A Fundamental Study of Relationship Between Vocal Tract Shapes with Diseases of Oral Cavity and Speech Spectra

Hironori Nishimoto

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 2001

Keywords: Diseases of Oral Cavity, Magnetic Resonance Imaging (MRI), Finite Element Method (FEM), 3-D Vocal Tract Model, Transfer Funcion of Vocal Tract, Formant Frequency .

1 Introduction

The purpose of this paper is to find handholds about relationship between vocal tract shape with diseases of oral cavity and voice spectra.

Disease of oral cavity or aftereffect occurs speech disorder. This is because vocal tract shape is different in comparison with that of normal persons, by cases of tongue and mouth floor resection. Palatal Augmentation Prosthesis (PAP) is a treatment for speech disorder. However, there is no standard for a diagnosis of speech sound distortion and design for PAP. Thus, it is important to find relationship between the vocal tract shape with disease of oral cavity and speech spectra.

In this research, Finite Element Method (FEM) is used for simulation to deal with three-dimensional characteristics of the vocal tract shape. Firstly, vocal tract shapes are measured from Magnetic Resonance Imaging (MRI). Secondly, three-dimensional vocal tract models are cunstructed on a computer from vocal tract shape data. Finally, transfer functions of vocal tracts are estimated by FEM. This paper shows estimated transfer functions of modified three-dimensional vocal trafct models. It shows comparisons simulated results by FEM with results of acoustic analisys. The purpose is to find relationship between vocal tract shapes and voice spectra.

2 Acoustic Analysis

Details of the target subjects in this research are shown in the Table 1 . Patient-A is a case of tongue and mouth floor resection, and PAP is applied. Patient-B is a case of mouth floor resection.

Table 1: Subjects

Subject	Age	Sex	Case
Normal-A	24	male	Normal
Normal-B	26	male	Normal
Patient-A	37	male	Tongue and mouth floor resection
Patient-B	72	male	Mouth floor resection

Japanese vowels /i/ utterd by each subjects were analysised, and means and S.D. of formant frequencies were estimated by unbiased estimation of log spectrum [2]. There were used for verification of the validity of simulated results.

3 Simulations with Normal Subjects

3.1 Simulation with Normal-A

Fig. 1 shows a vocal tract model of Normal-A and transfer functions of vocal tract model of Normal-A simulated by FEM. The three-dimensional vocal tract model has a hemispherical surface of radiation at the mouth-lip to express radiation from the mouth-lip.

The estimated transfer function is similar to the results of acoustic analisys except F_2. The number of peaks of the eistimated transfer function is different from the number of formants of the result of acoustic analysis. It causes that formant frequencies are estimated from spectrum envelops. If poles are close, a spectrum appears as one peak. Additionally, this vocal tract model is supposed that its wall is rigid and air does not have viscosity. Then, peaks of estimated transfer functions become sharply.

There is no influence of transfer functions on the vocal tract model with hemispherical surface of radiation below F_2. However, the peaks frequencies of the transfer functions over F_3 is changed by installing hemispherical surface of radiation. The difference between the results of acoustic analysis and the estimated transfer functions is tolerance except F_2 because of the following reasons.

- Measurement values from MR images have ± 5[%] error[1].
- Analysised formant frequencies have ± 10[%] range.

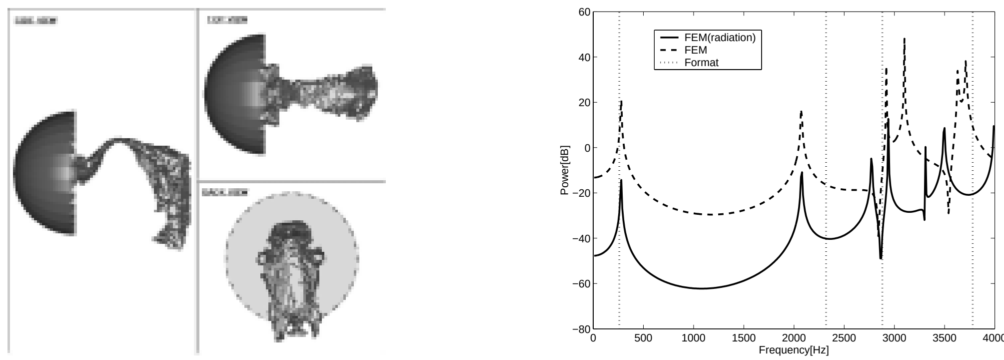

Figure 1: A 3-D vocal tract model of Normal-A with hemispherical surface of radiation and transfer functions of a Normal-A vocal tract model with hemispherical surface of radiation and without hemispherical surface of radiation.

3.2 Simulation with Normal-B

The estimated transfer functions of Normal-B vocal tract model approximate the results of acoustic analisys except F_1 and F_2. The fundamental frequency of Normal-B was analysised and it was $160[Hz]$. Therefore, the result of acoustic analysis of Normal-B speech is close to the component of the second harmonic.

These simulated results lead that this simulation method is useful to estimate transfer functions of a three-dimensional vocal tract model except F_2.

4 Influence on a Sound Spectrum by Removing Pyriform Fossas

The back cavity of the vocal tract affects F_2 when utteriig Japanese vowel /i/. Fig. 2 shows that the transfer function of the vocal tract model removed pyriform fossa approaches a result of acoustic analysis. Pyriform fossa is an organ that is expanded or contracted during utterance. When taking MR images, pyriform fossa shape is deformed. Then, pyriform fossa is not actually played as branches of the vocal tract. Therefore, the cases that pyriform fossas are removed like this model correspond to actual voice spectra.

