Numerical Simulation of Acoustic Characteristics of Vocal-tract Model with 3-D Radiation and Wall Impedance

Hiroki Matsuzaki* and Kunitoshi Motoki†
* Hokkaido Institute of Technology, Sapporo, Japan
E-mail: hiroki@hit.ac.jp Tel/Fax: +81-11-688-2327
† Hokkai-Gakuen University, Sapporo, Japan
E-mail: motoki@eli.hokkai-s-u.ac.jp Tel: +81-11-841-1161 Fax: +81-11-551-2951

***Abstract*—Numerical simulations of acoustic characteristics of vocal-tract model are performed with three-dimensional radiation and wall loss. A method to calculate transfer characteristics suitable for three-dimensional vocal-tract models is proposed. And the simulation results show that the simulation under the lossless condition potentially causes unrealistic result, and wall loss is needed for the three-dimensional simulation.**

I. Introduction

Numerical simulations of acoustic characteristics of vocal-tract models have been performed based on MRI (Magnetic Resonance Imaging) data using FEM (Finite Element Method) [1], [2]. In the three-dimensional simulation, following two problems arise: First, in the three-dimensional model of the vocal-tract, it is difficult to clearly define the output point of acoustic signal of the vocal-tract as the contours of the lips do not exist on a single plane, and the acoustic field in the vicinity of the lips is influenced by the radiation from the nostrils. Second, the vocal-tract has a variety of wall conditions, such as teeth, tongue, etc. Each part has a different wall impedance, and influences the sound wave propagation in the vocal-tract. It is obviously questionable to evaluate the transfer characteristics under a lossless condition.

In this paper, a method to evaluate the resonance characteristics of the three-dimensional vocal-tract model based on the radiated power is presented. As the radiated power can be calculated without specifying a fixed output point, the proposed method can be used for the configuration with a complex shape in the radiation. The sound wave propagation of the three-dimensional vocal-tract model is simulated under rigid wall (lossless) and soft wall (loss) condition. The transfer characteristics calculated using the proposed method and distributions of active sound intensities are presented. The results show the following aspects: under the rigid wall condition, a lot of sharp peaks and zeros whose bandwidths are extremely narrow occur in the transfer characteristics, and strange sound energy circulation occurs in the vocal-tract model. Under the soft wall condition, the bandwidths of the peaks and zeros are enlarged and fewer peaks and zeros occur in the transfer function than those under the rigid wall condition. And the sound energy circulation between the nasal tract and the oral cavity is suppressed.

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

II. Geometrical vocal-tract models

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.

A three-dimensional volume of radiation[3] with a radius of 4 cm, which is spherical in shape, is attached to the face covering the lips and nostrils. A specific acoustic impedance of spherical waves is used as a boundary condition on the round surface of the three-dimensional volume of radiation. The surface mesh of the vocal-tract model is shown in Fig. 2.

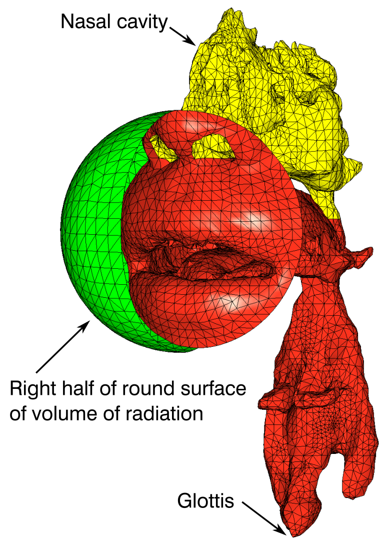

Fig. 2. Surface meshes of the vocal-tract model.

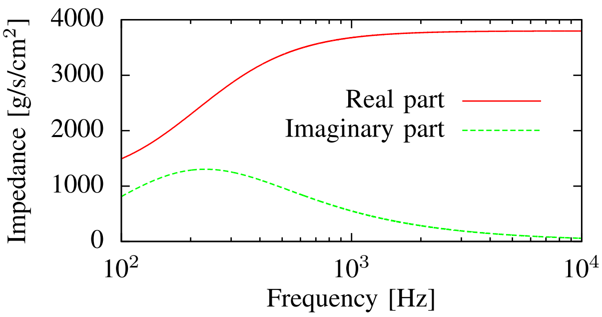

Fig. 3. Frequency characteristics of wall impedance.

