Contributions of the piriform fossa of female speakers to vowel spectra

Congcong Zhang[1], *Kiyoshi Honda*[1], *Ju Zhang*[1], *Jianguo Wei*[1,*]

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

congcongzhang11@126.com, khonda@sannet.ne.jp

Abstract

The bilateral cavities of the piriform fossa are the side branches of the vocal tract and produce anti-resonance(s) in the transfer function. This effect has been known for male vocal tracts, but female data were few. This study investigates contributions of the piriform fossa to vowel spectra in female vocal tracts by means of MRI-based vocal-tract modeling and acoustic experiment with the water-filling technique. Results from three female subjects indicate that the piriform fossa generates one or two dips in the frequency region of 4-6 kHz. Experiment for filling each cavity separately demonstrated that the spectral patterns vary greatly across the conditions for filling both, left and right cavities. The result suggests that the cavities interact with each other, and the phenomena contribute to determining spectral envelopes in the higher frequencies.

Index Terms: speech production, model simulation, piriform fossa, anti-resonance, vocal tract.

1. Introduction

The piriform fossa is a part of the vocal tract located near the larynx, consisting of a pair of bilateral small cavities (Figure 1). It offers a side branch to the vocal tract and contributes to expressing individual vocal characteristics and adjusts vocal quality by cavity deformation [1-5].

In previous studies, using magnetic resonance imaging (MRI), Dang and Honda [1] investigated acoustic characteristics of the piriform fossa in mechanical models and humans. In the experiment, injecting water in the piriform fossa of the lower part of the vocal-tract models, it was observed that the piriform fossa generates anti-resonance in the narrow frequency region (4-5 kHz) as a side branch of the vocal tract. Honda, et al. [2] replicated the acoustic experiment using male and female vocal-tract models to analyze the effect of the piriform fossa on vowel spectra with the result that the piriform fossa causes a regional effect above 4 kHz, while the female fossa generates the wider spectral change. However, because only one female subject was used, whether this is a general observation or not is uncertain. Takemoto, et al. [3] built 3D computational models of the vocal tract based on a male speaker and presented the existence of acoustic interaction between the left and right piriform fossa using finite-difference-time-domain (FDTD) method. Judging from the above studies, however, there is no study characterizing the piriform fossa of female speakers.

The purpose of the present study is to explore acoustic contributions of female piriform fossa to the spectra of Chinese vowel /a/ and /i/. 3D vocal-tract models of four subjects (three female and one male) were built based on MRI data obtained from the subjects. Acoustic experiments were conducted to analyze the effects of each piriform fossa on sound propagation in vocal tract, and further examined acoustic interaction between the left and right cavities of the piriform fossa using the water-filling technique.

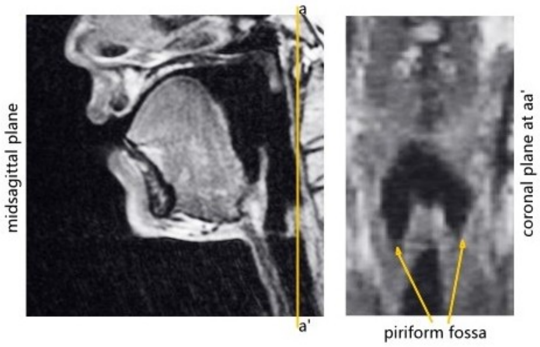

Figure 1: *MRI data of the vocal tract. The coronal plane shows the pear shape of piriform fossa.*

2. Materials and methods

Exploring the acoustic roles of piriform fossa requires morphological and acoustic measurement. MRI technique makes it possible to print realistic 3D models of the vocal tract so as to measure the volumes of the left and right cavities of the piriform fossa.

2.1. Subjects

The three female subjects and one male subject are labeled CR, LH, SC and WS in this experiment. All the subjects are native Chinese speakers of 25-27 years. No subject reported any history of speech or language disorders.

