**TECHNICAL REPORT**

# Transfer functions of solid vocal-tract models constructed from ATR MRI database of Japanese vowel production

Tatsuya Kitamura[1,*], Hironori Takemoto[2], Seiji Adachi[3] and Kiyoshi Honda[2]

[1]*Faculty of Intelligence and Informatics, Konan University,*
*8–9–1, Okamoto, Higashinada-ku, Kobe, 658–8501 Japan*
[2]*ATR Cognitive Information Science Laboratories,*
*2–2–2, Hikaridai, Keihanna Science City, Kyoto, 619–0288 Japan*
[3]*Fraunhofer Institute for Building Physics,*
*Nobelstrasse 12, 70569 Stuttgart, Germany*

(*Received 1 September 2008, Accepted for publication 3 February 2009*)

**Abstract:** The ATR MRI database of Japanese vowel production was used to evaluate the acoustic characteristics of the vocal tract for the five Japanese vowels through the measurements of frequency responses from solid vocal-tract models formed by a stereolithographic technique. The database includes speech sounds as well as volumetric magnetic resonance imaging (MRI) data, but the speech sounds were recorded separately from the acquisition MRI data; therefore, their speech spectra are not appropriate for use as the reference for the transfer functions of the vocal tract. A time-stretched pulse signal generated from a horn driver unit was introduced into the physical model at the lips, and the response signals of the models were recorded at the model's glottis. In the measurements, the glottis of the models was sealed with a plastic plate, and the response signals were measured from a small hole in the plate using a probe microphone. This method permits accurate measurement of the transfer functions of the vocal tract under a closed-glottis condition. The resulting transfer functions of the five Japanese vowels provide a benchmark for testing numerical analysis methods that have been used to study vocal-tract acoustics, although the solid wall decreases the frequencies of lower resonances.

**Keywords:** Vocal tract, Vowel, Transfer function, Solid models, Magnetic resonance imaging

**PACS number:** 43.70.Bk, 43.58.Gn [doi:10.1250/ast.30.288]

## 1. INTRODUCTION

Investigation of the relationship between the detailed shape and acoustic properties of the vocal tract is of importance in voice production studies. On the basis of recent advances in magnetic resonance imaging (MRI) for measuring the precise three-dimensional shape of the vocal cavities during speech production, the transfer functions not only for a vocal tract approximated by cylindrical tubes but also for a three-dimensional vocal tract have been investigated by numerical analysis methods such as the finite element method (FEM) and the finite-difference time-domain (FDTD) method; however, in these studies, the accuracy of the simulation was not thoroughly examined. One reason for this is the absence of a baseline for the acoustic properties of the vocal tract. In response, in the present study we measure the transfer functions of physical vocal-tract models constructed from the ATR MRI database of Japanese vowel production, which includes volumetric MRI data and speech data of the five Japanese vowels produced by a male native Japanese speaker. The speech sounds were recorded separately from the MRI scanning; therefore, their speech spectra are not appropriate for use as the reference for the transfer functions of the vocal tracts. The solid vocal-tract models were formed by a stereolithographic technique with a higher degree of accuracy than the spatial resolution of MRI, and their acoustic characteristics are thus sufficiently reliable for examining the accuracy of the simulation. The solid models can also provide the transfer functions of the vocal tract free from the glottal source characteristics, which cannot be separated from speech spectra. In addition, the database has been released with the aim of providing a common research base for speech science. Therefore, the acoustic characteristics of the models can provide a benchmark for testing numerical analysis methods.

[*]e-mail: t-kitamu@konan-u.ac.jp

The transfer functions of the vocal tract can also be measured directly by external excitation techniques. Van Den Berg [1] reported those of a hemilaryngectomized subject for eleven vowels. He excited the vocal tract of the subject at the opening of the throat using a throat loudspeaker and measured the response at the lips using a microphone. He succeeded in determining the vocal-tract transfer functions, but it is obvious that this method cannot be used for subjects with a normal larynx. Fujimura and Lindqvist [2] and Djeradi *et al.* [3] measured the acoustic characteristics of the vocal cavities by transdermic excitation of the vocal tract through the laryngeal wall using a vibrator and loudspeaker, respectively. On the basis of another theoretical concept, Sondhi and Gopinath [4] devised a method for measuring the vocal-tract area function from the impulse response of the vocal tract measured at the lips. Using these methods, the acoustic characteristics of the vocal tract including the acoustical effects of the losses in the cavities and the radiation at the lips can be determined; however, the measurements have low reproducibility because there is inter-repetition variation in the vocal-tract shape even for a subject producing the same vowel. In addition, the methods cannot be used to determine the relationship between the transfer function and the actual shape of the vocal tract. For the same reasons, speech spectra are not a satisfactory benchmark for the numerical simulation of vocal-tract acoustics. In the present study, we therefore measure the acoustic properties of the physical models formed from the given MRI data.

In this paper, we first describe the formation of the solid vocal-tract models and the experimental procedures used in the present study. We then give the transfer functions of the models and offer concluding remarks.

## 2. METHOD

### 2.1. Solid Vocal-Tract Models

Solid vocal-tract models of the five Japanese vowels, /a/, /e/, /i/, /o/, and /u/, were constructed from the ATR MRI database of Japanese vowel production. The database includes MRI data and speech data of the five Japanese vowels produced by a male native Japanese speaker. The subject had a Tokyo dialect, produced the vowel /u/ as a midvowel /ɯ/, and had a nasalized vowel /a/.

