Acoust. Sci. & Tech. **26**, 4 (2005)

TECHNICAL REPORT

An experimental study of acoustic characteristics of hypopharyngeal cavities using vocal tract solid models

Satoru Fujita* and Kiyoshi Honda†

ATR Human Information Science Laboratories,
2–2–2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619–0288 Japan

(*Received 3 June 2004, Accepted for publication 30 November 2004*)

Abstract: Solid models of the vocal tract with hypopharyngeal cavities were molded with a stereolithographic technique based on MRI data obtained from a male speaker during the production of Japanese vowels /a/ and /o/. A vowel synthesis experiment conducted with the models revealed a relatively good agreement in the second and third formants, as well as in anti-resonance at 4–5 kHz. The elimination of the models' piriform fossa resulted in the disappearance of the anti-resonance and shifts of the adjacent formants. The modification of the laryngeal cavity into a uniform tube caused spectral changes in the frequency range of 1.5–7.0 kHz. These acoustic effects of hypopharyngeal cavities were dependent on vocal tract shapes.

Keywords: Vocal tract model, MRI, Hypopharynx, Piriform fossa, Laryngeal cavity

PACS number: 43.70.−h [DOI: 10.1250/ast.26.353]

1. INTRODUCTION

The hypopharynx is a part of the vocal tract that consists of three small cavities: the laryngeal cavity and the bilateral cavities of the piriform fossa. The laryngeal cavity forms an "hourglass" shape with a constriction between the laryngeal ventricle and a gradual opening to the pharynx. The piriform fossa, named after the inverted shape of a "pear," is a pair of closed tubes behind the laryngeal cavity that gradually narrows toward the bottom. The acoustic effects of these cavities on spectral characteristics of vowels [1] and voice quality [2] have already been studied. Recently, their acoustic roles were reevaluated based on accurate morphological data obtained by magnetic resonance imaging (MRI) [3]. Kitamura *et al.* [4] reported that the hypopharyngeal shape in static MRI data is relatively constant across vowels and also examined the effects of its individual differences on vowel spectra. Takemoto *et al.* [5] also observed the geometrical stability of the laryngeal cavity based on dynamic MRI data obtained during vowel sequences and described its acoustic effects on vowel spectra. As for the piriform fossa, Dang and Honda [6] reported that these cavities cause troughs in vowel spectra at around 5 kHz, based on acoustic experiments using solid models of the hypopharynx and human vocal tracts.

The purpose of this study is to directly measure the acoustic effects of the hypopharyngeal cavities using realistic models of the entire vocal tract without relying on numerical calculations. Solid models were formed according to volumetric MRI data to accurately represent three-dimensional (3D) vocal tract shapes.

2. METHOD

2.1. Acquisition of MRI Data

MRI data were acquired from a Japanese male subject during the production of the five Japanese vowels (/a/, /e/, /i/, /o/, /u/) using a clinical scanner (Shimadzu, SMT-100GUX 1.0 T) installed at the Takanohara Chuo Hospital in Nara City. The imaging sequence was a sagittal Spin Echo series 3.5 mm thick, a 256 × 256 mm field of view, and a 256 × 256 pixel size. Then, the data sets were interpolated using a bilinear algorithm to form 3D data with cubic voxels of 0.5 × 0.5 × 1.75 mm.

The subject was positioned supine in the MRI unit and instructed to steadily repeat phonation, to breath gently through his mouth, and keep the soft palate elevated during scanning. The subject was not instructed to maintain voice fundamental frequency during vowel production. To visualize dental images, the subject's upper and lower teeth were covered by dental imaging plates about 1-mm thick with an intermediate layer of liquid imaging agent [7].

*e-mail: sfujita@jaist.ac.jp
Present address: Japan Advanced Institute of Science and Technology, 1–1, Asahidai, Nomi, Ishikawa, 923–1292 Japan
†e-mail: honda@atr.jp

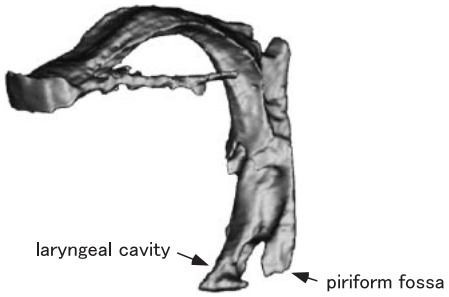

Fig. 1 Extracted vocal tract for /e/.