2.2. MRI Data

The MRI data include static scans for Chinese vowels /a/ and /i/ with teeth scans. The data were obtained by a Siemens Verio (3.0 Tesla) at the ATR Brain Activity Imaging Center (ATR-BAIC) in Japan. During data acquisition, subjects were asked to lay supine in the MRI scanner and keep the head still.

The MRI vowel data were obtained by a synchronized scan technique to visualize sagittal images with 2.0-mm slice thickness, 35 slices and 512×512 pixels.

The MRI teeth data were static 3D images in the sagittal plane with 1.0-mm slice thickness, 88 slices and 512×512 pixels. During the scan, the subjects kept firm tongue-teeth contact.

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

2.3. 3D vocal-tract model

The vocal-tract region from the glottis to the lips (without nasal cavity and teeth) was extracted manually from the transverse plane of the MRI vowel data slice by slice. And the lower and upper teeth were extracted from the sagittal plane of the MRI teeth data. The extracted data for the vocal-tract region and teeth were converted into the STL format using Materialise MIMICS. Then, the teeth data were superimposed onto the vowel data by using the dental pulp as landmarks [6]. After the teeth superimposition, the interdental space was extracted from the vowel data. In this way, the vocal-tract region was extracted completely. Finally, the extracted region was converted into the STL format using MIMICS.

The procedure for building the 3D vocal-tract models is similar to a previous study [7]. A wall of 3-mm thickness was built outside the vocal tract. The lower boundary of the vocal tract and the face region were separated from the whole structure. Then, the 3D vocal-tract models were printed with the accuracy of 0.05-mm by a 3D laser printer (Formlabs Form 1+) using light-sensitive liquid resin material.

2.4. Experiment on the piriform fossa

The water-filling method used by Dang and Honda [1] is used to examine the contributions of the piriform fossa in vowel spectra. The glottis of the 3D vocal-tract model was placed on a glottal coupler having a hole of 1.2-mm diameter, which is adhered on a horn driver unit (UNI-PEX P-800N, 8Ω).

The experiment was carried out in a sound-proofing room with the temperature of 25 ℃ . Continuous white noise generated from a PC-based signal generator was amplified by an audio power amplifier to drive the horn driver unit. The output sounds from the solid vocal-tract models were recorded by a condenser microphone (Behringer ECM8000), kept 0.1-m from the radiating end of the solid model, and a USB audio driver interface (Roland Duo Capture EX) connected to a PC at a sampling rate of 44.1 kHz. During the experiment, Plasticene was used to prevent the sound leakage at the joint.

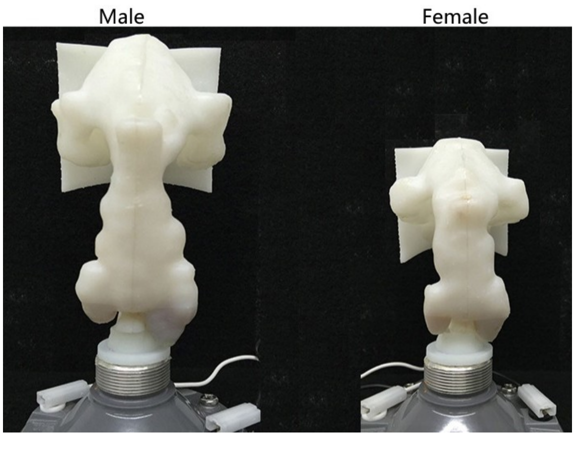

Figure 2: *Acoustic experiments of the vocal-tract models using water-filling method.*

Figure 2 shows the water-filling experiment for analyzing the piriform fossa in the condition of both cavities filled with colored water. To examine the acoustic effects of the piriform fossa and the interaction between the left and right cavities, acoustic recording was performed under following four conditions:

(1) Both cavities open (NF: natural condition)

(2) Left cavity filled (LF)

(3) Right cavity filled (RF)

(4) Both cavities filled (BF)

3. Results

3.1. Geometry of the left and right cavities of the piriform fossa

As shown in Figure 1, the piriform fossa is a pair of bilateral cavities behind the larynx above the esophageal entrance. The open end of the piriform fossa is determined by the arytenoid apex plane and the lateral wall of the pharynx [1]. Based on the above morphological features of the piriform fossa, difference in water-filled volume of each cavity for each subject was examined by calculating the volume below the plane at the level of the arytenoid apex using MIMICS.

