

ACOUSTICAL LETTER

Study of acoustic characteristics of vocal tract with nasal cavity during phonation of Japanese /a/

Hiroki Matsuzaki and Kunitoshi Motoki*

Department of Electronics and Information Engineering, Faculty of Engineering, Hokkai-Gakuen University

(*Received 16 June 2006, Accepted for publication 21 September 2006*)

Keywords: Acoustic analysis, Three-dimensional FEM, Vowel, Nasal cavity, Vocal tract, Wall impedance

PACS number: 43.70.Bk [doi:10.1250/ast.28.124]

1. Introduction

MRI data on the vocal tract during the production of Japanese vowels indicate coupling between the oral cavity and the nasal cavity. A large number of studies have been carried out on the nasalized vowels on the basis of a one-dimensional speech production model. Acoustic analysis of the three-dimensional nasal tract has also been performed by the finite element method (FEM) [1,2]. Moreover, acoustic coupling between the oral and nasal cavities in the radiation space has been studied using simplified finite element models [3]. However, as for the case with three-dimensional radiation from the lips and nostrils, the acoustic characteristics of vocal-tract shape based on MRI data are not clear. Simulations based on a detailed vocal-tract shape with proper boundary conditions are important for studying the acoustic characteristics of speech, especially at high frequencies where individual information possibly exists.

In this paper, using vowel MRI data of the vocal tract with the nasal cavity during phonation of the Japanese /a/, we examine the effects of the nasal cavity on the acoustic characteristics of the speech production system. The acoustic analysis of three-dimensional geometrical vocal-tract models is performed by the FEM. Two types of models are composed on the basis of MRI data. One is a model with a three-dimensional nasal cavity, and the other, for the purpose of comparison, is without a nasal cavity. The nasal cavity is also coupled to the oral cavity through a space between the lips and the nostrils in a three-dimensional volume of radiation. The effects of the wall impedance of the vocal tract are also examined.

2. Geometrical vocal-tract models

Two types of three-dimensional geometrical vocal-tract models were designed. The first one was the model with the nasal cavity. The second one was the model without the nasal cavity, which was made by removing the nasal cavity from the model with the nasal cavity.

ATR Human Information Science Laboratories provided the Japanese vowel MRI data on which the dental shape data were superimposed. Among these MRI data, we used a data set taken during the phonation of /a/ to shape the three-dimensional geometrical vocal-tract models because the data set of /a/ clearly shows the coupling between the oral and nasal cavities. An MR image of the mid-sagittal plane of the data set is shown in Fig. 1. The subject of these data has a history of an operation on his paranasal sinuses. Therefore, the nasal cavity is deformed from the normal shape, which was observed as a coupling between the inferior nasal meatus and the sinus maxillaris. Note that simulation results in the following sections have the possibility of reflecting the effects of the shape of an operated nasal cavity. However, the detailed acoustic field in the vicinity of the branch and the effects of external coupling between the nostrils and the mouth can be investigated with an operated nasal cavity.

A three-dimensional volume of radiation [4] with a radius of 4 cm, which is spherical in shape, was attached to the face covering the lips and nostrils. A specific acoustic impedance of spherical waves was used as a boundary condition on the round surface of the three-dimensional volume of radiation. The boundary condition in the radiation space is an important parameter for the simulation. The use of the characteristic impedance of air on the surface of radiation space is a simple method for the simulation. However, it requires a relatively large radiation space that leads to a marked increase in the number of finite elements. As we have confirmed from the measurement results that the wavefront in the vicinity of the lips is well approximated by spherical waves [5], the acoustic impedance density of the spherical waves is specified on the surface of the radiation space. With this radiation model, the size of the radiation space can be effectively reduced compared with the use of the characteristic impedance. The surface meshes of the models are shown in Fig. 2. A boundary between the oral and nasal cavities was assumed on the horizontal plane at the narrowest position between the velum palatinum and the pharynx wall on the mid-sagittal plane. The space above this boundary plane was removed for the model without the nasal cavity. The nostrils were also closed with the surface mesh being based on the nodes surrounding the nostrils.

