FEM Analysis on Acoustic Characteristics of Vocal Tracts Shape with Different Geometrical Approximation

Hiroki Matsuzaki* and **Kunitoshi Motoki***

Abstract

The acoustic features of speech characterized by 3-D vocal tract shapes are investigated by a 3-D FEM simulation. Simulation models have cascaded structures of 36 cross sections based on MRI data for the Japanese vowel /a/ of an adult male. To compare simplified structures with a non-simplified structure of the vocal tract, a reference model is used as the non-simplified structure and rectangular models and two elliptical models are used as simplified structures. Vocal tract transfer functions (VTTFs) are computed from the simulation results. It is concluded that the elliptical model with bent configuration is useful for the analysis of the acoustic characteristics of the vocal tract by the FEM below 6kHz.

1 INTRODUCTION

We have examined the acoustic characteristics of the vocal tracts using a 3-D FEM which is a simulation method suitable for computing the acoustic field inside an arbitrary 3-D shape. As a first step, we used a simple simulation model of the vocal tract constructed from cascading acoustic tubes with elliptical shapes [1, 2]. Although these simulation results have shown that the acoustic characteristics of the 3-D models are different from those of the traditional 1-D equivalent circuit model [3], it is still not clear whether the simplified 3-D models are valid.

The purpose of this study is to investigate the validity of the simplification of the vocal tracts with regard to the cross-sectional shape. A 3-D finite element method (FEM) applied to an acoustic wave equation in a steady state is used to obtain the distributions

---

* Faculty of Engineering, Hokkai-Gakuen University
E-mail : {matsu, motoki}@ell.hokkai-s-u.ac.jp

of sound pressure in 3-D vocal tract models. The 3-D vocal tract models have cascaded structures of 36 cross sections and are based on magnetic resonance imaging (MRI) data of the vocal tract for the Japanese vowel /a/ of an adult male. To compare simplified structures with a non-simplified structure of the vocal tract, the following models are used: a reference model, a rectangular model, and two elliptical models. The reference model is composed as close as possible to that of the original MRI data, and is used as a basis for the other models. The cross-sectional shapes for the rectangular and elliptical models are rectangular and elliptical, respectively. The areas and perimeters of these cross sections coincide with those of the original MRI data. The rectangular model and one of the elliptical models are configured without a bend of the vocal tract, while the other elliptical model has a bent geometry. The distributions of sound pressure are obtained by the simulation, and the VTTFs are computed from these distributions. The simulation results are summarized as follows: (1)The formant frequencies of the rectangular model and the elliptical model with straight configuration are lower than those of the reference model. (2)The shifts are smaller for the elliptical model. (3)The formant frequencies of the elliptical model with bent configuration are in good agreement with those of the reference model, even up to the sixth formant frequency. (4)In the higher frequencies above 6 kHz, the VTTFs of the simplified models are largely different from that of the reference model. From these results it is concluded that the elliptical model with the bend is useful for the analysis of the acoustic characteristics of the vocal tract by the FEM below 6 kHz.

2 SIMULATION METHOD

It is well known that an acoustic wave equation in a steady state is represented using velocity potential ϕ as

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

where k is the wavelength constant. A Sound pressure p is obtained by

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

where ρ is the air density. The 3-D FEM is applied to Eq.(1)[4].

3 3-D VOCAL TRACT MODELS

3-D vocal tract models have cascaded structures of 36 cross sections, from the glottis to the lips, and are based on MRI data of the vocal tract for the Japanese vowel /a/ of an adult male.

3.1 Reference Model

The reference model is composed as close as possible to that of the original MRI data, and is used as a basis for the other models. The original MRI data have branches such as the pyriform fossa and the epiglottis. These branches, however, are omitted to facilitate the creation of finite elements for each model. The finite element(FE) mesh of the reference model is shown in Figure 1. A rigid wall is assumed. To simulate the 3-D radiation into a free space, the 3-D radiational model with a radius of 4cm, which is hemispherical in shape, is attached to the aperture surface (the lips side)[6]. A specific acoustic impedance of spherical waves is used as a boundary condition on the round surface

