Title	Effects of complicated vocal tract shape on vocal tract transfer function
Author(s)	Nishimoto, Hironori; Akagi, Masato
Citation	Journal of Signal Processing, 10(4): 267-270
Issue Date	2006-07
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/4890
Rights	Copyright (C) 2006 Research Institute of Signal Processing Japan. Hironori Nishimoto and Masato Akagi, Journal of Signal Processing, 10(4), 2006, 267-270.
Description	

Journal of Signal Processing, Vol. 10, No. 4, pp. 267–270, July 2006

SELECTED PAPER

Effect of Complicated Vocal Tract Shape on Vocal Tract Transfer Function

Hironori Nishimoto and Masato Akagi

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Phone:+81-761-51-1699 (ext.1391) FAX:+81-761-51-1390
Email: {nhiro; akagi}@jaist.ac.jp

Abstract

To investigate the effects of vocal tract shape on the formant frequency of speech sounds below 4000 Hz, in this study, we compare the estimated transfer functions and sound pressure distributions of three dimensional (3-D) vocal tract models with complicated and simple shapes. A 3-D vocal tract model (Model-O) with a complicated shape caused by oral lesions is designed by magnetic resonance imaging (MRI). Four deformed vocal tract models are constructed on the basis of Model-O. These deformed models have the same cross-sectional-area function and different cross-sectional shapes. The transfer functions of these models are estimated by means of the finite element method (FEM) and the equivalent circuit model using vocal tract transfer function. FEM can estimate the number of peaks of the transfer functions of Model-O, the same as the number of speech formants. However, for a transfer function estimated by the equivalent circuit model, even the number of peaks is different from that of speech formants. These results suggest that FEM can estimate transfer functions accurately. Even in the low-frequency region, not only a cross-sectional-area function but also the shape of the vocal tract affects the vocal tract transfer function. The transfer functions of the deformed models replaced with the oral area by cylindrical tubes estimated by FEM are also different from speech spectra. This is because the sound pressure contour of such models shows that sound wave is planar in the oral cavity.

1. Introduction

People who have oral lesions have distorted speech sounds. Generally, their vocal tract shapes are complicated because of operations and disorders in their articulatory functions [1]. These shapes are one of the causes of distorted speech sounds. Therefore, the estimation of the transfer functions of a vocal tract with a complicated shape is important in determining what causes distorted speech and how to treat speech disorders.

To determine the effects of complicated vocal tract shape on the vocal tract transfer function in low-frequency regions, we discuss in this paper the transfer functions of vocal tract models with complicated and simple shapes below 4000 Hz. The vocal tract model with a complicated shape is obtained from magnetic resonance imaging (MRI) data of articulatory disorders caused by oral lesions and mouth and tongue floor resection in a male patient. Four deformed models that have the same cross-sectional-area function as the original model are constructed the basis of this model. The cross sections of these deformed models are circular; therefore they are different from that of the original one. The vocal tract transfer functions of these models are compared to investigate the effect of vocal tract shape on the vocal tract transfer function. Moreover, the sound pressure distributions of the deformed models are estimated to investigate the acoustical characteristics of complicated vocal tract shapes. In this paper, the finite element method (FEM) is applied to the estimation of vocal tract transfer functions and sound pressure distributions.

2. Vocal Tract Models

2.1 MRI data

A 3-D vocal tract models derived from MRI data are constructed. MRI data are the Japanese vowel /i/ uttered by a Japanese male subject with oral lesions because of an acquired disease. These MRI data are obtained using the MRI scanner MAGNETOM VISION Ver31B (Siemens). The scanning parameters are as follows: sagittal slice plane, 1.95 mm slice thickness, 0.98×0.98 mm pixel image size, 32 slices, 256×256 pixel size. Speech sounds are recorded in a sound-proof room after taking MRI data.