3. FEM simulation

The three-dimensional FEM was applied to the wave equation in a steady state to obtain a velocity potential. The applied FEM was used to simulate acoustic wave propagation in the geometrical vocal-tract models. The glottis, as the driving surface, was driven with a sine wave. To examine the effect of the boundary condition of the wall, two types of boundary conditions were assumed on the walls of the vocal tract and face. A rigid wall condition was assumed in the first simulation and a soft wall condition in the second simulation.

*e-mail: {matsu, motoki}@eli.hokkai-s-u.ac.jp

Fig. 1 MR image of mid-sagittal plane during phonation of /a/.

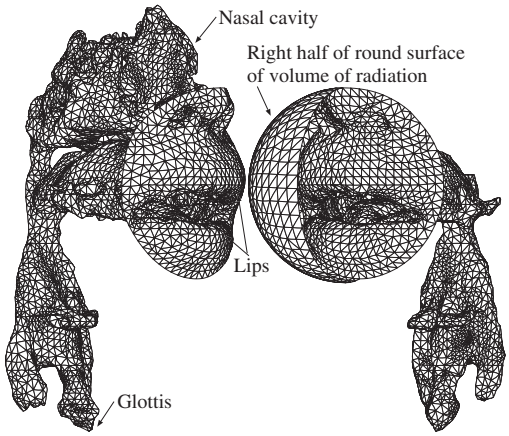


Fig. 2 Surface meshes. Left: model with nasal cavity. Right: model without nasal cavity.

The wall impedance proposed by Kamiyama *et al.* [6] was assumed as the boundary condition of the soft wall with a thickness of 2 cm. Sound pressure and particle velocity were computed from the velocity potential. The simulation was carried out in a driving frequency range of 100 Hz to 10 kHz at intervals of 10 Hz and then 1 Hz in the vicinity of peaks and zeros.

4. Results and discussion

We obtained vocal-tract transfer functions and active sound intensity using the sound pressure and particle velocity. The transfer function $H(\omega)$ is defined as

$$H(\omega) = K\left|\frac{\sqrt{W_{\mathrm{rad}}}}{u_{\mathrm{g}}}\right|, \tag{1}$$

where W_{rad} is a radiation power equivalent to the total active intensities on the surface of the three-dimensional volume of radiation, u_{g} is the source volume velocity, and K is a constant for $H(\omega)$ to be dimensionless.

4.1. Rigid wall condition

4.1.1. Vocal-tract transfer functions

The vocal-tract transfer functions for the rigid wall condition are shown in Fig. 3. In the transfer function of the model with the nasal cavity, an additional peak appears at

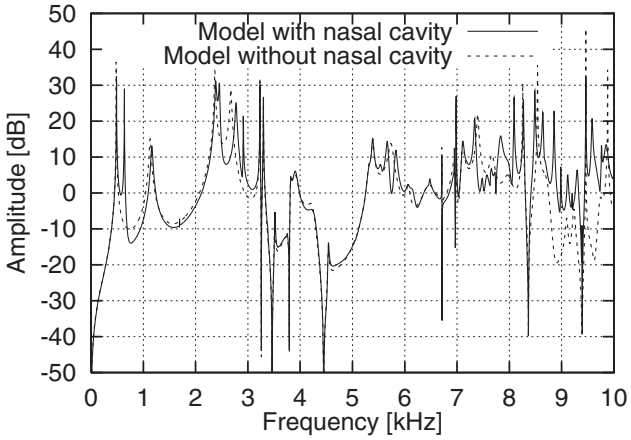


Fig. 3 Vocal-tract transfer functions of models with (solid line) and without (dotted line) nasal cavity under rigid wall condition.