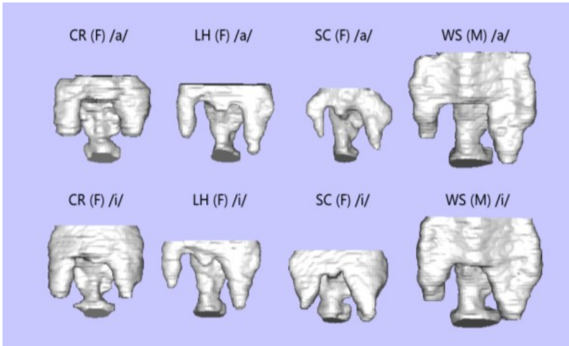

Figure 3: *Back views of the hypopharyngeal cavities of 3D vocal-tract models.*

Figure 3 shows back views of the hypopharyngeal cavities extracted from volumetric MRI data for all the subjects. The left-right asymmetry of the cavities is seen together with a slight to large difference of the shapes between /a/ and /i/. The volumes of the left and right cavities of the piriform fossa are indicated for vowel /a/ and /i/ in Figure 4. Except CR, the right cavity is larger than the left cavity. A marked difference in size between /a/ and /i/ is observed in SC.

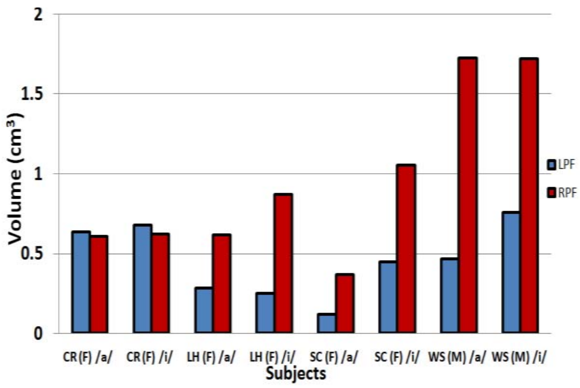

Figure 4: *Water-filled volumes of the left and right cavities of piriform fossa (LPF: left piriform fossa, RPF: right piriform fossa).*