2.2 Vocal tract models

A 3-D vocal tract model (Model-O) is constructed from the MRI data of the subject uttering the Japanese vowel /i/. The interval of nodes in the surface of the model is 1.0 mm. To explain the radiation from the lips, hemispheric volume (r=40mm) [2] is connected to the lips of the model. Figure 1 shows the Model-O. The model has an asymmetrical and

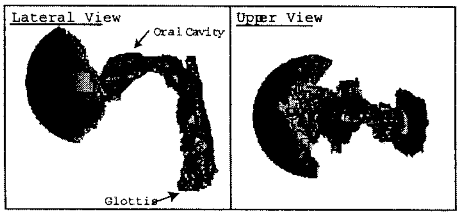

Figure 1: A 3-D vocal tract model uttering Japanese vowel /i/ (Model-O)

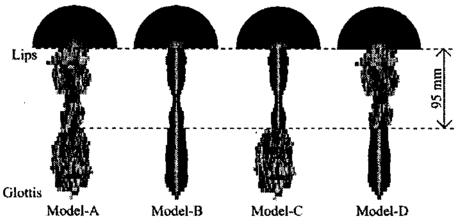

Figure 2: Four 3-D vocal tract models (From left side, Model-A, Model-B, Model-C and Model-D)

complicated cross section. The cross-sectional-area function is measured by Tiede and Honda's method [3].

To observe the effects of oral cavity shape on vocal tract transfer functions, four deformed 3-D vocal tract models (Model-A, Model-B, Model-C and Model-D) are constructed. Their cross-sectional-area functions are equal to that of Model-O. Figure 2 shows these models.

- Model-A: The cross-sectional shape of Model-A is the same as that of Model-O.
- Model-B: Model-B is a connected-cylinder model.
- Model-C: Model-A and Model-B are divided into two components by a section 95 mm from the lips. Model-C consists of the component on Model-A's glottis side and that on Model-B's lip side.
- Model-D: Model-D consists of the component on Model-B's glottis side and that on Model-A's lip side.

A cross-sectional-area function of the subject is measured to estimate a vocal tract transfer function by the equivalent circuit model. In this paper, this function is called Model-1D.

3. Acoustical Analysis

Speech sounds of the subject are recorded in a sound-proof room. Unbiased estimation by the log spectrum method [4] is applied to the analysis of spectral envelopes of sound waves. Figure 3 shows spectral envelopes of the subject uttering the Japanese vowel /i/. The formant frequency of the speech sound is estimated using the spectral envelopes. The F2 of the subject exists at 1,520 Hz and is different from that of a normal spectrum. This is because the subject cannot sufficiently move his tongue and this cause of distorted speech sounds.

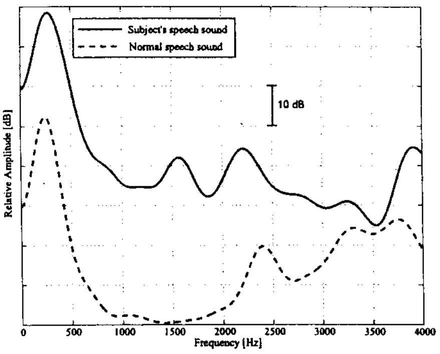

Figure 3: Spectral envelopes of subject uttering Japanese vowel /i/ (solid line) and that of standard speech sound (dashed line)

4. Methods of Estimating Vocal Tract Transfer Function

Vocal tract transfer functions of 3-D vocal tract models are estimated by FEM. An acoustic wave equation in a steady state is represented as

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

where ϕ is the velocity potential and $k(=\omega c)$ is the wave number (ω is the angular frequency and c is the sound velocity). From this equation, p (sound pressure) and u (particle velocity) are represented as

$$\begin{aligned} p &= j\omega\rho c && (2) \\ u &\doteq -\nabla\phi && (3) \end{aligned}$$

where ρ is the air density. In this paper, the parameters of these equations are c=353.46 m/s and $\rho = 1.1421$ kg/m^3. Input surfaces (glottis) are driven by sinusoidal waves with a volume velocity. The impedance of radiation surface Z is obtained using

$$Z = \rho c \frac{jkr}{1 + jkr} \tag{4}$$

where r is the radius of the hemispheric volume. The vocal tract wall and face are assumed as the rigid walls (impedance $Z = \infty$). The vocal tract transfer function H is calculated as

$$H = \frac{\frac{A_{out} \sum I_{out}}{n_{out}}}{\frac{A_{in} \sum I_{in}}{n_{in}}} \tag{5}$$

where $A_{in}(A_{out})$ is the area of input (output) surfaces, $I_{in}(I_{out})$ is the particle velocity of input (output) surfaces

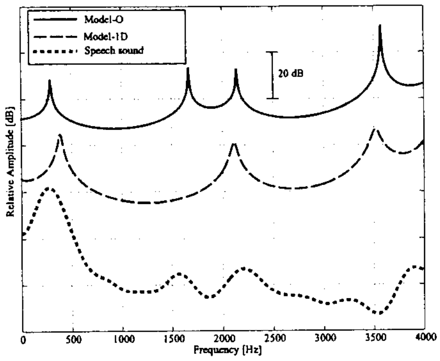

Figure 4: Vocal tract transfer functions of Model-O (FEM) and Model-1D (equivalent circuit model), and spectral envelope of speech sound

and $n_{in}(n_{out})$ is the number of nodes in the input (output) surface. LMS SYSNOISE is used for this analysis.