In the measurement of the MRI data, we used phonation synchronized scanning [5,6] and bone-conducting stimulus presentation [7] to acquire high-quality MRI data from the subject. In these methods, the subject is presented with a harmonic complex tone with a cyclic sequence (four harmonic complex tones in a 3 s cycle) using a piezoelectric bone-conducting speaker during scanning. The fundamental frequency of the harmonic complex tone is 120 Hz, which is close to the mean pitch frequency of the subject's voice so that he can easily keep the pitch frequency constant during the data acquisition. The subject repeatedly breathes in at the first harmonic complex tone and phonates during the following three harmonic complex tones in exact timing with the sequence. This technique allows scanning only during production to avoid motion artifacts due to inhalation. The technique also makes it easy for subjects to sustain a steady pitch frequency of speech. Although this technique has the above advantages, it imposes a greater strain on the subject; in the measurement, the subject was asked to repeat steady phonation 64 times for each vowel, which takes approximately seven minutes.

During the experiments, the subject wore earplugs to monitor his own voice by bone conduction feedback during the intense scan noise. This makes it possible for the subject to avoid the distortion of vowel gestures due to exposure to the noise. The bone-conducting presentation of the harmonic complex tone is to allow the subject to listen to the sound while wearing earplugs.

The MRI data were obtained using a Shimadzu-Marconi MAGNEX ECLIPSE 1.5 T Power Drive 250 system at the ATR-Promotions Brain Activity Imaging Center. An atlas array coil was used for acquiring MRI data of the subject's head and neck regions. The imaging sequence was a sagittal fast spin echo series with 2.0 mm slice thickness, no slice gap, no averaging, a $256 \times 256$ mm$^2$ field of view, a $512 \times 512$ pixel image size, 51 slices, a 90° flip angle, 11 ms echo time, and 3,000 ms repetition time.

The MRI data were converted into volume data with a $0.5 \times 0.5 \times 0.5$ mm$^3$ voxel resolution by linear interpolation. The volume data of the upper and lower jaws were then superimposed on the original volume data of each vowel by the method of Takemoto *et al.* [8]. The vocal tract region from the glottis to the lips was next extracted from the volume data by a thresholding method together with manual editing of the detailed boundaries. The velopharyngeal port was assumed to be closed in the models, even though a small opening was observed in the case of the vowel /a/ for our subject; the nasal cavities were therefore excluded from the vocal-tract region.

The extracted vocal-tract region was converted into STL file format by a three-dimensional image editor, Materialise MIMICS, and was smoothed using an STL editor, Materialise MAGICS. A 3 mm wall was then extruded outside the vowel-tract region, and the face region was trimmed except for the upper and lower lip regions (the height of each lip region is approximately 30 mm). Lastly, solid vocal-tract models as shown in Fig. 1 were formed with 0.2 mm accuracy by a stereolithographic technique. The density of the epoxy resin (D-Mec Ltd., SCR735) used for the solid models is 1.19 g/cm$^3$ and the speed of sound in the material is 2,352 m/s. The former

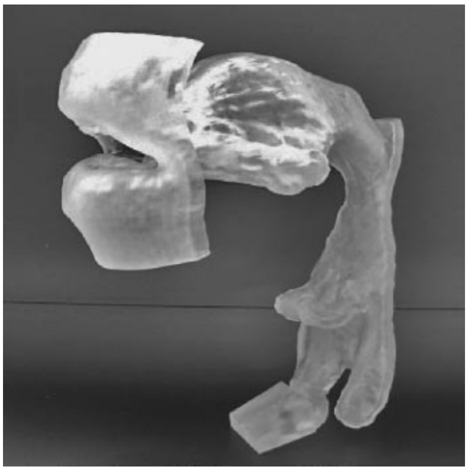

**Fig. 1** Solid vocal-tract models of a male subject producing the Japanese vowel /a/.

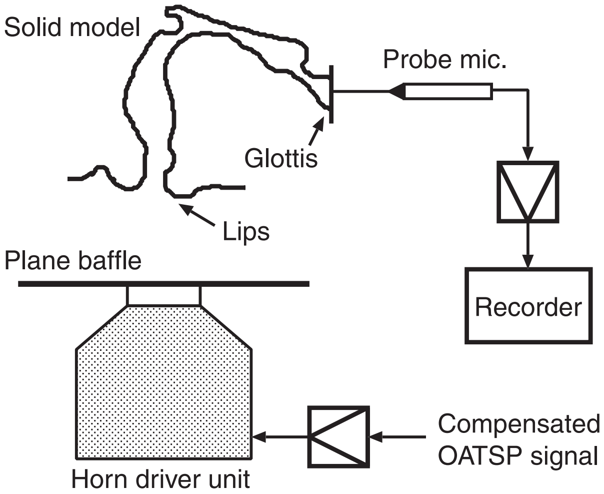

**Fig. 2** Diagram of measurement setup.

was measured at a temperature of 20.0°C at the Industrial Research Institute of Ishikawa and the latter was measured at a temperature of 25.5°C at Hyogo Prefectural Institute of Technology. Upon excitation by Rosenberg glottal waves [9] with a suitable pitch frequency for our subject, the solid vocal-tract models produce clear lifelike vowels similar to those uttered by the subject. This clearly suggests that the models precisely replicate the detailed shape of the vocal tract of the subject, and that the shape of the vocal tract is one of the sources of the speaker characteristics.