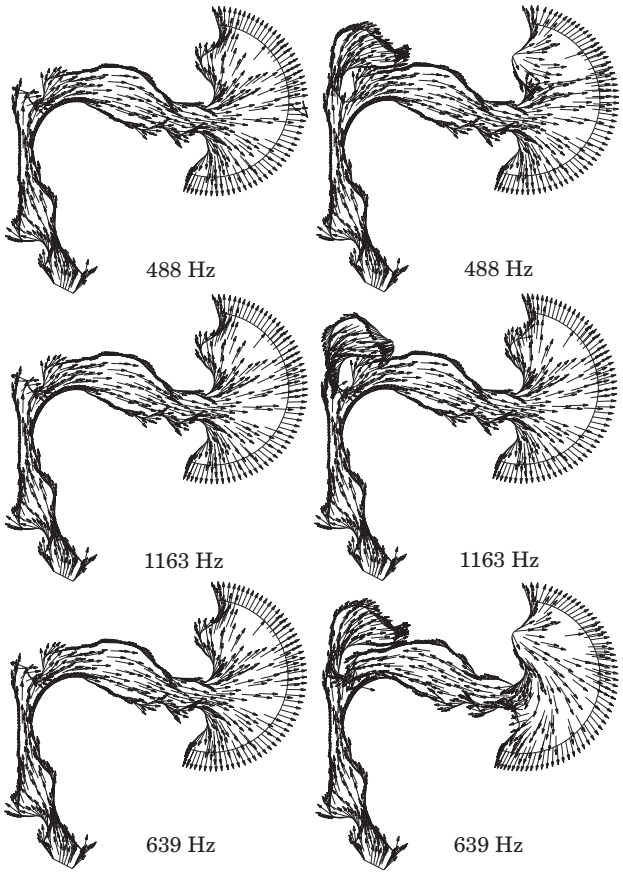
639 Hz between the first formant (F_1) and the second formant (F_2). An additional zero also appears at 761 Hz. These must be a pole–zero pair, which is one of the acoustic characteristics of the nasalized vowel. Additional peaks at 2,455 Hz and 2,912 Hz also appear in close vicinity to the third formant (F_3) and fourth formant (F_4), respectively. Note that the transfer function of the model with the nasal cavity agrees with that of the model without the nasal cavity in the range from 3 kHz to 7 kHz. The implication is that the spectrum envelope is mainly determined by the main vocal tract only, and the nasal cavity has little influence in this frequency range. Although we cannot make a generalization about these results because the nasal cavity in the model has sections which have been operated on, the results indicate that the nasal cavity coupled with the oral cavity contributes to the appearance of peaks in the transfer functions below 3 kHz.

Table 1 shows F_1 to F_4 frequencies and their percentages of shifts relative to the formant frequencies of the model without the nasal cavity. The coupling of the nasal cavity to the oral cavity makes the F_1 to F_4 frequencies higher, although the percentage of shifts is low.

Kitamura *et al.* [7] reported that the F_1 to F_4 frequencies of the average spectral envelope of speech data, which are recorded from the same subject, are 563, 1,047, 2,578 and 3,016 Hz, respectively. The percentages of the shifts of the F_1 to F_4 frequencies of the model with the nasal cavity relative to the formant frequencies of speech data are about −13.3, 11.1, −7.4 and −8.1%, respectively. There are large differences between the formant frequencies of the model with the nasal cavity and those of the speech data. The lossless condition potentially causes the differences, especially the wall boun-

Table 1 From F_1 to F_4 frequencies [Hz] and percentages of the shifts.

	F_1	F_2	F_3	F_4
Model with nasal cavity	488	1,163	2,386	2,773
Model without nasal cavity	482	1,132	2,369	2,675
Percentage of shifts	1.2	2.7	0.7	3.7



(a) Model without nasal cavity (b) Model with nasal cavity

Fig. 4 Distributions of active sound intensity vectors on mid-sagittal plane under rigid wall condition.

dary condition, as it is well known from one-dimensional modeling that the yielding wall has an effect of upward shifts of lower formant frequencies.

4.1.2. Active sound intensity

Distributions of the active sound intensity vectors on the mid-sagittal plane are shown in Fig. 4. The frequencies 488 and 1,163 Hz correspond to the F_1 and F_2 frequencies of the model with the nasal cavity. 639 Hz is the frequency of the peak which appeared only in the transfer function of the model with the nasal cavity. As shown in Fig. 4(a), a uniform flow of the active intensity vectors can be observed, and there is no appreciable difference dependent on frequencies for the model without the nasal cavity. On the other hand, as shown in Fig. 4(b), different aspects of the sound energy flow are obtained depending on the difference in frequencies for the model with the nasal cavity. At F_1, a sound energy flow from the glottis bifurcates into the oral and nasal cavities. At F_2, the sound energy flow from the glottis meets the energy flow from the nasal cavity at the top of the pharynx, and starts to flow into the oral cavity. At 639 Hz, a sound energy flow radiates from the nostril and drains into the oral cavity. These curious distributions of energy flow may have possibly emerged from the assumption of the lossless condition in the geometrical vocal-tract model. The effects of the wall boundary condition are examined in the next section.