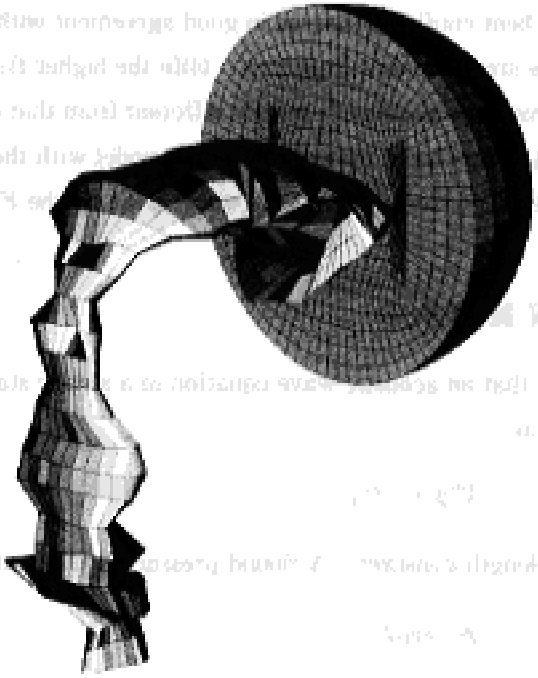

Figure 1: FE mesh of the reference model. A hemispherical volume is the 3-D radiational volume with a radius of 4cm.

of the hemisphere, and a rigid wall is assumed on the cut surface which is regarded as a suface of a plane baffle. The driving surface (the glottis side) is driven by particle velocity $\exp(j\omega t)$.

3.2 Simplified Models

Each cross section of simplified models is obtained by converting the MRI data into elliptical or rectangular shape. The area and perimeter of each cross section coincide with those of the original MRI data. An example of the converted cross-sectional shapes is shown in Figure 2.

As shown in Figure 3 the interval length between two cross sections is defined as the distance between the central points of the coadjacent cross sections. The central point is determined as the center of the intersection line of the cross section and the mid-sagittal plane of the reference model. The total length of the vocal tract is 17.836cm.

The rectangular model and one of the elliptical models are configured without a bend of the vocal tract, while the other elliptical model has a bent geometry. The angle of the bend is approximately the same as that of the reference model. For the straight configuration, each cross section is aligned symmetrically with respect to the straight central line. The FE meshes are shown in Figure 4 for the rectangular model and in Figure 5 for the elliptical models. These models also have the same 3-D radiation volume and boundary conditions as those of the reference model.

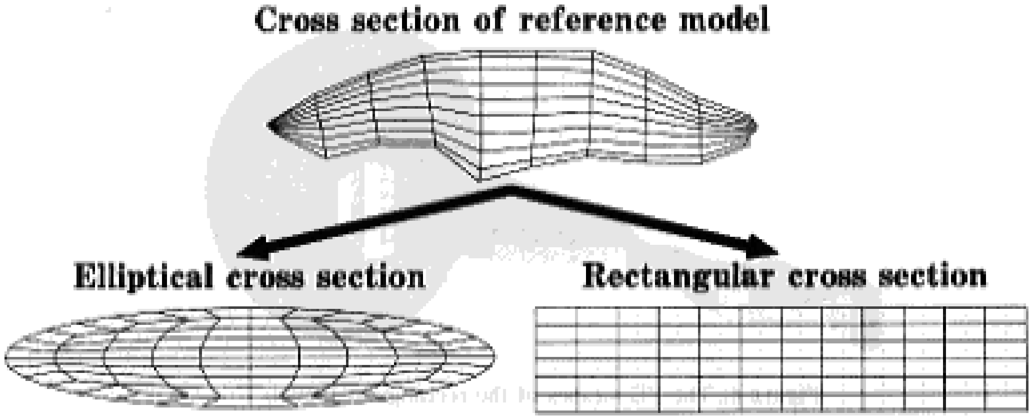

Figure 2: An example of the converted cross-sectional shapes. The simplified cross sections have the same area and perimeter as those of the original MRI data.

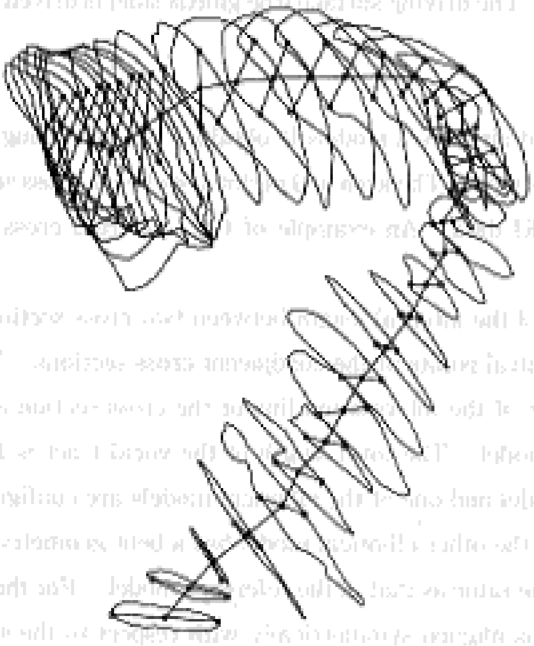

Figure 3: This figure shows the cross sections, the intersections of the cross sections and the mid-sagittal plane, and the lines between the central points of the coadjacent cross section.