### 2.2. Measurement Method

Figure 2 shows a diagram of the measurement setup. The lips of the solid model faced the radiating end of a horn driver unit (ALE Acoustic, 7550DE) with a 400 mm × 400 mm plane baffle. The distance between the lip end of the solid model and the plane baffle is 200 mm. The measurement was carried out in an anechoic chamber while keeping the temperature at 25°C. The glottis of the solid models was sealed with a plastic plate with a 1.2 mm hole in the center (B&K UA0929), and the tip of the probe tube of a microphone (B&K 4182) was inserted into the hole. This increases the glottal impedance and suppresses the acoustic effects of glottal opening on the acoustic characteristics of the vocal tract, and thus the laryngeal cavity resonance, which disappears in open-glottis periods during vowel production, can be measured [10–12].

In the measurement of the acoustic characteristics of the solid vocal-tract models, a 0.34 ms optimized Aoshima's time-stretched pulse (OATSP) signal [13] $p_s(n)$ was introduced into the solid model at the lips from the horn driver unit via an amplifier (Accuphase E-305). Here, $n$ is the discrete time index. The sound pressure at the glottis $p_g(n)$ was recorded at a sampling rate of 48 kHz with 16-bit resolution using the probe microphone, a microphone amplifier (B&K 5935), and a solid-state recorder (Marantz PMD-670). Ten responses to the OATSP signal $p_g(n)$ were synchronously averaged in the time domain, and the impulse response of the solid models $h_m(n)$ was then calculated. $h_m(n)$ is obtained by convolving the time-averaged response $\bar{p}_g(n)$ with the time-reversed OATSP signal $p_s(N-n)$,

$$h_m(n) = \bar{p}_g(n) * p_s(N-n), \qquad (1)$$

where $N$ is the length of $p_s(n)$. The impulse response in the frequency domain, or the transfer function, $H_m(\omega)$ is the result of the discrete Fourier transform (DFT) of $h_m(n)$. The transfer function $H_m(\omega)$ measured in this study is $[1 + Z_r(\omega)/Z_0]p_s(\omega)$ times greater than the volume velocity transfer function of the models $H(\omega)$, where $p_s(\omega)$ is the sound pressure of the OATSP signal, $Z_r(\omega)$ is the radiation impedance of the solid models, and $Z_0$ is the characteristic impedance of a plane wave. Note that the resonance and antiresonance frequencies are constant for the transfer functions $H_m(\omega)$ and $H(\omega)$ (see Appendix A for details).

Prior to the measurement, we compensated the impulse response of the measurement setup by carrying out inverse filtering. The impulse response of the measurement system without a solid model $h_s(n)$ was first measured by the method described above. The transfer function of the measurement system $H_s(\omega)$ was next calculated by a fast Fourier transform (FFT), and the frequency response of the inverse filter of the measurement setup is obtained by

$$H_s^{-1}(\omega) = \frac{1}{H_s(\omega)}. \qquad (2)$$

The coefficients of the finite impulse response (FIR) filter of $H_s^{-1}(\omega)$ were next calculated using the frequency sampling method [14]. The number of taps in the inverse filter is 1,024 in this study. A 0.34 ms OATSP signal was then filtered with the inverse filter $H_s^{-1}(\omega)$. We compensated the impulse response of the measurement setup by

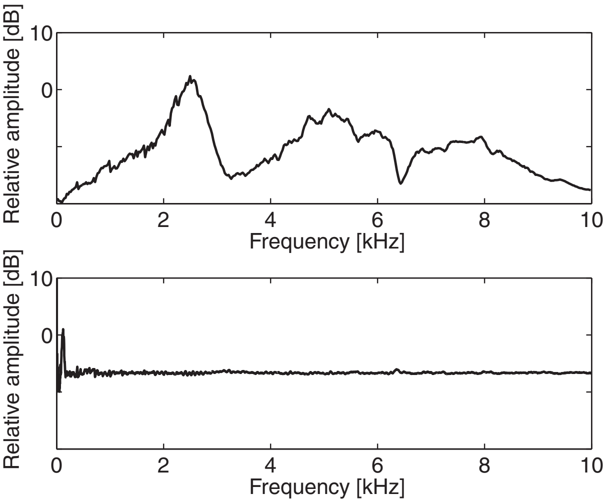

**Fig. 3** Impulse response of the measurement setup. The upper and lower panels show the original and compensated responses, respectively.

inputting the resultant signal to the measurement setup. The original and compensated impulse responses of the measurement setup are shown in Fig. 3. Because of distortion below 160 Hz in the compensated impulse response, we will show the acoustic characteristics of the solid models above 160 Hz, where the impulse response is relatively flat.

## 3. RESULTS AND DISCUSSION

Figure 4 depicts the acoustic characteristics of the solid vocal-tract models, showing that the method permits measurement not only of the resonance peaks but also of the antiresonance dips of the acoustic characteristics. The frequencies and bandwidth of the first four resonances are listed in Tables 1 and 2, respectively. The resonance frequencies were identified by a peak-picking method. The first resonance frequency of the solid model for the vowel /i/ could not be determined because of the distortion of the measurement setup below 160 Hz.