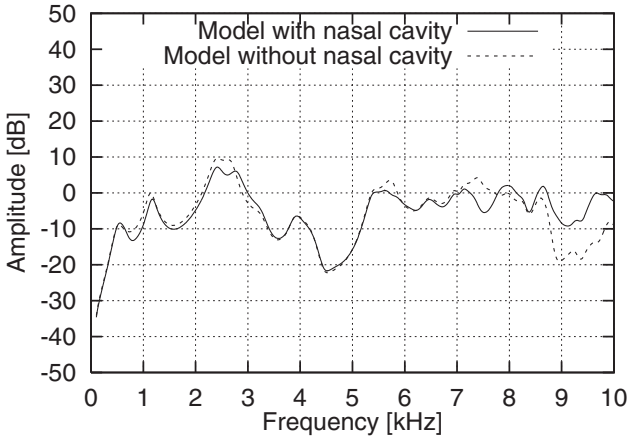


Fig. 5 Vocal-tract transfer functions of models with (solid line) and without (dotted line) nasal cavity. The whole vocal-tract wall was assumed to be soft.

Table 2 From F_1 to F_4 frequencies [Hz] and percentages of shifts.

	F_1	F_2	F_3	F_4
Model with nasal cavity	554	1,173	2,417	2,758
Model without nasal cavity	535	1,140	2,418	2,625
Percentage of shifts	3.6	2.9	−0.0	5.1

4.2. Soft wall condition

4.2.1. Vocal-tract transfer function

The vocal-tract transfer functions for the soft wall condition are shown in Fig. 5. The sharp peaks and zeros shown in Fig. 3 are not observed and the bandwidths are enlarged because of wall loss [8]. The additional peaks at 639, 2,455 and 2,912 Hz in Fig. 3 have disappeared for the model with the nasal cavity. The broad bandwidth of each peak possibly merges these additional peaks with the F_1, F_3 and F_4, respectively.

Table 2 shows the F_1 to F_4 frequencies and their percentages of shifts relative to the formant frequencies of the model without the nasal cavity. Compared with the formant frequencies for the rigid wall condition, the upward shifts of F_1, F_2 and F_3 can be confirmed. The percentages of shifts of the F_1 to F_4 frequencies of the model with the nasal cavity relative to the formant frequencies of speech data are about −1.6, 12.0, −6.2 and −8.6%, respectively. The difference in the formant frequencies between the model with the nasal cavity and speech data decreased greatly at F_1, and slightly at F_3, while increasing slightly at F_2 and F_4, compared with those for the rigid wall condition described in Sect. 4.1.1. These discrepancies may be reduced by adjusting the wall impedance since there are still many immature points left in the simulation; for example, the wall impedance distributed homogeneously and uniformly on the wall of the models.

4.2.2. Active sound intensity

Distributions of the active sound intensity vectors on the

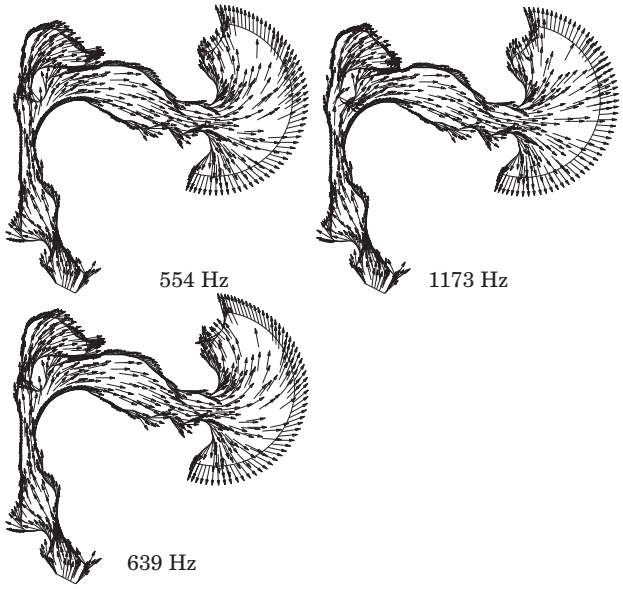


Fig. 6 Distributions of active sound intensity vectors on mid-sagittal plane under soft wall condition.