Fig. 2 Vocal tract model /o/ attached to a horn driver.

2.2. Extraction of Vocal Tract Shape

Vocal tract shapes were extracted using a commercial image processing tool (Materialise, MIMICS). The procedure included thresholding, manual segmentation of borders, vocal tract extraction with a region growing method, and hand editing of the detail boundaries. The entire region of the vocal tract from the vocal folds to the lips was extracted. The nasopharyngeal port was represented as closed in the models, even when a small opening was observed as is often the case for /a/. Therefore, the data included neither facial surfaces nor nasal cavities. Figure 1 shows the extracted vocal tract shape for the vowel /e/.

2.3. Modeling Solid Vocal Tracts

Each extracted vocal tract was converted into hollowed 3D data with a surrounding 3-mm thick wall using a commercial CAD tool (Materialise, MAGICS). Some of the models were then smoothed and divided into two pieces so that the internal tube shape could be modified. Finally, solid models of the vocal tract were built using a stereolithography system (CMET, SOUP 600GS). The solid models were formed by layers of liquid photosensitive resin, each 0.1 mm thick. The orientation of the models was carefully aligned to avoid the unexpected formation of islands within the vocal tract during the lithographic process.

There were two types of model made for vowels the /a/, /e/, and /o/: one with a smoothed wall and another with an unsmoothed wall. The models with a smoothed wall were made in two pieces to uncover the hypopharyngeal cavities for modification. The other vowels, /i/ and /u/, were omitted because their vocal tract shapes were unnatural due to extreme articulation with the dental imaging plate.

2.4. Synthesis Using Solid Vocal Tract Models

Vowel synthesis was performed using a horn driver unit (YL-551) as a source generator to excite the vocal tract models. To simulate glottal impedance, the horn driver's throat was filled with a sponge, and a thin metal plate with multiple holes 0.5 mm in diameter was placed at the throat opening. The model was attached to the driver unit using acrylic adhesive tape with a 3-mm hole. No baffle flange was attached to the models in the experiment.

The voice source signal was made using a Rosenberg wave generator program on a personal computer. The voice source parameters (*i.e.* F_0, OQ, and the fluctuation rate) were adjusted interactively while listening to synthesized sounds to obtain a natural vowel quality. This source signal was fed to the driver unit through a D/A converter unit (M-Audio, DUO) and an audio amplifier (Audio-technica, AT-MA50). The output sounds from the models were recorded in a soundproof room using a microphone system (B&K 4912 and B&K 5939) and digitized with an A/D converter (Edirol, UA-5) for storage on a computer. The sampling rate for the D/A and A/D conversions was 48 kHz. Figure 2 shows the vocal tract model attached to the horn driver unit.

2.5. Spectral Analysis

Vowel sounds from the subject in an upright position were also recorded in a soundproof room. A cepstrum analysis was performed on the synthesized and recorded vowels using a hamming window with a 1.0-s length. The spectral envelopes and formant frequency peaks below 5 kHz were compared across the sounds.

3. RESULTS OF VOWEL SYNTHESIS

3.1. Synthesis of Vowel /a/

Figure 3 shows the spectral envelopes of synthesized and natural vowels /a/. In the synthesized vowel, F_1 and F_3 were lower, while F_4 was higher than in the natural vowel. The differences in formant frequencies were:

F_1: -28%, F_2: $+6\%$, F_3: -6%, F_4: $+1\%$, F_5: $+1\%$

Differences in formant frequency peaks were observed, especially at F_1, F_2 and F_3, possibly due to sustained vowel production in a supine body posture during the MRI experiment and the slightly open condition of the velopharyngral port. In evidence, by making a hole with a 2 mm diameter on the wall to simulate vocal tract bifurcation at

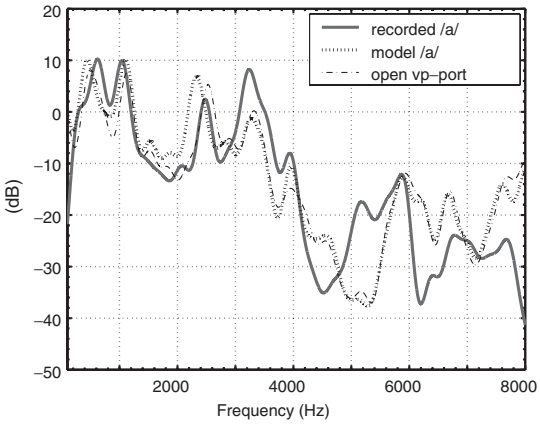

Fig. 3 Spectra of recorded and synthesized /a/.