III. FEM SIMULATION

A three-dimensional FEM was applied to the wave equation in a steady state

$$\nabla^2 \phi = -k^2 \phi \tag{1}$$

where ϕ is a velocity potential and k is a wavenumber. Sound pressure p is obtained by

$$p = j\omega\rho\phi \tag{2}$$

where ω is an angular frequency and ρ is an air density. Particle velocity $\mathbf{v}$ is computed by the following equation.

$$\mathbf{v} = -\nabla\phi \tag{3}$$

The three-dimensional FEM is again applied to this computataion. Acoustic wave propagation in the geometrical vocal-tract models was simulated by using the FEM. 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, namely, rigid and soft wall condition, were assumed on the walls of the vocal tract and face. The wall impedance proposed by Kamiyama *et al.* [4] was assumed as the boundary condition of the soft wall with a thickness of 2 cm, as shown in Fig. 3. The simulation was carried out in a driving frequency range of 100 Hz to 8 kHz at intervals of 10 Hz and then 1 Hz in the vicinity of peaks and zeros.

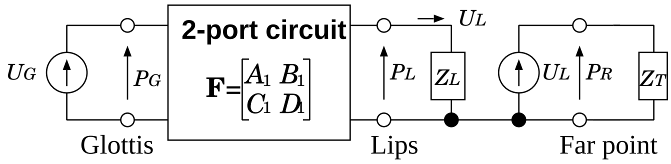

Fig. 4. Electrical equivalent representation of one-dimensional speech production model.

IV. ACOUSTIC OUTPUT AND TRANSFER CHARACTERISTICS

A. One-dimensional case

Fig. 4 shows an electrical equivalent representation of one-dimensional speech production model. The transfer function of the vocal-tract, H_1, can be defined as the ratio of the volume velocity at the lips, U_L, to that of the glottis, U_G. H_1 can be easily calculated using a cascade matrix $\mathbf{F}$ representing the vocal-tract as an equivalent 2-port electrical circuit as follows.

$$H_1 = \frac{1}{C_1 Z_L + D_1} \tag{4}$$

C_1 and D_1 are the components of $\mathbf{F}$, and Z_L is a radiation impedance at the lips. Then a sound-pressure at a far point, P_R, can be obtained as,

$$P_R = U_L Z_T = U_G H_1 Z_T \tag{5}$$

where Z_T is a transfer radiation impedance, and has frequency characteristics approximately 6dB/octave.

B. Acoustic field in 3D model

As shown in Fig. 2, the radiating area at the lips is spatially continuous which means that the clear terminal boundary of the vocal-tract is difficult to specify from the physical shape of the lips. In the vicinity of the lips, sound waves radiated from both the lips and the nostrils are superposed. As a result, extreme low sound-pressure areas can possibly appear in some frequencies as shown in Fig. 5. Furthermore, in higher frequencies, the radiation of higher-order modes may form a complicated sound-pressure distribution outside the vocal-tract [5]. The spatial averaged values of sound-pressure or particle velocity in the radiating area should not be used as the output of the vocal-tract.

C. Radiation power

Assume a sound-pressure $p(\mathbf{r})$ and particle velocity $\mathbf{v}(\mathbf{r})$ at the position $\mathbf{r}$. An active sound intensity $\mathbf{I}(\mathbf{r})$ is obtained as,

$$\mathbf{I}(\mathbf{r}) = \mathrm{Re}\{p(\mathbf{r})\mathbf{v}^*(\mathbf{r})/2\} \tag{6}$$

where $\mathrm{Re}\{\cdot\}$ represents to take a real part, and * denotes complex conjugate. The total acoustic power passing through an arbitrary plane S in the vocal-tract and a closed surface C surrounding the lips and nostrils as shown in Fig. 6 can be calculated as follows.

$$W_S = \iint_S \mathbf{I}(\mathbf{r}) d\mathbf{s} \geq \iint_C \mathbf{I}(\mathbf{r}) d\mathbf{s} = W_C \tag{7}$$

W_C represents a total radiation power, and can be approximately calculated using $p(\mathbf{r})$ and $\mathbf{v}(\mathbf{r})$ in the vicinity of the