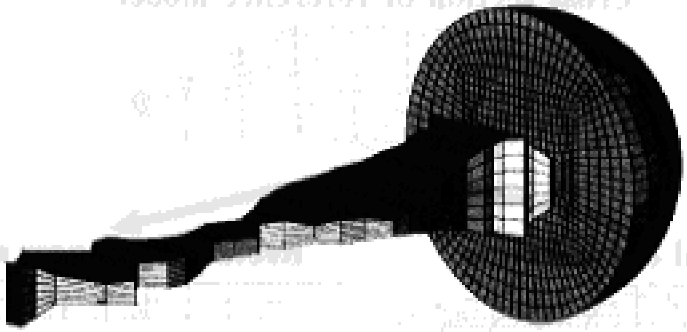

Figure 4: The FE meshes of the rectangular model.

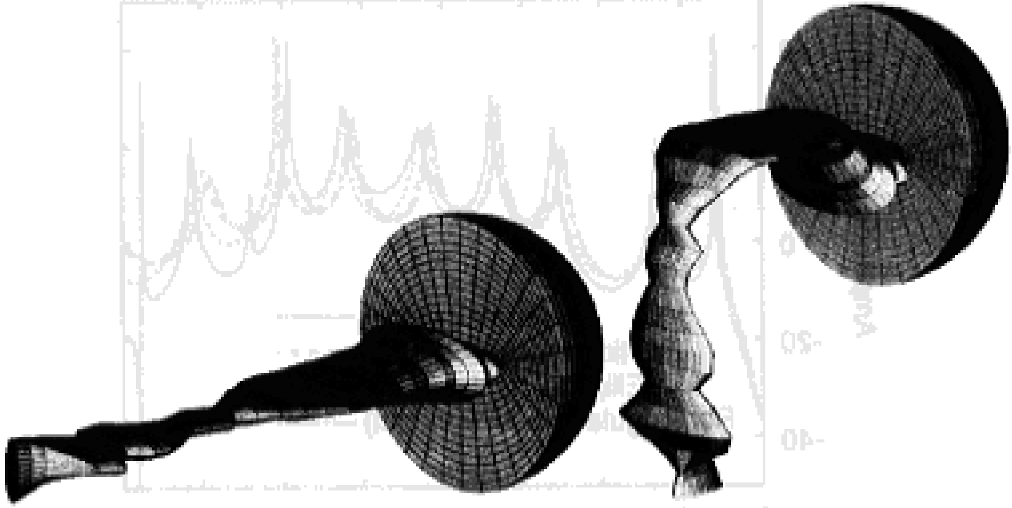

Figure 5: The FE meshes of the elliptical models. left: straight configuration, right: bent configuration.

4 VTTFS FROM FEM SIMULATION

From the distribution of sound pressure obtained from the FEM simulation, the VTTF $H(\omega)$ is defined as [5]

$$H(\omega)=K\left|\frac{\sqrt{W_{rad}}}{u_g}\right| \tag{3}$$

where W_{rad} is a radiation power equivalent to the total active intensities on the surface of the 3-D radiational volume. u_g is a source volume velocity. K is a constant for $H(\omega)$ to be dimensionless.

Figure 6 shows the computed VTTFs. The solid line shows the VTTF of the reference model. The VTTFs for elliptical models are shown as a broken line for the straight configuration and a dotted line for the bent configuration, respectively. The VTTF for the rectangular model is shown as a one-dot chain line.

Table 1 shows the formant frequencies and percentages of shifts of the formant frequencies relative to those of the the reference model. The percentage of the shift is defined as

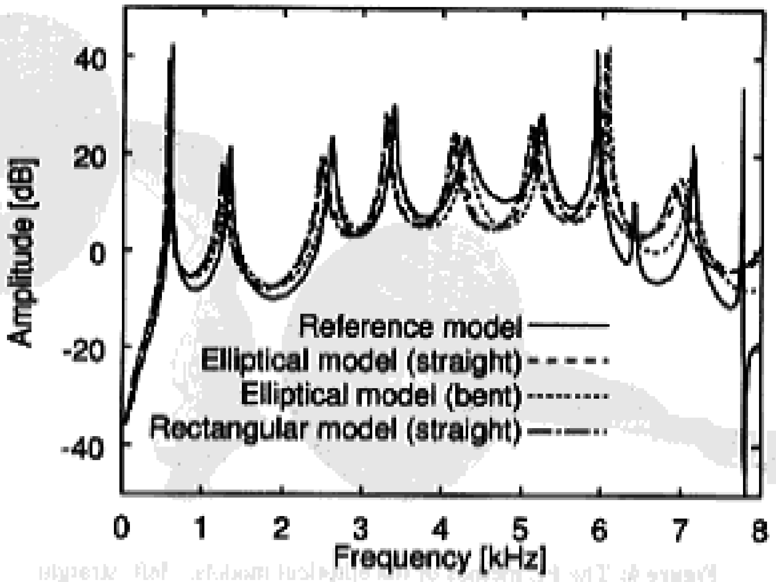

Figure 6: VTTFs from FEM simulation.

$$\frac{f_{\mathrm{Simp}} - f_{\mathrm{Ref}}}{f_{\mathrm{Ref}}} \times 100 \tag{4}$$

where f_{Simp} and f_{Ref} are the formant frequencies of the simplified models and of the reference model, respectively.

