Modeling of a speech production system based on MRI measurement of three-dimensional vocal tract shapes during fricative consonant phonation.

Takuya NIKAWA*1, Masafumi MATSUMURA*1, Takashi TACHIMURA*2, and Takeshi WADA*2

*1: Osaka Electro-Communication University, 18-8, Hatsu-cho, NEYAGAWA, 572-8530 JAPAN
E-mail: taku@amlab.osakac.ac.jp
*2: Division for Oral-Facial Disorders, Faculty of Dentistry, Osaka University, Yamada-Oka 1-8, SUITA, 565-0871 JAPAN

ABSTRACT

This study, based on the measurement of three-dimensional (3-D) vocal tract shapes during fricative consonant phonation, presents a realistic modeling of a human speech production system.

The 3-D shapes of a vocal tract and a dental crown were measured using Magnetic Resonance Imaging (MRI). A male subject was asked to produce the fricatives /s/ and /ʃ/ while wearing a dental crown plate that contained a contrast medium for MRI processing. 3-D MR images of the vocal tract for each sound were obtained while the subject’s tongue was kept still. The 3-D shapes and area functions of the vocal tract corresponding to respective sounds were computed using a gray level interpolation technique to form serial sections. The measured results suggest that there are individual differences in speech production and vocal tract shapes.

The airflow involved in the production of the fricatives /s/ and /ʃ/ was estimated in the 3-D vocal tract using the Finite Element Method (FEM). The shapes of the 3-D vocal tracts for the fricatives were reconstructed from the coronal MR images. The behavior of the airflow was determined from the vector diagram of the flow rate.

In this study, the vocal tract model with cascading circular tubes is called the VT model. A new acoustic model for the phonation of fricatives was proposed based on the VT model in which the sound source is a noise. Synthesized sounds of the Japanese fricatives /s/ and /ʃ/ were generated using this model. An auditory test demonstrated that the generated sounds were intelligible.

1. INTRODUCTION

Three-dimensional (3-D) data of vocal tract shapes are important for the construction of an articulatory model. Several techniques for investigating human speech production have been reported in speech science literature focusing on the vocal tract as a fundamental articulatory organ. In addition, attempts have been made to model the vocal tract and estimate its characteristics.

Magnetic Resonance Imaging (MRI) is widely acknowledged for its utility in investigating the geometry of the 3-D vocal tract shape, particularly because it does not involve any known radiation risks[1][2]. However, most previous MRI studies have been limited to the investigation of vowels. This is due to the difficulty of obtaining profiles of dental crown shapes that contain a small amount of water by conventional MRI techniques. Thus, there is insufficient data on 3-D shapes of the vocal tract during the phonation of consonants that are produced using the teeth in a steady state.

In order to resolve this problem, we have developed a method that uses a dental crown plate to enable simultaneous MR imaging of the dental crown and the vocal tract[3].

The purpose of the present study is to apply this method, the Finite Element Method (FEM), to estimation of air flow in the 3-D vocal tract during phonation of the fricative consonants /s/ and /ʃ/. We describe a method by which to reconstruct 3-D vocal tract shapes from 3-D MR images, and demonstrate an FEM analysis of the 3-D vocal tract model. Airflow in the 3-D vocal tract models is also illustrated.

In this study, we suggest that the noise source generates noise at the places where the flow rate is the fastest. In the final section, a new acoustic model for phonation of fricative consonants, based on the vocal tract model with cascading circular tubes in which the sound source is a noise, is proposed. Synthesized sounds of the Japanese fricatives /s/ and /ʃ/ are generated using this model, and their respective sound spectra are computed.

2. METHODS


A superconductive MR system (1.0T, MAGNEX100HP, Shimadzu Corp., Japan)[3] was used for MRI measurement in this study.

The mid-sagittal MR images were measured by the single-slice flip-back spin-echo imaging method at a high speed imaging. Each image was acquired with the repetition time TR=200 ms and the echo time TE=15 ms using an image matrix of 256×256 over a 25 cm field of view. The section thickness of the excited plane was 5 mm. The measurement time was 25 s. The coronal MR images were measured by a multi-slice T1-weighted spin-echo imaging method and were acquired with TR=200 ms and TE=15 ms using...
an image matrix of 256×256 over a 25 cm field of view.

3-D MR images consisting of MR images of 32 individual coronal sections, at intervals of 4 mm, from the tip of the nose to the atlas, were obtained. The measurement time was 142 s. The section thickness of the excited plane was 3.5 mm.

Data on dental crown shapes are required in order to analyze speech production and precisely estimate the acoustical characteristics of the vocal tract. In this study, these data were obtained by means of a dental crown plate, 0.6 mm thick, and containing a contrast medium for MR imaging, which was tightly attached to the subject’s dental crown by thermoforming. The subject was asked to produce fricatives while wearing a dental crown plate; during pronunciation, the subject’s tongue was kept still.

The experiments were performed using a Japanese adult male subject. MR images and the sounds uttered by the subject while in a supine position in the MR chamber were obtained. The subject’s voice was recorded using a high sensitivity condenser microphone.

2.2. 3-D vocal tract shape.

In this study, the 3-D vocal tract shapes during phonation of the fricative consonants /s/ and /ʃ/ were obtained from profiles of their sagittal sections.

First, sagittal MR images, taken at intervals of 4 mm, were estimated from coronal MR images formed using grey level interpolation. Second, air-tissue boundaries of the sagittal sections were obtained from each MR image by means of a threshold operation in which the threshold value was taken as the average of the grey level values at the air-tissue border points. Adjacent air-tissue boundaries were connected by spline interpolation, because the 3-D vocal tract shapes are constructed by a cascade connection of the air-tissue boundaries.