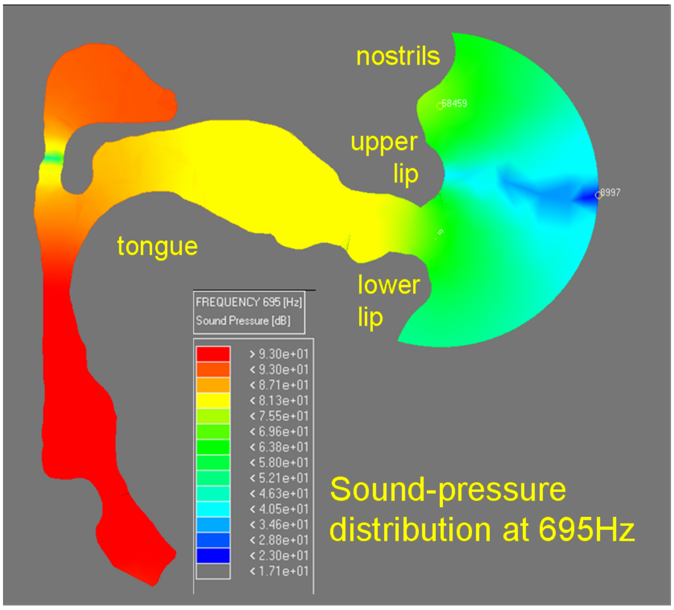

Fig. 5. Sound-pressure distribution on the mid-sagittal plane at 695 Hz.

lips. The difference between W_S and W_C is a power dispersion in the vocal-tract when W_S is evaluated at the glottis end. In the case of the loss-less vocal-tract model, W_C is always equal to W_S, which makes easier to calculate the radiation power.

D. Evaluation by transfer impedance

Using a sound-pressure P_R at a far point R, the total radiation power W_C can be written as,

$$W_C = \iint_C \mathbf{I}(\mathbf{r})ds \approx \frac{1}{2\rho c} \iint_C |P_R|^2 ds \tag{8}$$

where ρc is a characteristic impedance of air. Then, $|P_R|$ is obtained as,

$$|P_R| = K\sqrt{W_C} \tag{9}$$

where K includes directivity factor D of radiation, and is defined as,

$$K = \sqrt{\frac{2\rho c}{\iint_C D^2 ds}}. \tag{10}$$

The transfer characteristics of the model are evaluated in terms of the transfer impedance Z_P, the ratio of the sound-pressure P_R at a distant point to the given source volume velocity U_G.

$$|Z_P| = \left|\frac{P_R}{U_G}\right| = K\frac{\sqrt{W_C}}{|U_G|} \tag{11}$$

As the radiated power can be calculated without specifying a fixed output point, the transfer characteristics of the three-dimensional model can be evaluated using $|Z_P|$. In the following computation K is set constant although D has a frequency dependency.

V. Effect of Wall Boundary Condition

A. Transfer characteristics

Transfer characteristics of the three-dimensional vocal-tract model computed using (11) are shown in Fig. 7. For the

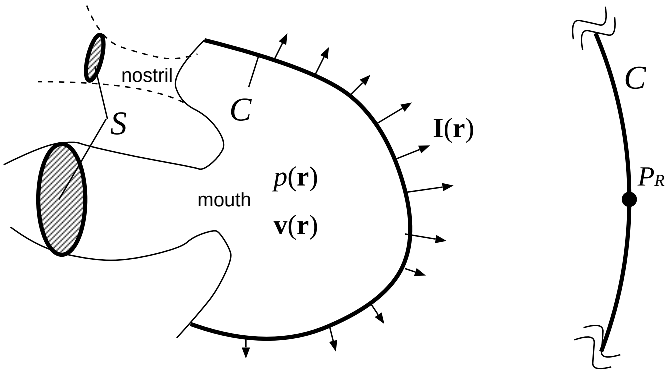

Fig. 6. Arbitrary section S in vocal-tract and closed surface C in free space.

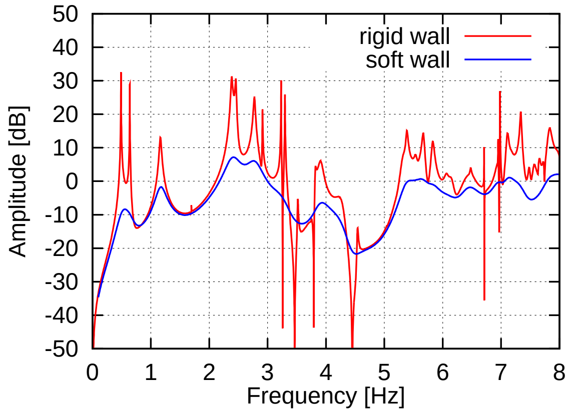

Fig. 7. Transfer characteristics of the vocal-tract model.

rigid wall condition, peaks and zeros are very sharp. The first formant (F1) and the second formant (F2) frequencies are 488 Hz and 1163 Hz, respectively. An additional peak caused by the nasal cavity appears at 639 Hz between F1 and F2. Kitamura *et al.* [6] reported that the F1 and F2 frequencies of the average spectral envelope of speech data, which are recorded from the same subject, are 563 Hz and 1047 Hz, respectively. There are large differences between the formant frequencies of the model and those of the speech data. The lossless condition potentially causes the differences, especially the wall boundary condition, as it is well known from one-dimensional modeling that the yielding wall has an effect of upward shifts of lower formant frequencies.