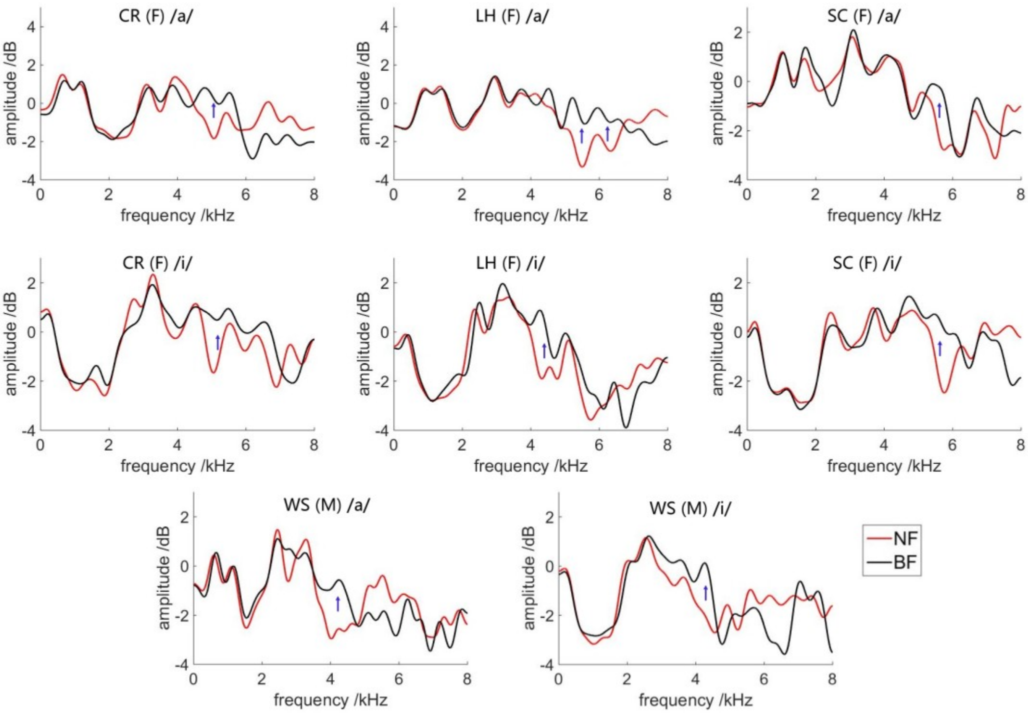

Figure 5: *Spectral envelopes from the vocal-tract models for vowels /a/ and /i/ in the conditions of no water (red line) and filled with water (black line) of the piriform fossa.*

3.2. Anti-resonances of the piriform fossa in female vowel spectra

The Imai's improved cepstral analysis method was used to estimate spectral envelopes of the recorded signals [8]. Figure 5 shows the spectral envelopes for vowel /a/ and /i/ obtained from the 3D vocal-tract models with the piriform fossa in the natural condition (no water filled) and the filled condition (filled with water). In the figure, the red lines represent the spectra in the natural condition, and the black lines represent the spectra in the filled condition. The differences between the two spectra are seen above 4 kHz, and the regions indicated by the arrows are thought to correspond to the anti-resonance(s) of the cavities. In those frequency regions, spectral dips (anti-resonances) in the natural condition are replaced by peaks when the cavities are filled with water. It is also observed that the anti-resonances tend to be coupled with poles as seen in the amplified regions in the higher frequency regions.

In the male subject (WS), the piriform fossa appears to generate dips in the frequency region of 4-5 kHz and peaks in the frequency region of 5-6 kHz. In the three female subjects (CR, LH and SC), the piriform fossa appears to generate dips at 4-6 kHz and peaks at 6-8 kHz. For SC in vowel /a/, the effects of the piriform fossa are not obvious due to the reduction of the cavity size. The spectrum for CR generates a single dip, suggesting acoustic interaction between the LPF and RPF, which show symmetrical shapes with the similar volumes.

3.3. Interaction between the left and right cavities

Acoustic interaction between the two cavities is thought to result in different dip frequencies across different conditions. Figure 6 shows the spectral envelopes for vowel /a/ and /i/ obtained from the 3D vocal-tract models in the conditions of no cavities filled (NF), left cavity filled (LF) and right cavity filled (RF) with water. The spectra of the above three conditions (NF, LF, and RF) are drawn by the red, blue and green lines, respectively. The figure shows large differences in these spectra, suggesting the acoustic interaction between the left and right cavities. Dips are seen in the frequency region of 4-6 KHz in the vowel spectra obtained under the three conditions.

Dip frequencies were measured in the conditions when the dips were identifiable in Figure 6, which are shown in Table 1. D_{NF}, D_{RF} and D_{LF} correspond to the dip frequencies in each condition. D_{NFH} is the second dip frequency in the higher frequencies in the natural condition. For CR, LH and WS in /a/, the D_{NF} is higher than D_{RF} and D_{LF}. For SC in /a/, the dips were unclear possibly because of the small volume of the cavities. There were also different cases showing that the D_{NF} is lower than D_{RF} and D_{LF}. For LH in /i/, the D_{NFH} is higher than D_{RF} but lower than D_{LF}. These facts suggest that the LPF and RPF interact with each other in generating complex dip patterns in vowel spectra.