mid-sagittal plane of the model with the nasal cavity are shown in Fig. 6. The frequencies 554 and 1,173 Hz correspond to the F_1 and F_2 frequencies. 639 Hz is the frequency of the peak which appeared only in the transfer function of the model with the nasal cavity under the rigid wall condition. Compared with Fig. 4(b), the radiation from the nostrils decreased greatly at F_1. The sound energy flow from the nasal cavity at F_2 does not drain into the oral cavity as observed under the rigid wall condition. Similarly, the sound energy flow from the three-dimensional volume of radiation does not drain into the oral cavity at 639 Hz. This result suggests that the boundary condition in the simulation based on the three-dimensional shape greatly affects the simulation results.

5. Conclusion

The transfer functions and the active sound intensities of a vocal-tract model with and without a nasal cavity were computed by the three-dimensional FEM on the basis of MRI data of the Japanese /a/. An oral cavity was also coupled with the nasal cavity in a three-dimensional volume of radiation. The effects of wall impedance were also examined.

The coupling of the nasal cavity to the oral cavity indicated the following aspects. Additional peaks appeared below 3 kHz for the lossless condition; however, they disappeared in the simulation under the soft wall condition. Sound energy circulation did not occur in the simulation under the soft wall condition. As for the effects of the wall boundary condition on the spectral envelope, the upward shifts of lower formant frequencies were confirmed on the basis of three-dimensional simulations. However, the disagreements of the formant frequencies between the simulation and real speech should be further investigated by adjusting the wall boundary condition to a more realistic one. As the presented results are only for one subject, simulations on different subjects are also required to draw general conclusions.

Acknowledgments

This study has been carried out using "ATR vowel MRI data". Part of this work has been supported by a research project of the High-Tech Research Center, Hokkai-Gakuen University.

References

[1] N. Takahashi, A. Ishida, T. Nakai and H. Suzuki, "Coupling of the nasal and the oral cavities based on acoustic tube model of vocal tract," *IEICE Tech. Rep.*, SP95-11, pp. 7–14 (1995).
[2] H. Suzuki, T. Nakai and H. Sakakibara, "Analysis of acoustic properties of the nasal tract using 3-D FEM," *Proc. ICSLP 96*, pp. 1285–1288 (1996).
[3] M. Kouchi, N. Takahashi, T. Nakai and H. Suzuki, "A basic study about the nasal sound production taking account of the acoustic combination between oral and nasal cavity in the radiation space," *IEICE Tech. Rep.*, EA96-13, pp. 9–12 (1996).
[4] H. Matsuzaki, N. Miki and Y. Ogawa, "FEM analysis of sound wave propagation in the vocal tract with 3-D radiational model," *J. Acoust. Soc. Jpn.* (*E*), **17**, 163–166 (1996).
[5] K. Motoki, P. Badin and N. Miki, "Measurement of acoustic impedance density distribution in the near field of the labial horn," *Proc. ICSLP 94*, pp. 607–610 (1994).
[6] N. Kamiyama, N. Miki and N. Nagai, "Study of the vocal tract wall impedance using viscoelastic model of the wall," *IEICE Trans. Fundam.*, **J75-A**, 1649–1656 (1992).
[7] T. Kitamura, H. Takemoto and K. Honda, "Acoustic analysis of the vocal tract by FEM with voxel meshing," *IEICE Tech. Rep.*, SP2004-78, pp. 41–46 (2004).
[8] J. Suzuki, "Discussions on vocal tract wall impedance," *J. Acoust. Soc. Jpn.* (*J*), **34**, 149–156 (1978).