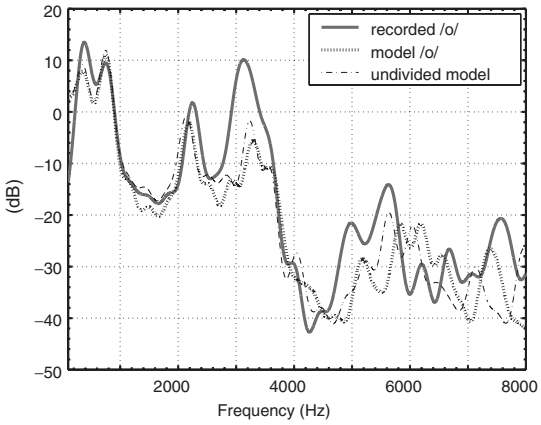

Fig. 4 Spectra of recorded and synthesized vowel /o/ with a spectrum in open velopharyngeal-port condition.

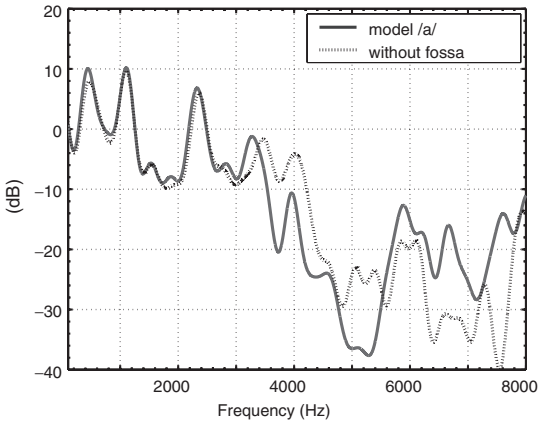

Fig. 5 Spectral envelopes of /a/ with and without piriform fossa.

the velopharyngeal port, F_1 and F_3 became closer to those of the natural vowel, as shown in Fig. 3. The solid wall of the model can be regarded as an additional factor for decreasing F_1, which was not examined in this work.

The troughs due to the piriform fossa were located in the higher frequencies in the model than in the natural vowel, which may be accounted for by the shorter length of the model's piriform fossa due to the smoothing procedure in modeling the wall.

3.2. Synthesized Vowel /o/

Figure 4 shows spectral envelopes of vowel /o/ for recorded sound and synthesized ones from smoothed and unsmoothed models. In the synthesized vowel from the smoothed model, F_1 and F_3 were lower, while F_4 was higher, than the natural vowel. The differences in formant frequencies in the same model were:

F_1: −8%, F_2: 0%, F_3: −5%, F_4: +4%, F_5: +5%

In this vowel, the spectral envelopes resemble each other in the frequency range below 5 kHz.

The spectral shapes were similar between the two models in the frequency range below 5 kHz, but the influences of the smoothing were observed at frequencies higher than 5 kHz.

3.3. Synthesized Vowel /e/

The synthesized sound from model /e/ lacked natural vowel quality, and its formant peaks differed significantly from the natural vowel /e/. Since it was assumed that sustained phonation in a supine position affected the articulation of /e/ more than other vowels, this model was extruded from the following experiment.

4. EFFECTS OF MODIFICATIONS ON THE CAVITIES

Modifications were applied to the piriform fossa and laryngeal cavity of the solid models for vowels /a/ and /o/, and acoustic effects of these cavities were examined by comparing synthesized sounds before and after each modification.

4.1. Effects of the Piriform Fossa

The bilateral cavities of the models' piriform fossa were filled with plasticine to observe its influence on sound spectra.

Figure 5 shows the spectral envelopes of the synthesized sounds from the original and modified models for /a/. The model without piriform fossa cavities demonstrated an elevation of the spectral envelope in the frequency region of 3.3–5.5 kHz. Spectral changes in the lower frequency range (below 3 kHz) were relatively small (about 3 dB), while those in the higher frequency range (above 3 kHz) were significant (about 13 dB). In the presence of the bilateral cavities, all the spectral peaks below 5 kHz shifted to within 6% of the lower frequency range, and the trough at around 5 kHz became distinctive.