For the soft wall condition, the sharp peaks and zeros shown in the vocal-tract transfer function under the rigid wall condition are not observed and the bandwidths are enlarged because of wall loss[7]. F1 and F2 frequencies are 554 Hz and 1173 Hz, respectively. Compared with the formant frequencies for the rigid wall condition, the upward shifts of F1 and F2 can be confirmed. The additional peak at 639 Hz has disappeared. The broad bandwidth of each peak possibly merges this additional peak with the F1. The difference in the formant frequencies between the model and speech data decreased greatly at F1, while increasing slightly at F2 compared with those for the rigid wall condition. These discrepancies may be reduced by adjusting the wall impedance since there are still many immature points left in the simulation; for example, the

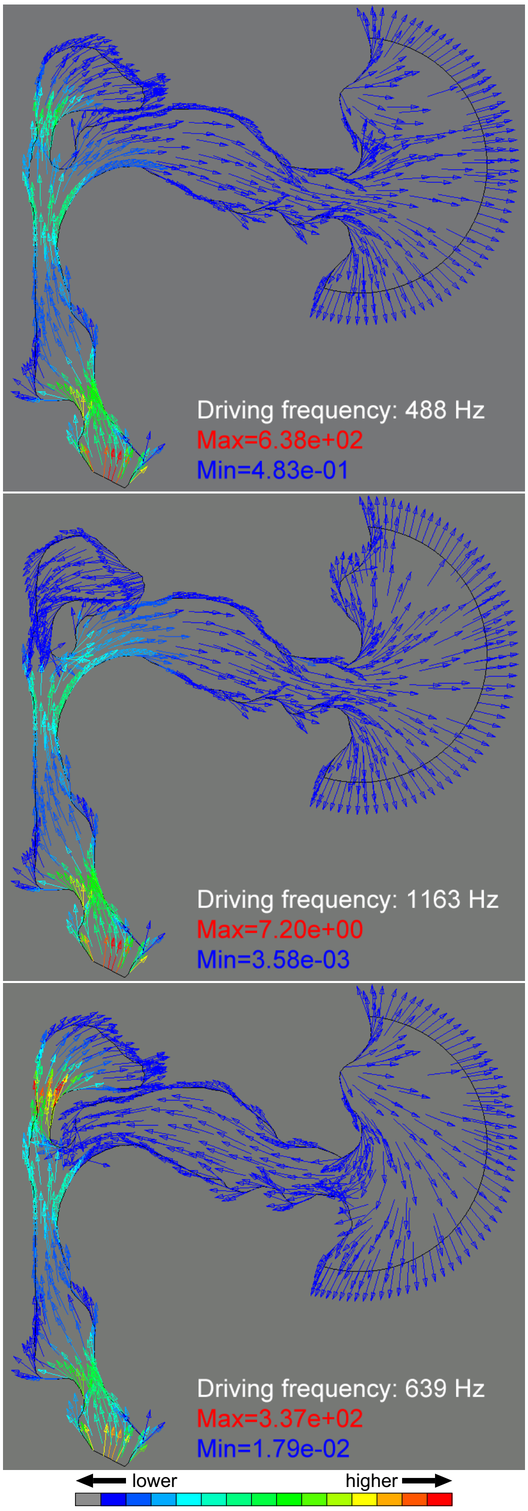

Fig. 8. Distributions of active sound intensity vectors on mid-sagittal plane under rigid wall condition.