Table 1 : Formant frequencies [Hz] and percentages of the shifts. The shifts are shown in parentheses.

	refrence model	rectangular model		elliptical model			
				straight		bent	
1st	588	551	(-6.59)	553	(-5.95)	563	(-4.25)
2nd	1332	1251	(-7.58)	1302	(-6.08)	1231	(-2.25)
3rd	2609	2479	(-4.98)	2514	(-3.64)	2609	(0.00)
4th	3390	3287	(-3.04)	3315	(-2.21)	3390	(0.00)
5th	4302	4138	(-3.81)	4188	(-2.65)	4302	(0.00)
6th	5240	5108	(-2.52)	5161	(-1.51)	5260	(0.38)
7th	5922	6091	(2.85)	6042	(2.05)	6082	(2.70)

The formant frequencies of the rectangular model and the elliptical model with straight configuration are lower than those of the reference model in the range of -1.5 percent — -7.6 percent. The length of each section for the straight configuration is set equal to the interval length of adjacent cross sections of the original MRI. As the rate of shift is always negative up to the 6th formant frequency, it can be said that the use of the interval length results in the overestimate of the vocal tract length for the straight configurations.

The shifts of the elliptical model are smaller than those of the rectangular model. Consequently, it can be said that for the simplification of the vocal tract the use of the elliptical cross sections gives better geometrical approximation of the real vocal tract than the use of the rectangular cross sections.

The formant frequencies of the elliptical model with bent configuration shows very good agreement with those of the reference model even up to the sixth formant frequency. From these results it is concluded that the elliptical model with bent configuration is useful for the analysis of the acoustic characteristics of the vocal tract by FEM below 6 kHz.

In the higher frequencies above 6 kHz, the VTTFs of the simplified models are largely different from that of the reference model. In the VTTFs of the reference model a small peak appears near 6.4kHz and a sharp peak near 7.8kHz, which might be effects of the asymmetrical shape.

5 CONCLUSION

Using the 3-D FEM we have simulated the acoustic wave propagation in the precise and simplified 3-D vocal tract models.

The formant frequencies of the elliptical model with bent configuration were in good agreement with those of the reference model. Therefore the elliptical model with bent configuration is more useful than other simplified models.

ACKNOWLEDGMENTS

We wish to express our gratitude to Prof. N.Miki, Future University-Hakodate, for valuable advice. Part of this work has been supported by a research project of High-Tech Research Center (Hokkai-Gakuen University), by a Grant-in-Aid for Scientific Research (12750345) from the Ministry of Education of Japan, and by Crest of JST (Japan Science and Technology).

References

[1] H. Matsuzaki, N. Miki and Y. Ogawa, "3-D FEM analysis of vocal tract models using elliptic tubes with volume radiation," The XIIIth International Congress of Phonetic Science, Stockholm, Sweden, vol.4, 82.12, pp.440−443 (1995).

[2] H. Matsuzaki, N. Miki and Y. Ogawa, "3-D Finite element analysis of Japanese vowels in elliptic sound tube model," Jpn. IEICE Trans(A), vol.J81-A, no.6, pp.899−906 (1998) (in Japanese).

[3] J.L. Flanagan, "Speech analysis synthesis and perception," 2nd Ed. pp.25−35, Springer-Verlag (1975).

[4] H. Matsuzaki, T. Hirohku, N. Miki and N. Nagai, "Analysis of acoustic characteristics in the vocal tract with inhomogeneous wall impedance using three dimensional FEM model," Jpn. IEICE Trans(A), J77-A, No.3, pp.325−333 (1994) (in Japanese).

[5] K. Motoki, "A method of constructing an acoustic model of vocal-tracts using higher-order modes," J. Acoust. Soc. Jpn. 54, 12, pp.850−856 (1998) (in Japanese).

[6] H. Matsuzaki, N. Miki and Y. Ogawa, "FEM analysis of sound wave propagation in the vocal tract with 3-D radiation model," J. Acoust. Soc. Jpn. (E) 17, 3, pp.163−166 (1996).