The resonance frequencies in Table 1 are different from the average formant frequencies of male speakers. There are a couple of possible reasons for the difference. The main reason is probably that the shapes of the vocal tract were distorted during the measurement. Table 3 lists the formant frequencies of the vowels obtained during the MRI data acquisition. The speech sound was recorded by an optical microphone (Phone-Or SOM) and a solid-state recorder (Marantz PMD-670) at a sampling rate of 44.1 kHz with 16-bit resolution. One of the 64 utterances that was not lost into MRI scanning noise was selected for each vowel, and was downsampled to 12 kHz. Note that this speech data are not those in the database. Log spectral envelopes of the vowel sounds were then calculated by the unbiased log spectral estimation [15], and averaged with respect to voiced frames. The frame length was 32 ms, the frame period was 16 ms, the order of the cepstrum was 60, and the number of iterations was 3. The formants were then identified by a peak-picking method. Table 4 shows the formant frequencies of the five Japanese vowels in the database, which were recorded in a supine posture. The method for identifying the formant frequencies is the same as that described above. The differences in formant frequencies in Tables 3 and 4 suggest that the articulatory gesture of the subject was deformed in the experiment, probably because the subject needed to keep the vocal-tract shape steady during the measurement (approximately seven minutes for each vowel).

It is possible that the difference between the solid models and vocal tracts in terms of the hardness of their wall causes the differences between the resonance frequencies of the transfer functions (Table 1) and the formant frequencies of the speech sounds (Table 3). The yielding walls of acoustic tubes cause the resonance frequencies to shift to higher frequencies; this effect is most pronounced at low frequencies [16]. In Appendix B, we estimate the acoustic effects of a yielding wall on the transfer functions of the vocal-tract of the five Japanese vowels using a transmission line model. The differences between the resonance frequencies of the vocal-tract transfer functions calculated under rigid and yielding wall conditions suggest that the yielding wall causes an increase in the resonance frequency in the range from 0.2% to 54.4% depending on the frequency (see Appendix B for details). The acoustic characteristics shown in Fig. 4 thus cannot be compared directly with the vocal-tract transfer functions measured by the methods based on the external excitation of the vocal tract [1–4].

The laryngeal cavity resonance occurs during vowel production. The resonance appears under the closed-glottis condition and disappears under the open-glottis condition [10–12]. The lower formants are also damped when the glottis is open [12,16]. In this study, the glottal impedance of the models was thus set to be high in the measurement so that the closed-glottis condition is simulated to obtain the transfer functions with the laryngeal cavity resonance and undamped lower resonances. Kitamura *et al.* [17,18] attempted to determine the pressure-to-velocity transfer function for solid vocal-tract models, but the peak of the laryngeal cavity resonance was attenuated in the measured transfer functions. In their method, the solid vocal-tract models were connected to a driver unit with a uniform tube of 5 mm diameter, and the transfer functions were derived from the sound pressure at the output end of the models and the sound velocity at the input end. In this case, the glottal impedance of the models was not high, that is, the measurements were carried out under the open-glottis condition. This may be the reason why the laryngeal cavity resonance was not clearly observed in their previous study.