Figure 6 shows the spectral envelopes of the synthesized sounds from the original and modified models for /o/. In the model without the piriform fossa, the spectral

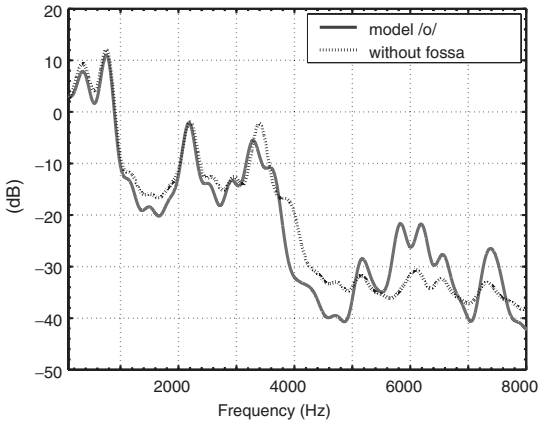

Fig. 6 Spectral envelopes of /o/ with and without piriform fossa.

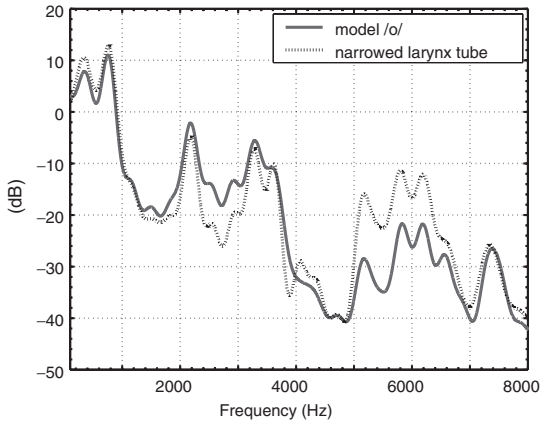

Fig. 8 Spectral envelopes of /o/ with normal and narrowed laryngeal tube.

envelope in the region of 3.3–5.0 kHz was raised and the trough at about 4.7 kHz was diminished, although the differences were smaller than in /a/. In this vowel, the directions of the shifts of formant peaks in the lower frequency region were not always uniform because F_2 changes in the opposite direction of the shift of other formants.

4.2. Effects of Modification on the Laryngeal Cavity

For its influence to be clearly observed, the laryngeal cavity of the models was narrowed with plasticine to form a uniform tube 18 mm long with a diameter of about 3-mm.

Figure 7 shows the spectral envelopes of the synthesized sounds from the original and modified models for /a/. By narrowing the laryngeal tube, the spectral envelope indicated a fall in the frequency region of 1.5–3.3 kHz and a rise in the higher frequency region of 5–7.5 kHz. Frequency changes in the lower three formants were small (within 3%), and F_3 was weakened.

Figure 8 shows the spectral envelopes of the synthesized sounds from the original and modified models for /o/. The spectral changes were essentially similar to those

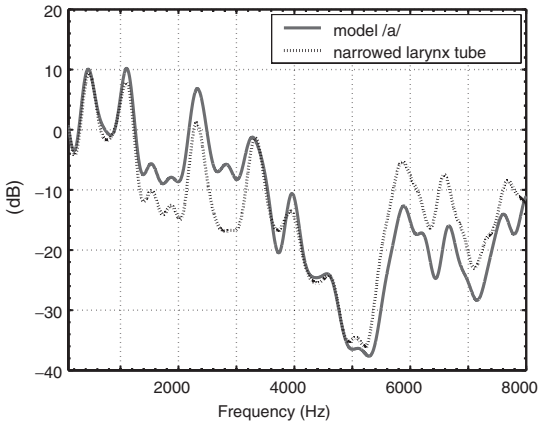

Fig. 7 Spectral envelopes of /a/ with normal and narrowed laryngeal tubes.

of the model for /a/, though in the narrowed cavity configuration, the spectral envelope showed a drop in the range of 1.5–3.2 kHz and a rise in the range of 5–6.5 kHz.