5 Simulation with a Subjects of Tongue and Mouth Floor Resection

Simulations by the three-dimensional vocal tract model of tongue and mouth floor resection were done.

Simulatied results of vocal tract models with hemispherical surface of radiation were closer to acoustic analysis results with each subject than those without surface of radi-

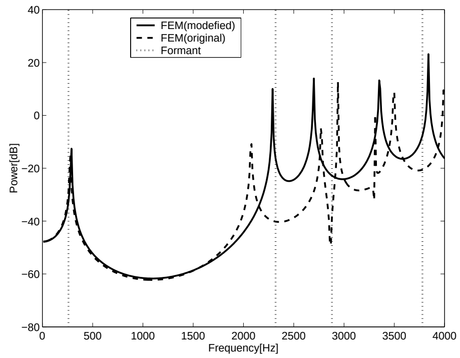

Figure 2: A transfer function of a Normal-A vocal tract removed pyriform fossas.

ation spherical. However, when formant frequencies over F_3 were compared, simulatied results did not approximate with the results of acoustic analysis. In patient cases, patient articulation is not fixed. As the result, there were differences between the simulated results and the results of acoustic analysis.

5.1 Simulation Results by FEM and 1-D Model

Fig. 3 is simulated results of Patient-A (PAP (−)) by one-dimensional model [3] and FEM. There are four peaks of the estimated transfer function of the vocal tract by FEM. From acoustic analysis, there are four formants in $0 - 4000[Hz]$ range. However, the estimated transfer function of the vocal tract by one-dimensional model has only 3 peaks of estimated transfer function in same frequensy range. This is because one-dimensional model did not express a sound distortion caused of the abnormal shape of oral cavity.

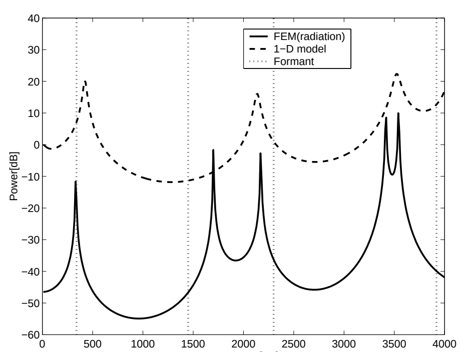

Figure 3: A transfer function of a Patient-A(PAP(−)) vocal tract model simulated by FEM and 1-D model.

5.2 A Simulation by the modefied Patient-A(PAP(−)) model

Breath of Patient-A (PAP (−)) goes out from right hand side toward the center shaft of the vocal tract. Deformation is given to a part of the three-dimensional vocal tract model of Patient-A(PAP(−)). The model is made to slide based on the center shaft of the vocal tract. There are not differences between the estimated transfer functions in $0 - 4000[Hz]$ range. This suggests that the vocal tract shape of Patient-A (PAP(−)) is little biased against the center shaft of the vocal tract. Therefore, if deformation is given to a part of the vocal tract model, a route of breathing is not changed. On the other hand, there is a difference between simulation results over $4000[Hz]$. Smoothness on a surface of a vocal tract shape and its continuity are often taken influence. Therefore, when PAP of Patient-A(PAP(−)) is designed, the breathing route of the vocal tract shape is not important from the above result.

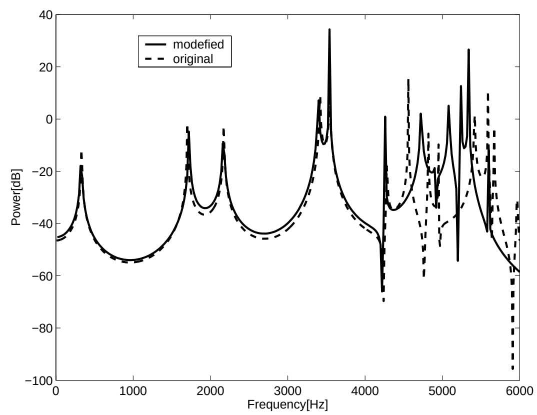

Figure 4: A transfer function of a modefied vocal tract model of Patient-A (PAP(−)).

6 Conclusion

This paper showed comparisons simulation results by FEM with results of acoustic analysis. It showed that simulation results by FEM approximate better a sound distortion than those by 1-D model. This is because the three-dimensional characteristics of the vocal tract. As the result, this method is useful to simulate spectra from vocal tract Moreover, it showed that a relationship between a patient vocal tract shape and voice spectrum by modified a patient vocal tract model.

References

[1] H. Saito, et al. “3-Dimensional Measurement of Vocal Tract Shape Using MRI −Methodology and Area Function in Normal Subjects−”, J. Jpn. Stomatol. Soc.

49(2), pp.92-101, 2000

[2] S. Imai, D. Furuichi, “Unbiased Estimation os Log Spectrum”, IEICE, Vol.J70-A, No.3, pp.471-480, 1987

[3] Man Mohan Sondhi, JUERGEN SCHROETHER, “A hybrid time-frequency domain articulatory speech synthesizer”, IEEE Transactions on acoustics, speech, and signal processing, Vol. ASSP-35, No. 7, July, pp955-967, 1987