Table 1.*Dip frequencies in vowel spectra in the three conditions.*

Subjects	D_{NF}	D_{RF}	D_{LF}	D_{NFH}
CR (F) /a/	5072	4372	4864	—
LH (F) /a/	5500	4344	5012	6329
SC (F) /a/	—	—	—	—
WS (M) /a/	4028	3996	3888	4434
CR (F) /i/	5053	5085	—	—
LH (F) /i/	4311	4674	4847	4716
SC (F) /i/	5763	5994	—	—

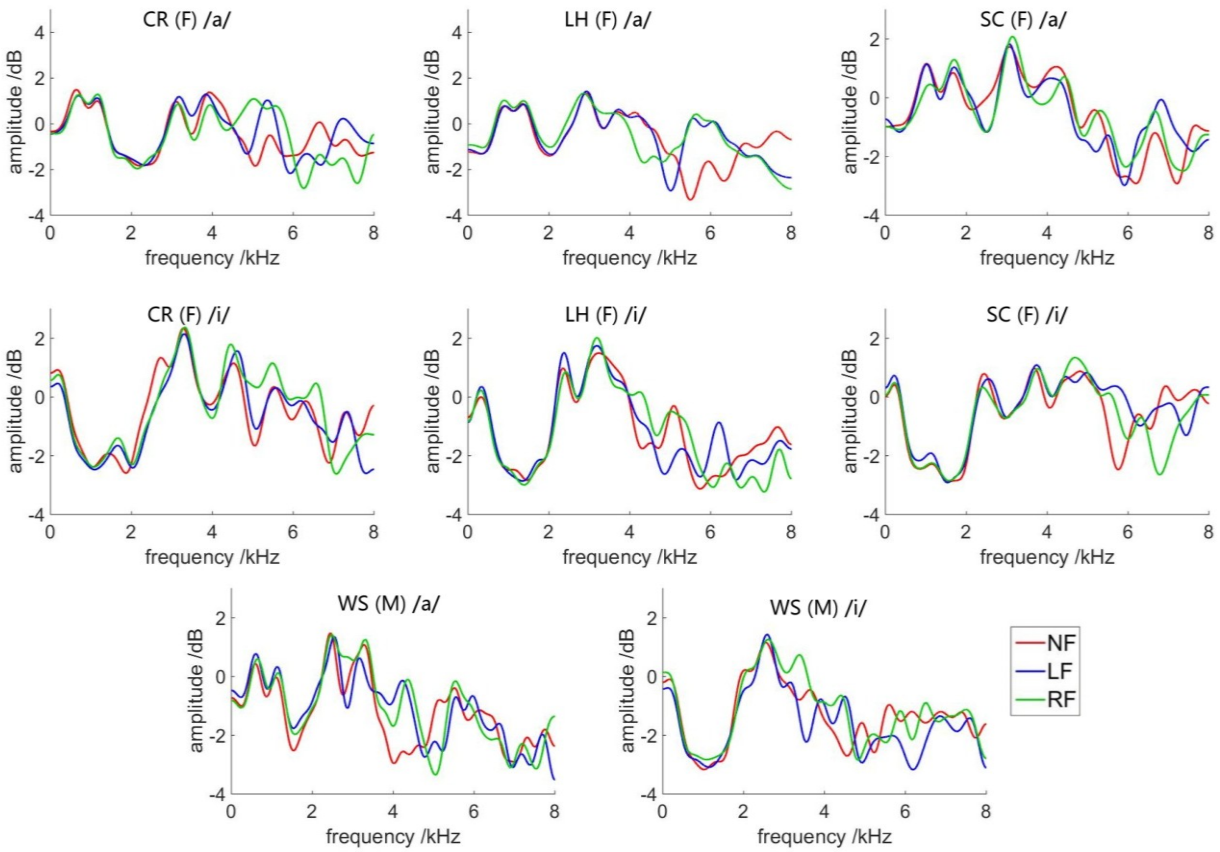

Figure 6: *Spectral envelopes from the vocal-tract models for vowels /a/ and /i/ in the conditions of no cavities filled (red line), left cavity filled (blue line) and right cavity filled (green line).*