5. DISCUSSION

The acoustic effects of hypopharyngeal cavities have been examined by others using numerical calculations of the vocal tract. Fant [1] described systematic effects of these cavities based on a simplified vocal tract model: narrowing the laryngeal tube increases the density of spectral peaks in the frequency region of 3–5 kHz (at about F_4) to rapidly decrease the spectral slope above 5 kHz; adding the piriform fossa further sharpens the cutoff at 5 kHz due to anti-resonance just above this frequency. The effects of the piriform fossa were also examined in a few recent studies. Fant and Båvegård [8] as well as Dang and Honda [6] note that the presence of the fossa decreases the lower formants to various degrees depending on vocalic configurations. The effects of the piriform fossa obtained from the present study essentially agree with the previous studies. A few differences from the previous studies can be noted, however, as seen in anti-resonance effects over the broad frequencies down to 3 kHz, shifts of the nearby peaks spread apart from the anti-resonance center, and lack of large shift of the lower formants. Causal factors of these discrepancies need to be investigated in the future.

Sundberg [2] reported that lowering the larynx and narrowing the laryngeal cavity relative to the pharynx contributes a rise to vowel spectra in the region of 2.5–5 kHz, as often seen in singing voices. Takemoto *et al.* [5] reported that the bilateral pouches of the laryngeal ventricle make the cavity resemble a Helmholtz resonator, which causes a resonance peak at 3–3.5 kHz. The results from the current experiment on the laryngeal cavity in the vowels /a/ and /o/ are qualitatively different from the above studies. By uniformly narrowing the laryngeal cavity, the bilateral pouches of the laryngeal ventricle were eliminated

from the cavity. This modification reduced the original resonance of the laryngeal cavity at 1.5–3 kHz and created a new resonance above 5 kHz, with no significant changes in the lower formant frequencies. The acoustic change observed in the broad frequency range of 1.5–3 kHz differs from the effect expected from a Helmholtz resonator (i.e., a resonance peak at 3–3.5 kHz). Further, the spectral rise in the range above 5 kHz in the narrowed cavity configuration implies a simple addition of a $1/4$ wavelength resonance of a closed tube about 17 mm long, which nearly agrees with the actual length of the tube (18 mm). Therefore, the modification from the natural to narrowed laryngeal cavity configuration was reflected only by local changes in spectral level and resulted no obvious frequency shifts in the higher formants bellow 5 kHz. Further acoustic and aerodynamic studies may be necessary to account for the detailed acoustic manifestation of the laryngeal cavity.

6. SUMMARY

Acoustic effects of the hypopharyngeal cavities, *i.e.* the piriform fossa and laryngeal cavity, were examined by modifying the cavities in the vocal tract models. Major effects of the piriform fossa were seen in the frequency range of 3.3–5.0 kHz as an anti-resonance caused by the side branches to the vocal tract. The effects of the laryngeal cavity were found in the spectral amplitude near 3 kHz and over 5 kHz. These effects seem to contribute to forming realistic spectral envelopes of vowels that were not represented by traditional vocal tract models.

ACKNOWLEDGMENTS

This research was conducted as part of ‘Research on Human Communication’ with funding from the National Institute of Information and Communications Technology.

REFERENCES

[1] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).

[2] J. Sundberg, *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, Ill., 1987).

[3] K. Honda, H. Takemoto, T. Kitamura, S. Fujita and S. Takano, “Exploring human speech production mechanisms by MRI,” *IEICE Trans. Inf. Syst.*, **E87-D**, 1050–1058 (2004).

[4] T. Kitamura, K. Honda and H. Takemoto, “Individual variation of the hypopharyngeal cavities and its acoustic effects,” *Acoust. Sci. & Tech.*, **26**, 16–26 (2005).

[5] H. Takemoto, K. Honda, S. Masaki, Y. Shimada and I. Fujimoto, “Measurement of temporal changes in vocal tract area function during a continuous vowel sequence using a 3D cine-MRI technique,” *Proc. 6th Int. Semin. Speech Production, Sydney*, pp. 284–289 (2003).

[6] J. Dang and K. Honda, “Acoustic characteristics of the piriform fossa in modesl and humans,” *J. Acoust. Soc. Am.*, **101**, 456–465 (1997).

[7] M. Wakumoto, S, Masaki, J. Dang, K. Honda, Y. Shimada, I. Fujimoto and Y. Nakamura, “Visualization of dental crown shape in an MRI-based speech production study,” *Int. J. Oral Maxillofacial Surgery*, **26**, 189–190 (1997).

[8] G. Fant and M. Båvegård, “Parametric model of VT area functions: Vowels and consonants,” *KTH, TMH-QPSR*, **38**(1), 1–20 (1997).