higher resonance frequency appears due to the uniform shape with the shorter length of the laryngeal cavity.

In our acoustic experiment on models, four glottal conditions were tested to simulate open and closed glottis. In the results in both vowels, the glottal aperture with 1.2-mm diameter was found to be adequate for the closed glottis. The aperture of 3.0-mm diameter was judged suitable to simulate the open glottis from spectral attenuation at 3-4.5 kHz in the female data. The aperture size of 4-mm diameter appears to cause the large effects on spectra in the wider frequency range in female, demonstrating spectral deformation in the region of the first formant and amplification in the higher frequencies at 5.5-6.5 kHz. Thus, this aperture size was inappropriate to simulate the open glottis.

Finally, female laryngeal cavity resembles a straight tube with no visible ventricles in comparison with male laryngeal cavity of a Helmholtz-like resonator. Although this contrast derives partly from the lack of sufficient MRI resolution, the acoustic effect of the very small ventricles if any is thought to be negligibly small, and the short closed tube produces resonance with a wider frequency band change than in male's case. Those details should be examined in the future studies.

5. Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.61471259 and No.61573254).

6. References

[1] T. Kitamura, K. Honda, and H. Takemoto, "Individual variation of the hypopharyngeal cavities and its acoustic effects," *Acoustical science and technology*, vol. 26, no. 1, pp. 16-26, 2005.

[2] T. Kitamura, H. Takemoto, S. Adachi, et al. "Cyclicity of laryngeal cavity resonance due to vocal fold vibration," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2239-2249, 2006.

[3] H. Takemoto, S. Adachi, T. Kitamura, et al. "Acoustic roles of the laryngeal cavity in vocal tract resonance," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2228-2238, 2006.

[4] K. Honda, T. Kitamura, H. Takemoto, et al. "Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling," *Computer methods in biomechanics and biomedical engineering*, vol. 13, no. 4, pp. 443-453, 2010.

[5] K. Honda and J. Zhang. "Teeth visualization in MRI during vowel production for three-dimensional vocal tract modeling", unpublished.

[6] S. Imai and Y. Abe, "Spectral envelope extraction by improved cepstral method," *Electro. And Commun. in Japan*, vol. 62-A, no. 4, pp. 441-444, 1978.

[7] H. Takemoto, K. Honda, S. Masaki, et al. "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1037-1049, 2006.

[8] S. Adachi and M. Yamada, "An acoustical study of sound production in biphonic singing, Xöömij," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2920-2932, 1999.

[9] R. Caussé, J. Kergomard, and X. Lurton, "Input impedance of brass musical instruments—comparison between experiment and numerical models," *The Journal of the Acoustical Society of America*, vol. 75, no. 1, pp. 241-254, 1984.

[10] M. A. Redford, *The handbook of speech production*, John Wiley & Sons, 2015.

[11] J. L. Flanagan, "A difference limen for vowel formant frequency," *The journal of the Acoustical Society of America*, vol. 27, no. 3, pp. 613-617, 1955.