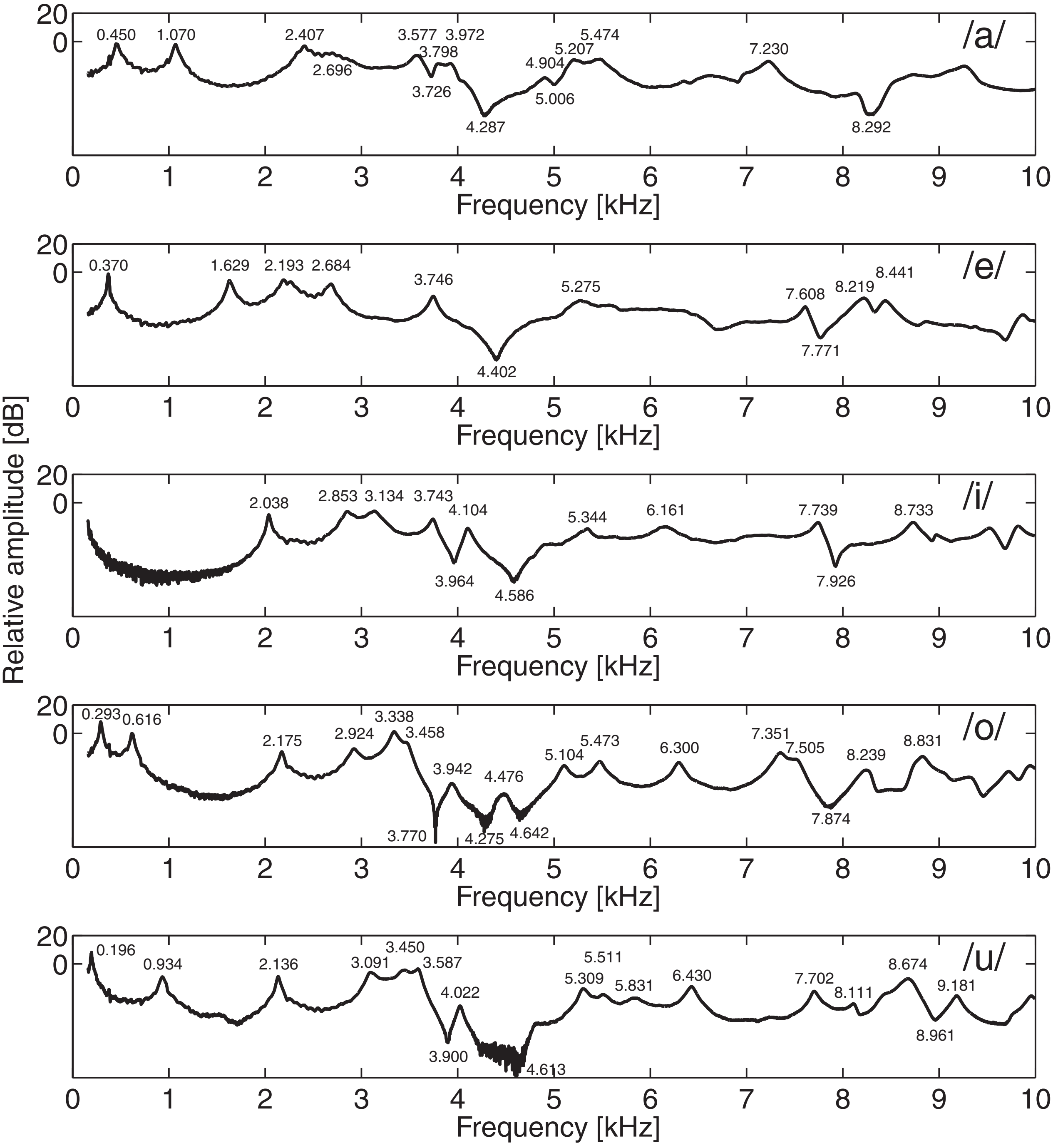

**Fig. 4** Transfer functions of solid vocal-tract models of the five Japanese vowels with numerical values representing the frequencies of peaks and dips in Hz. The panels represent, from top to bottom, the vowels /a/, /e/, /i/, /o/, and /u/.

**Table 1** The first to fourth resonance frequencies (R1–R4) of the transfer functions of the solid models (Fig. 4) in Hz.

| Vowel | R1 | R2 | R3 | R4 |
|---|---|---|---|---|
| /a/ | 450 | 1,070 | 2,407 | 2,696 |
| /e/ | 370 | 1,629 | 2,193 | 2,684 |
| /i/ | — | 2,038 | 2,853 | 3,134 |
| /o/ | 293 | 616 | 2,175 | 2,924 |
| /u/ | 196 | 934 | 2,136 | 3,091 |

**Table 2** The 3-dB-down bandwidths (B1–B4) of the first to fourth resonances of the transfer functions of the solid models (Fig. 4) in Hz.

| Vowel | B1 | B2 | B3 | B4 |
|---|---|---|---|---|
| /a/ | 43 | 41 | 72 | — |
| /e/ | 19 | 44 | 129 | 82 |
| /i/ | — | 29 | 114 | 132 |
| /o/ | 19 | 31 | 40 | 83 |
| /u/ | 16 | 50 | 32 | 104 |

**Table 3** The first to fourth formant frequencies (F1–F4) of the five Japanese vowels recorded during the measurement of the MRI data.

| vowel | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| /a/ | 516 | 1,055 | 2,250 | 2,648 |
| /e/ | 375 | 1,641 | 2,320 | 3,094 |
| /i/ | 234 | 2,039 | 2,672 | 2,977 |
| /o/ | 375 | 656 | 2,227 | 2,977 |
| /u/ | 304 | 984 | 2,203 | 3,023 |

**Table 4** The first to fourth formant frequencies (F1–F4) of the five Japanese vowels in the ATR MRI database of Japanese vowel production. The speech sounds were recorded separately from the measurement of the MRI data.

| Vowel | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| /a/ | 633 | 1,078 | 2,672 | 3,047 |
| /e/ | 375 | 1,711 | 2,391 | 3,047 |
| /i/ | 211 | 2,039 | — | 3,070 |
| /o/ | 398 | 750 | 2,414 | 2,883 |
| /u/ | 234 | 1,266 | 2,133 | 3,000 |

Although we introduced the OATSP signal into the solid models from the lips and measured the responses at the glottis in the present study, the response can also be measured at the lips if the input signal is introduced from the glottis by setting the glottal impedance to be high. Honda *et al.* [10] examined the acoustic effects of the glottal area and piriform fossa of solid vocal-tract models by this method. They demonstrated that the transfer functions of the solid models under the closed-glottis condition can be determined by setting the glottal area to be small (radius of the glottal area $r = 0.75$ mm), that is, by setting the glottal impedance to be high.

Fujita and Honda [19] carried out a vowel synthesis experiment using solid vocal-tract models. They mounted the models on a horn driver unit, introduced Rosenberg waves from the glottis, and measured the spectra of the resulting vowel sounds. They also showed that modification of the shape of the laryngeal cavity does not have a significant effect on the spectra, although this is possibly due to the insufficient glottal impedance. In their experiments, the horn driver unit's throat was filled with a sponge, a thin metal plate with multiple holes was placed at the throat opening, and the diameter of the glottal area was set to 3 mm. Considering that the laryngeal cavity resonance is generated when the glottis is closed, the result may have been caused by the small glottal impedance in their experiments.