Vocal tract transfer functions are also calculated by means of an equivalent circuit model using cross-sectional-area function of a vocal tract [5]. A chain matrix are described as

$$\begin{pmatrix} P_{out} \\ U_{out} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P_{in} \\ U_{in} \end{pmatrix} \tag{6}$$

where $P_{in}(P_{out})$ is the input (output) pressure, $U_{in}(U_{out})$ is the input (output) volume velocity, and A, B, C and D are derived from the cross-sectional-area functions, section length and characteristics of air. The transfer function $H(\omega)$ is given by

$$H(\omega) = \frac{1}{A - CZ_L} \tag{7}$$

where Z_L is the impedance of lips.

The vocal tract transfer functions of the models are estimated by these two methods. Calculation range is from 20 Hz to 4,000 Hz in 20 Hz steps.

5. Results

5.1 Vocal tract transfer functions of Model-1D and Model-O, and a speech sound

Figure 4 shows the estimated vocal tract transfer functions of Model-O and Model-1D and spectral envelope of speech sound. The number of peaks and the peak frequencies of the vocal tract transfer function of Model-O correspond to those of the spectral envelope. For the vocal tract transfer function of Model-1D, the peak of the transfer function at approximately 1,500 Hz does not exist and this is different from that of the spectral envelope. These results suggest that FEM should be applied to the estimation of the transfer functions of a vocal tract with a complicated shape.

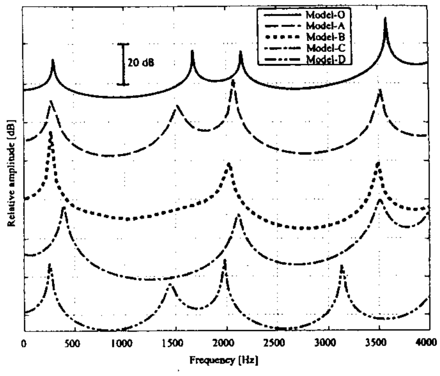

Figure 5: Vocal tract transfer functions of Model-O, Model-A, Model-B, Model-C and Model-D

5.2 Vocal tract transfer function of deformed models

The vocal tract transfer functions of Model-O, Model-A, Model-B, Model-C and Model-D are estimated by FEM. Figure 5 shows the transfer functions of the models. The numbers of peaks of Model-O and Model-A's transfer functions are the same and the frequencies correspond to the formant frequencies of speech sound. This result shows that bent center lines of the vocal tract models do not affect the transfer functions. On the other hand, there are three peaks of Model-B's transfer function; the peaks at approximately 1,500 Hz do not exist. This result corresponds to that of Model-1D. These results indicate that not only the cross-sectional-area function but also the cross-sectional shape of the vocal tract affects the transfer function of vocal tracts.

The peak at approximately 1,500 Hz of Model-C's transfer function does not exist, and that of Model-D's transfer function exists. These results indicate that the oral cavity shape of the subject produces F2 at approximately 1,500 Hz in the speech sound spectrum in uttering the /i/ sound.

5.3 Sound pressure and phase distributions

In order to investigate the difference in transfer function between Model-A and Model-B, sound pressure distributions estimated by FEM are observed. Figure 6 shows the sound pressure distributions of Model-A at the peak frequencies of the estimated vocal tract transfer function. Figure 7 shows the sound pressure distributions of Model-B at the peak frequencies of the estimated vocal tract transfer function. In the oral cavity of Model-A, the sound pressure contours are not composed of plane waves at 1,540 Hz. The sound pressures at the upper and lower surfaces of the oral cavity of Model-A is greater than that on the middle surface. The sound pressure distribution of Model-D in the oral cavity is similar to that of Model-A.