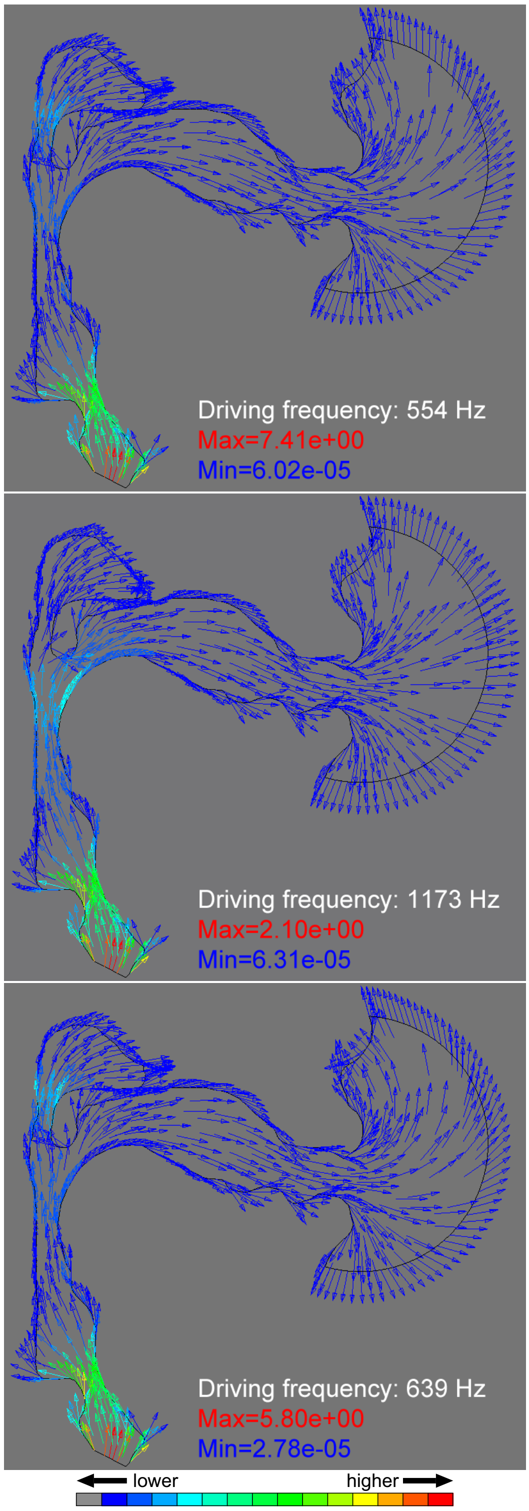

Fig. 9. Distributions of active sound intensity vectors on mid-sagittal plane under soft wall condition.

wall impedance distributed homogeneously and uniformly on the wall of the models.

B. Active sound intensity

Distributions of active sound intensity vectors on mid-sagittal plane under rigid wall condition are shown in Fig. 8. The frequencies 488 and 1163 Hz correspond to F1 and F2 frequencies, and 639 Hz is the frequency of the peak caused by the nasal cavity. At F1 frequency, a sound energy flow from the glottis bifurcates into the oral and nasal cavities. At F2 frequency, 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.

Distributions of active sound intensity vectors on mid-sagittal plane under soft wall condition are shown in Fig. 9. The frequencies 554 Hz and 1173 Hz correspond to F1 and F2 frequencies, and 639 Hz is the frequency of the peak which appeared in the transfer function under the rigid wall condition. Compared with Fig. 8, the radiation from the nostrils decreased greatly at F1 frequency. The sound energy flow from the nasal cavity at F2 frequency 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.

VI. Conclusion

The acoustic characteristics of the vocal-tract model based on MRI data of the Japanese /a/ were computed by the three-dimensional FEM. The oral cavity was also coupled with the nasal cavity in the three-dimensional volume of radiation.

We showed that the transfer characteristics of the three-dimensional model can be evaluated using the transfer impedance regardless of the physical shape of the lips. And we showed that the simulation under the wall loss condition does not cause strange results. 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

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

References

[1] H. Matsuzaki, A. Serrurier, P. Badin, and K. Motoki, “3D and plane wave acoustic propagation comparison for a Japanese and a French vowel /a/ with nasal coupling,” *Proc. Autumn. meeting Acoust. Soc. Jpn.*, 2-4-15, pp.471-474, 2007.

[2] H. Matsuzaki, and K. Motoki, “Study of acoustic characteristics of vocal tract with nasal cavity during phonation of Japanese /a/,” *Acoustical Science and Technology*, vol. 28, 2, pp.124-127, 2007.

[3] 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)*, vol. 17, pp. 163–166, 1996.

[4] N. Kamiyama, N. Miki, and N. Nagai, “Study of the vocal tract wall impedance using viscoelastic model of the wall,” *IEICE Trans. Fundamentals*, vol. J75-A, pp. 1649–1656, 1992.

[5] K. Motoki, and N. Miki, “Effects of higher-order modes on sound radiation characteristics at mouth opening,” *Proc. Third Joint Meeting of Acoust. Soc. Am. and Acoust. Soc. Jpn.*, 2pSC42, pp.895-898, 1996.

[6] 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.

[7] J. Suzuki, “Discussions on vocal tract wall impedance,” *J. Acoust. Soc. Jpn. (J)*, vol. 34, pp. 149–156, 1978.