## 4. CONCLUSION

In this study we evaluated the acoustic characteristics of solid vocal-tract models constructed from the ATR MRI database of Japanese vowel production. The results provide a benchmark for testing the numerical analysis methods used to study vocal-tract acoustics and for evaluating the methods of extracting the vocal-tract cross-sectional area function, because the database is available to the public for speech science research.

It should be noted, however, that the obtained functions are different from the actual transfer functions of the human vocal tract in the following ways: (1) the models have a solid wall, (2) the radiation characteristics at the lips of the models are different from those of humans owing to the lack of facial surface of the models, and (3) the velopharyngeal port of the solid model for the vowel /a/ was closed, whereas the subject opened the port during the production of the vowel. Taking such differences into account is crucial when applying the results of the present study.

## ACKNOWLEDGEMENTS

This research was partly supported by SCOPE (071705001) of the Ministry of Internal Affairs and Communications, Japan and Kakenhi (18700193).

The MRI data used in this study are from the ATR MRI database of Japanese vowel production, and were accumulated at and released by ATR Human Information Science Laboratories under the project "Research on Human Communication" funded by the National Institute of Information and Communications Technology, Japan. The use of the database and the release of results are under a licensed agreement with ATR-Promotions Co., Ltd.

The authors wish to thank Drs. Shinobu Masaki and Hironori Nishimoto from ATR-Promotions, Dr. Satoru Fujita from AI Inc., and Dr. Yutaka Ichinose from Kyushu Sangyo University for their helpful comments.

## REFERENCES

[1] JW. Van Den Berg, "Transmission of the vocal cavities," *J. Acoust. Soc. Am.*, **27**, 161–168 (1953).

[2] O. Fujimura and J. Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," *J. Acoust. Soc. Am.*, **49**, 541–558 (1971).

[3] A. Djeradi, B. Guérin, P. Badin and P. Perrir, "Measurement of the acoustic transfer function of the vocal tract: A fast and accurate method," *J. Phonet.*, **19**, 387–395 (1991).

[4] M. M. Sondhi and B. Gopinath, "Determination of vocal-tract shape from impulse response at the lips," *J. Acoust. Soc. Am.*, **49**, 1867–1873 (1971).

[5] S. Masaki, M. Tiede and K. Honda, "MRI-based speech production study using a synchronized sampling method," *J.*

*Acoust. Soc. Jpn.* (*E*), **20**, 375–379 (1999).

[6] S. Takano, K. Honda and K. Kinoshita, "Measurement of cricothyroid articulation using high-resolution MRI and 3D pattern matching," *Acta Acustica united with Acustica*, **92**, 725–730(6) (2006).

[7] Y. Nota, T. Kitamura, H. Takemoto, H. Hirata, K. Honda, Y. Shimada, I. Fujimoto, Y. Syakudo and S. Masaki, "A bone-conduction system for auditory stimulation in MRI," *Acoust. Sci. & Tech.*, **28**, 33–38 (2007).

[8] H. Takemoto, T. Kitamura, H. Nishimoto and K. Honda, "A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions," *Acoust. Sci. & Tech.*, **25**, 468–474 (2004).

[9] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Am.*, **49**, 583–590 (1971).

[10] K. Honda, T. Kitamura, H. Takemoto, Y. Nota, H. Hirata, S. Masaki and S. Fujita, "Solid modeling and acoustic analysis of vocal tract based on vowel production MRI data," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 313–314 (2005).

[11] H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari and K. Honda, "Acoustic roles of the laryngeal cavity in vocal tract resonance," *J. Acoust. Soc. Am.*, **120**, 2228–2238 (2006).

[12] T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari and K. Honda, "Cyclicity of laryngeal cavity resonance due to vocal fold vibration," *J. Acoust. Soc. Am.*, **120**, 2239–2249 (2006).

[13] Y. Suzuki, F. Asano, H.-Y. Kim and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, **97**, 1119–1123 (1995).

[14] L. R. Rabiner, C. A. McGonegal and D. Paul, "FIR windowed filter design program - WINDOW," in *Programs for Digital Signal Processing* (IEEE Press, New York, 1979), Sections 5.2, 5.2-1–5.2-19.

[15] S. Imai and C. Furuichi, "Unbiased estimator of log spectrum and its application to speech signal processing," *Trans. IEICE*, **J70-A**, 471–480 (1987).

[16] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals* (Prentice Hall, Englewood Cliffs, N. J., 1978), Chap. 3, pp. 38–115.

[17] T. Kitamura, H. Nishimoto, S. Fujita and K. Honda, "Comparison of measured and simulated transfer functions of vocal tract model," *Tech. Rep. IEICE* (*SP*), **103**, 37–42 (2003).

[18] T. Kitamura, S. Fujita, K. Honda and H. Nishimoto, "An experimental method for measuring transfer functions of acoustic tubes," *Proc. ICSLP 2004*, TuB602 (2004).

[19] S. Fujita and K. Honda, "An experimental study of acoustic characteristics of hypopharyngeal cavities using vocal tract solid models," *Acoust. Sci. & Tech.*, **26**, 353–357 (2004).

[20] A. D. Pierce, *Acoustics—An Introduction to Its Physical Principles and Applications* (Originally published in 1981. Reprinted by Acoust. Soc. Am., New York, 1989).

[21] S. Adachi and M. Yamada, "An acoustical study of sound production in biphonic singing Xöömij," *J. Acoust. Soc. Am.*, **105**, 2920–2932 (1999).