At each frequency, the sound pressure contours of Model-B are at a 90° to the center line of this model. This result

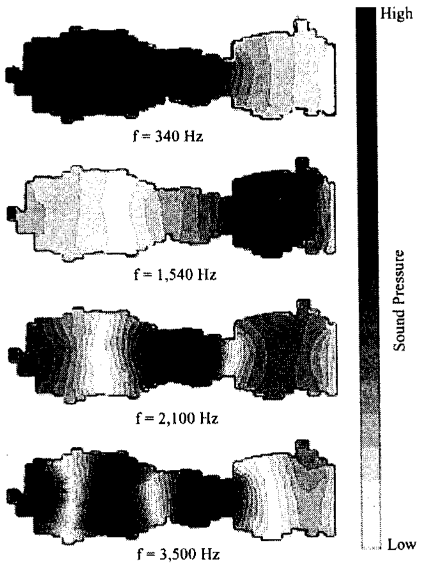

Figure 6: Sound pressure distributions of Model-A: Hemispheric volume is not displayed.

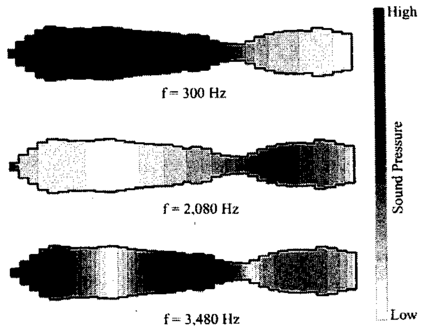

Figure 7: Sound pressure distributions of Model-B: Hemispheric volume is not displayed.

indicates that the sound pressure propagation of Model-B is by planar waves, which corresponds to that of Model-1D. The sound pressure contours of Model-C in the oral cavity are also composed of planar waves.

Figure 8 shows the phase distribution of Model-A at 1560 Hz. The same phase range in the oral cavity exists. From the sound pressure and phase distributions, there are no plane waves but only waves oscillating in the vertical direction in the oral cavity.

These results indicate that the oral cavity shapes of Model-A and Model-C affect vocal tract transfer functions.

6. Conclusion

To investigate the effects of vocal tract shape on vocal tract transfer functions, the vocal tract transfer functions and sound pressure distributions of deformed 3-D vocal tract models with complicated or simple shapes are estimated. The vocal tract transfer functions of the models with original oral cavity shapes have the peaks at approximately 1,500 Hz. On the contrary, those of the models with deformed oral cavity shapes as cylindrical tubes do not have peaks at such frequencies. The sound pressure distributions of Model-B are composed of planar waves. However the sound pressure distribution of Model-A at 1,540 Hz is not composed of planar waves in the oral cavity. These results indicate that the shape of the subject's oral cavity produces the peak of the vocal tract transfer function because the sound pressure contour is not planar.

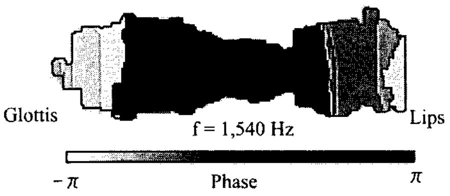

Figure 8: Phase distributions of Model-A: Hemispheric volume is not displayed.

These results indicate that the shape of the subject's oral cavity affects vocal tract transfer functions. Moreover, they suggest that the transfer functions of vocal tracts with asymmetrical shape are best estimated by FEM in the treatment of speech disorders.

References

[1] H. Nishimoto, M. Akagi, T. Kitamura and N. Suzuki: Estimation of transfer function of vocal tract extracted from MRI data by FEM, Proc. 18th ICA, Kyoto, Tu5. H. 1, 2004.

[2] 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., Vol. 17, No. 3, pp.163-166, 1996.

[3] M. K. Tiede and K. Honda: Improved area function estimation from midsagital distances on 3D MRI DATA, The Meeting of the Acoustical Society of Japan, 2-3-5, pp. 239-240, 1999.

[4] S. Imai and C. Furuichi: Unbiased estimation of log spectrum, IEICE Trans. A, Vol. J70-A, pp.471-480, 1987.

[5] M. M. Sondhi and J. Schroeter: A hybrid time-frequency domain articulatory speech synthesizer, IEEE Trans. on Acoustics, Speech and Signal Proc., Vol. 35, No. 7, pp.955-967, 1987.