Figure 1 shows the 3-D vocal tract shapes during fricative consonant /s/ and /ʃ/ phonation. The shapes show a narrow oral cavity formed by the upper and lower teeth in conjunction with the frontal tongue body during fricative consonant phonation.

2.3. FEM analysis of the vocal tract.

When FEM is used for fluid analysis, the analytic space is approximated as the aggregate of several elements, and the differential equation corresponding to the fluid equation is solved. From the law of conservation of mass we obtain the continuity equation:

$$\frac{\partial \rho \chi}{\partial t} + \frac{\partial (\rho \chi)}{\partial x} + \frac{\partial (\rho \chi)}{\partial y} + \frac{\partial (\rho \chi)}{\partial z} = 0$$  \hspace{1cm} (1)

Generally, Navier-Stokes' equation considering of external force, pressure and friction is given in the fluid equation. In this study, the internal area of the vocal tract is supposed to be ideal fluid:

$$\frac{\partial \chi}{\partial t} + \chi \frac{\partial \chi}{\partial x} + \chi \frac{\partial \chi}{\partial y} + \chi \frac{\partial \chi}{\partial z}$$

$$= F_x - \frac{1}{\rho} \frac{\partial P}{\partial x} + \chi \left( \frac{\partial^2 \chi}{\partial x^2} + \frac{\partial^2 \chi}{\partial y^2} + \frac{\partial^2 \chi}{\partial z^2} \right)$$ \hspace{1cm} (2-1)

$$\frac{\partial \chi}{\partial t} + \chi \frac{\partial \chi}{\partial y} + \chi \frac{\partial \chi}{\partial y} + \chi \frac{\partial \chi}{\partial z}$$

$$= F_y - \frac{1}{\rho} \frac{\partial P}{\partial y} + \chi \left( \frac{\partial^2 \chi}{\partial x^2} + \frac{\partial^2 \chi}{\partial y^2} + \frac{\partial^2 \chi}{\partial z^2} \right)$$ \hspace{1cm} (2-2)

$$\frac{\partial \chi}{\partial t} + \chi \frac{\partial \chi}{\partial z} + \chi \frac{\partial \chi}{\partial z} + \chi \frac{\partial \chi}{\partial z}$$

$$= F_z - \frac{1}{\rho} \frac{\partial P}{\partial z} + \chi \left( \frac{\partial^2 \chi}{\partial x^2} + \frac{\partial^2 \chi}{\partial y^2} + \frac{\partial^2 \chi}{\partial z^2} \right)$$ \hspace{1cm} (2-3)
where $F$ is external force, $\rho$ is density, $\nu$ is dynamic viscosity, $t$ is time, $(V_x, V_y, V_z)$ are the components of the velocity vector in the $x$, $y$ and $z$ directions, respectively, and $P$ is internal stress. Our FEM formulation is based on the above equations.

The 3-D vocal tract shape with a radiational space is divided into a tetrahedral elements (/s/: 28,686 elements, 7,010 nodes, /ʃ/: 31,681 elements, 7920 nodes). Figure 2 shows a 3-D FEM vocal tract model during fricative consonant /s/ phonation. The air is supposed to flow into the vocal tract vertical to the glottis surface.

3. RESULTS

3.1. Airflow in the vocal tract.

Figure 3 shows air flow in the vocal tract during phonation of the fricative consonants /s/ and /ʃ/.

As can be seen, the flow rate was high at the narrow space made between the upper central incisors and the tongue surface. The distances from the lips were 23 mm (/s/) and 44 mm (/ʃ/) at the points where the velocity was highest. Generally, a fricative consonant is produced when the vocal tract is sufficiently constricted somewhere along its length; a noise is produced when air is forced through the constriction. Our results suggest that noise is generated at the site of an abrupt obstacle such as the teeth, to the airflow.

3.2. Electric circuit models for phonation of the fricative consonants /s/ and /ʃ/.

The findings in this study suggest that the noise source generates noise at the place where the flow rate is the fastest. In this study, a new acoustic model for phonation of the fricative consonants /s/ and /ʃ/ was proposed based on the vocal tract model (VT model) with cascading circular tubes in which the sound source is a noise. Figure 4 shows the electric circuit model for fricative consonant phonation. The circuit parameters were calculated from the vocal tract area function (Figure 5) estimated from the 3-D vocal tract shape. Synthesized sounds of the Japanese fricative consonants /s/ and /ʃ/ were generated using this model. An auditory test demonstrated that the generated sounds were intelligible.

Figure 6 shows a comparison between the spectra of the /s/ and /ʃ/ uttered by the subject and the computed results from the electric circuit models. The spectra show highly similar characteristics. However, it is necessary to consider the characteristics of the noise source fully in order to construct a precise model.

4. CONCLUSIONS

This study was performed in order to construct a new model for fricative consonant phonation. It is based on the MRI measurement of 3-D vocal tract shapes during phonation of the fricative consonants /s/ and /ʃ/. Air flow in the 3-D vocal tract was estimated using FEM. The 3-D FEM vocal tract models for fricative consonant phonation was obtained from coronal MR images. Based on these data, the location of the noise source was estimated, then the electric circuit model was constructed based on the estimated
location. An auditory test demonstrated that the generated sounds were intelligible. Finally, we compared the spectra of /s/ and /ʃ/ uttered by the subject with the computed results from the electric circuit model. The spectra show highly similar characteristics. We conclude that this method is potentially applicable for other fricative consonants.

**5. REFERENCES**