[22] H. Takemoto, K. Honda, S. Masaki, Y. Shimada and I. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *J. Acoust. Soc. Am.*, **119**, 1037–1049 (2006).

[23] R. Caussé, J. Kergomard and X. Lurton, "Input impedance of brass musical instruments—Comparison between experiment and numerical models," *J. Acoust. Soc. Am.*, **75**, 241–254 (1984).

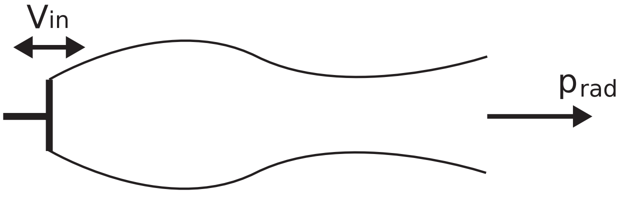

**Fig. 5** Acoustic tube excited by a piston at the input end.

## Appendix A: Relationship between the Measured Transfer Function and the Volume Velocity Transfer Function

Let us assume a linear response of an acoustic tube such as that of the vocal-tract models fabricated in this study. In the frequency domain, the input pressure $P_{\mathrm{in}}(\omega)$ and volume velocity $V_{\mathrm{in}}(\omega)$ are related to the output pressure $P_{\mathrm{out}}(\omega)$ and volume velocity $V_{\mathrm{out}}(\omega)$ as

$$\begin{bmatrix} P_{\mathrm{in}}(\omega) \\ V_{\mathrm{in}}(\omega) \end{bmatrix} = \begin{bmatrix} m_{11}(\omega), & m_{12}(\omega) \\ m_{21}(\omega), & m_{22}(\omega) \end{bmatrix} \begin{bmatrix} P_{\mathrm{out}}(\omega) \\ V_{\mathrm{out}}(\omega) \end{bmatrix}, \tag{A·1}$$

where $M(\omega) = [m_{ij}(\omega)]$ is a $2 \times 2$ transmission matrix. By assuming the reciprocity [20] in this system, the determinant of $M(\omega)$ becomes unity.

Figure 5 shows that an acoustic tube is excited by a piston providing nonzero volume velocity $V_{\mathrm{in}}(\omega)$ at the input end. The volume velocity transfer function is defined as

$$H(\omega) = \frac{V_{\mathrm{out}}(\omega)}{V_{\mathrm{in}}(\omega)}. \tag{A·2}$$

At the output end, no external signal is introduced into the tube and only sound radiating from the output exists. In this case,

$$\frac{P_{\mathrm{out}}(\omega)}{V_{\mathrm{out}}(\omega)} = Z_r(\omega) \tag{A·3}$$

is satisfied, where $Z_r(\omega)$ is the radiation impedance. The volume velocity transfer function then becomes

$$H(\omega) = \frac{1}{m_{21}(\omega)Z_r(\omega) + m_{22}(\omega)}. \tag{A·4}$$

The measurements in the present study took place using the slightly different setup shown in Fig. 6. First, $V_{\mathrm{in}}(\omega)$ is zero when the input end is closed. From Eq. (A·1), we obtain

$$m_{21}(\omega)P_{\mathrm{out}}(\omega) + m_{22}(\omega)V_{\mathrm{out}}(\omega) = 0. \tag{A·5}$$

Second, the OATSP signal $p_s(\omega)$ is introduced into the tube to induce excitation. In this case, both $p_s(\omega)$ entering to the tube and the sound $p_{\mathrm{rad}}(\omega)$ radiating from the output exist outside the tube. These traveling pressure waves are accompanied by the volume velocity waves

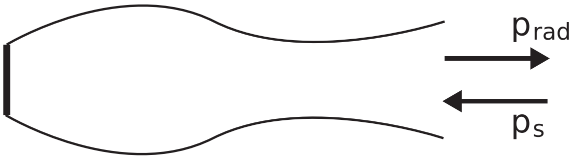

**Fig. 6** Closed acoustic tube with a sound $p_s$ introduced from the outside.

$-p_s(\omega)/Z_0$ and $p_{\rm rad}(\omega)/Z_r(\omega)$, respectively, where $Z_0$ is the characteristic impedance of a plane wave. Note here that $p_s(\omega)$ is a plane wave and the direction of motion is towards the tube. Because both the pressure and the volume velocity should be continuous at the output end, we have

$$P_{\rm out}(\omega) = p_{\rm rad}(\omega) + p_s(\omega), \tag{A·6}$$

$$V_{\rm out}(\omega) = \frac{p_{\rm rad}(\omega)}{Z_r(\omega)} - \frac{p_s(\omega)}{Z_0}. \tag{A·7}$$

Eliminating $p_{\rm rad}(\omega)$ from these equations, we obtain

$$P_{\rm out}(\omega) - Z_r(\omega)V_{\rm out}(\omega) = \left[1 + \frac{Z_r(\omega)}{Z_0}\right]p_s(\omega). \tag{A·8}$$

From Eqs. (A·5) and (A·8), we have

$$\begin{bmatrix} P_{\rm out}(\omega) \\ V_{\rm out}(\omega) \end{bmatrix} = \frac{\left[1 + Z_r(\omega)/Z_0\right]p_s(\omega)}{m_{21}(\omega)Z_r(\omega) + m_{22}(\omega)} \begin{bmatrix} m_{22}(\omega) \\ -m_{21}(\omega) \end{bmatrix}. \tag{A·9}$$