4. Discussions and conclusion

In this study, the 3D solid vocal-tract models were built, and the individual characteristics of the piriform fossa in female speakers were examined for the geometry and acoustic effects. Acoustic analysis was made by comparing the differences in spectra of the acoustic signals recorded during water-filling experiments in the four conditions. In the morphological analysis, the asymmetry of the two cavities was a common finding. The volumes of the both cavities of the piriform fossa of the three female subjects were 0.7-1.6 cm^3, which is significantly smaller than the data for the male subject used as a reference. The results from the acoustic analysis indicate that the piriform fossa generates dips at 4-6 kHz and peaks at 6-8 kHz in the female vowel spectra (Figure 5). This result differs from the previous study using the similar method [2]: the dip frequency of the piriform fossa of a female subject was lower, and the acoustic effect was observed in the wider frequency range. This is partly because the volume of the cavities of the female subject was as large as that of the male subject in the previous study.

When the two cavities are asymmetric (RPF > LPF), only one obvious dip is seen in the data for SC and WS (Figure 5). Also, in another case of the asymmetrical cavities showing two dips, frequencies of the two dips do not correspond to those measured for each cavity separately (LH in Figure 6). Those variations suggest acoustic interaction between the left and right cavities of the piriform fossa. Takemoto et al. [3] report that the bilateral cavities give rise to a coupled oscillation, suggesting that the anti-resonance frequencies are determined by the geometry of the two cavities. In this study, the interaction between the LPF and RPF was analyzed by comparing the dip frequencies of the spectra in the three conditions for filling both cavities and each cavity (Table 1): The results suggest a complex manner of interaction with no clear tendencies in dip frequency shifts. The cases were seen that, when one cavity is filled, the dip frequency falls in agreement with [3], but, in other cases, the dip frequency rises. Thus, no definitive tendency was found from our experiment.

In summary, the piriform fossa of our three female subjects exhibited the acoustic effects in the higher frequency regions: dips are at 4-6 kHz and peaks are at 6-8 kHz. The interaction between the left and right cavities was observed to vary greatly between subjects and vowels. Thus, phenomena and mechanisms of the acoustic interaction must be detailed in the future studies.

5. Acknowledgements

The research is supported by the National Natural Science Foundation of China (No. 61471259; No. 61573254).

6. References

[1] J. Dang and K. Honda, "Acoustic characteristics of the piriform fossa in models and humans," *Journal of the Acoustical Society of America*, vol.101, no.1, pp.456-465, 1997.

[2] 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.

[3] H. Takemoto, S Adachi, P Mokhtari, et al. "Acoustic interaction between the right and left piriform fossae in generating spectral dips," *Journal of the Acoustical Society of America,* vol.134, no.4, pp.2955-2964, 2013.

[4] B. Delvaux and D. Howard. "A new method to explore the spectral impact of the piriform fossae on the singing voice: benchmarking using MRI-based 3D-printed vocal tracts," *Plos One*, vol.9, no.7, e102680, 2014.

[5] T. Kitamura, et al. "Individual variation of the hypopharyngeal cavities and its acoustic effects," *Acoustical Science & Technology*, vol.26, no.26, pp.16-26, 2005.

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

[7] T. Kitamura, H. Takemoto, S. Adachi, et al. "Transfer functions of solid vocal-tract models constructed from ATR MRI database of Japanese vowel production," *Acoustical science and technology,* vol.30, no.4, pp.288-296, 2009.

[8] 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.