From Eqs. (A·1) and (A·9), $P_{\rm in}(\omega)$ becomes

$$P_{\rm in}(\omega) = \frac{\left[1 + Z_r(\omega)/Z_0\right]p_s(\omega)}{m_{21}(\omega)Z_r(\omega) + m_{22}(\omega)}, \tag{A·10}$$

where $\det M(\omega) = 1$ is used. This equation implies that the transfer function $P_{\rm in}(\omega)$ measured in this study is $[1 + Z_r(\omega)/Z_0]p_s(\omega)$ times greater than the volume velocity transfer function $H(\omega)$. Note that the resonance and antiresonance frequencies are constant for the two transfer functions, $H(\omega)$ and $P_{\rm in}(\omega)$.

## Appendix B: Acoustic Effects of Yielding Wall on the Vocal-Tract Transfer Functions

The acoustic effects of a yielding wall on the transfer function of the vocal tract were estimated using a transmission line model [21]. In this simulation, we used the vocal-tract area functions of the five Japanese vowels measured from the MRI data in the database by Takemoto *et al.*'s method [22], and the vocal-tract transfer functions were calculated under rigid and yielding wall conditions.

The transfer functions were calculated for the frequencies up to 4 kHz by considering the radiation impedance at the mouth. The radiation impedance of the vocal tract $Z_R$ was approximated by the following equation proposed by Caussé *et al.* [23]:

$$\begin{aligned}\frac{Z_R}{\rho c} = \frac{z^2}{4} &+ 0.0127z^4 + 0.082z^4 \ln z - 0.023z^6 \\ &+ j(0.6133z - 0.036z^3 \\ &+ 0.034z^3 \ln z - 0.0187z^5),\end{aligned} \tag{B·1}$$

$$z = kr, \tag{B·2}$$

where $k$ is the wave number, $r$ is the radius of the open end, $\rho$ is the air density, and $c$ is the speed of sound.

In the transmission line model [21], the shunt admittance per unit length due to the yielding wall $Y_w$ is modeled by

$$Y_w = \frac{i\omega}{\rho c^2} \frac{\omega_0^2}{b + i\omega a - \omega^2}, \tag{B·3}$$

where $\omega_0 = 406\pi$ rad/s is the lowest angular resonance frequency of the tract when closed at both ends, $a = 130\pi$ rad/s is the ratio of the wall resistance to the mass, and $b = (30\pi)^2\,(\text{rad/s})^2$ is the square of the angular frequency of the mechanical resonance. We assumed $\rho = 1.12\,\text{kg/m}^3$ and $c = 352.9$ m/s. $Y_w$ for the rigid wall condition was set to zero and that for the yielding wall condition was set to the right side of Eq. (B·3) in the simulation.

**Table 5** The first to fourth resonance frequencies (R1–R4) of vocal-tract transfer functions with a rigid wall for the five Japanese vowels in Hz.

| Vowel | R1 | R2 | R3 | R4 |
|---|---|---|---|---|
| /a/ | 528 | 1,128 | 2,602 | 2,985 |
| /e/ | 411 | 1,806 | 2,544 | 3,132 |
| /i/ | 171 | 2,232 | 3,093 | 3,303 |
| /o/ | 351 | 722 | 2,352 | 3,182 |
| /u/ | 238 | 1,088 | 2,340 | 3,232 |

**Table 6** The first to fourth resonance frequencies (R1–R4) of vocal-tract transfer functions with a yielding wall for the five Japanese vowels in Hz.

| Vowel | R1 | R2 | R3 | R4 |
|---|---|---|---|---|
| /a/ | 566 | 1,146 | 2,610 | 2,991 |
| /e/ | 458 | 1,818 | 2,552 | 3,138 |
| /i/ | 264 | 2,241 | 3,100 | 3,309 |
| /o/ | 405 | 750 | 2,361 | 3,188 |
| /u/ | 311 | 1,107 | 2,349 | 3,239 |

Tables 5 and 6 list the first four resonance frequencies of the vocal-tract transfer functions estimated under the rigid and yielding wall conditions, respectively. The differences between the first resonance frequency for the two conditions are 38 Hz (7.2%) for /a/, 47 Hz (11.4%) for /e/, 93 Hz (54.4%) for /i/, 54 Hz (15.4%) for /o/, and 73 Hz (30.7%) for /u/. The differences between the second resonance frequency for the two conditions are 18 Hz

(1.6%) for /a/, 12 Hz (0.7%) for /e/, 9 Hz (0.4%) for /i/, 28 Hz (3.9%) for /o/, and 19 Hz (1.7%) for /u/. The differences between the third resonance frequency for the two conditions are 8 Hz (0.3%) for /a/, 8 Hz (0.3%) for /e/, 7 Hz (0.2%) for /i/, 10 Hz (0.4%) for /o/, and 9 Hz (0.4%) for /u/. The difference between the fourth resonance frequency for the two conditions are 6 Hz (0.2%) for /a/, 6 Hz (0.2%) for /e/, 5 Hz (0.2%) for /i/, 6 Hz (0.2%) for /o/, and 7 Hz (0.2%) for /u/. The results indicate that the yielding wall shifts the resonance frequencies towards higher frequencies and that the effect is most notable at low frequencies.