Acoust. Sci. & Tech. **44**, 1 (2023)

PAPER

One-dimensional static vocal tract model in the time domain

Hironori Takemoto[1,*], Seiji Adachi[2] and Natsuki Toda[1]

[1]*Chiba Institute of Technology,*
2–17–1 Tsudanuma, Narashino, Chiba, 275–0016 Japan
[2]*Tezukayama University,*
7–1–1 Tezukayama, Nara, 631–8501 Japan

(*Received 30 April 2022, Accepted for publication 1 September 2022*)

Abstract: The vocal tract can be modeled as an acoustic tube in the low-frequency region because the plane wave propagation is dominant. Further, it can be considered static for a limited short period during running speech, such as vowels. Thus, its acoustic properties have been examined mainly using the transmission line model (TLM), that is, the one-dimensional static model in the frequency domain. In the present paper, we propose a one-dimensional static model in the time domain based on the finite-difference time-domain method. In this model, the vocal tract is represented by the cascaded acoustic tubes of different cross-sectional areas. The pressure and wall vibration effects are simulated at the center of each tube. On the other hand, the volume velocity is calculated at the labial end. According to the leapfrog algorithm, the pressure and volume velocity are sequentially computed. As a result, the impulse responses of the vocal tracts for the five Japanese vowels were calculated, and the corresponding transfer functions agreed well with those calculated by the TLM in the low-frequency region. The mean absolute percentage difference of the lower four peaks for the five vowels was 2.3%.

Keywords: Vocal tract model, FDTD, Wall vibration, Vowel production, Time domain

1. INTRODUCTION

The vocal tract is an internal space in the human body from the glottis to the lips or nostrils. It consists of laryngeal, pharyngeal, oral, and nasal cavities, and has a complex three-dimensional shape that rapidly changes during speech.

In the low-frequency region, the vocal tract, excluding the nasal cavity, can be approximated by an acoustic tube; that is, it is a one-dimensional (1D) model because of the dominant plane wave propagation [1]. The shape of the 1D model is determined by the vocal tract area function, which represents the changes in the cross-sectional area of the vocal tract along the long axis from the glottis to the lips. The area is measured discretely at equal intervals to divide the vocal tract into a series of cascaded short acoustic tubes with different cross-sectional areas. The pressure and volume velocity corresponding to the voltage and current, respectively, can be calculated for each tube using an electric circuit equivalent to the tubes. Thus, the 1D model is referred to as the equivalent circuit model or transmission line model (TLM). The 1D model is useful for examining the acoustic properties of the vocal tract and synthesis of the speech sounds; however, the upper limit frequencies for which the 1D models without side branches, such as the piriform fossae, are valid remain debatable: 4 kHz estimated by Flanagan [2] and 5 kHz by Stevens [1].

Many 1D models simulate wave propagation in the static vocal tract in the frequency domain [e.g., 1–3]. The vocal tract can be considered static over short time periods during an utterance, such as vowels. The transfer function for a static vocal tract can be easily calculated using the TLM, which considers the losses caused by the thermal exchange, viscous resistance, and wall vibration on the vocal tract wall. Therefore, the acoustic properties of the static vocal tract have been examined using this model [1–3]. However, it is difficult to visualize the wave propagating process in the vocal tract using this model.

Several 1D models simulate wave propagation in the dynamic vocal tract in the time domain, such as the well-known Maeda's model [4] and wave-reflection model [5]. Although they are modeled as electric circuits, the area and length of each acoustic tube can be changed over time; some areas even become zero at the stop consonant. The effects of these changes must be calculated in addition to the pressure and volume velocity of the traveling waves

[*] e-mail: hironori.takemoto@p.chibakoudai.jp

[doi:10.1250/ast.44.9]

at each simulation step. Therefore, while these models are very complex and difficult to implement, they can synthesize continuous speech sounds, including consonants.

The purpose of the present study was to develop a simple 1D model that simulates wave propagation in the static vocal tract in the time domain to better understand vocal tract acoustics. This model can calculate the impulse response when the impulse is input and synthesize vowel sounds when the glottal waves are input. Thus, students can easily confirm that the signals obtained by convolving the impulse response into the glottal waves match those obtained by inputting the same glottal waves into the model, thereby improving the understanding of linear time-invariant systems. Furthermore, this model can visualize the wave propagating process, such as the resonance modes of the pressure or volume velocity in the vocal tract, when the sinusoidal waves with the resonance frequencies are input.

In the following sections, the governing equations considering the wall vibration effects are formulated first because these effects cause a non-negligible frequency shift of the first peak of the transfer function [1]. Then, the equations are discretized based on the 1D finite-difference time-domain (FDTD) method, and the calculation procedure is described with a pseudocode. Finally, the transfer functions calculated using the proposed model are compared with those calculated using the TLM to validate the proposed model.

2. FORMULATION

On the vocal tract wall, thermal exchange, viscous resistance, and wall vibration cause acoustic energy losses [1–3]. Consequently, the peaks of the transfer function reduce in level; that is, the bandwidths increase. Of these three causes, the wall vibration effects non-negligibly increase the first peak frequency [1]. Thus, only the wall vibration effects are considered in the following formulation. Meanwhile, the visco–thermal effects are approximated by incorporating a pressure attenuation term.

Figure 1 shows an acoustic tube representing the vocal tract. The cross-section is a circle whose area gradually changes along the long axis from the glottis to the lips; that is, the x-axis. Let $A(x,t)$, $p(x,t)$, and $U(x,t)$ be the cross-sectional area, pressure, and volume velocity, respectively, at position x and time t. Additionally, let Δx and Δt be sufficiently small in terms of the distance and time. Furthermore, the wall surface is assumed to displace only in the radial direction perpendicular to the x-axis owing to the wall vibration. Let r_0 be the original radius at position x, and r be the displacement at time t.

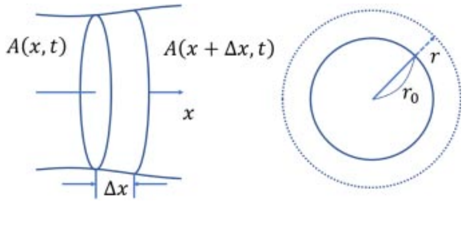

Fig. 1 A small section of the vocal tract (left panel) and wall displacement (right panel).

2.1. Continuity Equation

According to the definition of the bulk modulus, the pressure fluctuation Δp can be represented as:

$$\Delta p = -\rho c^2 \frac{\Delta V}{V}, \tag{1}$$

where ρ is the density of the medium, c is the speed of sound in the medium, V is the volume of the small section, and ΔV is the sum of the volume flow out of the section during Δt and the volume increment owing to the wall displacement caused by the wall vibration. Because Δx is sufficiently small, the volume of a truncated cone can be approximated by that of a tube as follows:

$$\begin{aligned}\Delta p = &-\rho c^2 \frac{(U(x+\Delta x,t) - U(x,t))\Delta t}{A(x,t)\Delta x} \\ &- \rho c^2 \frac{(A(x,t+\Delta t) - A(x,t))\Delta x}{A(x,t)\Delta x}.\end{aligned} \tag{1$'$}$$

Because $\Delta p = \dot{p}\Delta t$, this equation can be simplified as:

$$\dot{p} = -\rho c^2 \frac{U'}{A} - \rho c^2 \frac{\dot{A}}{A}, \tag{2}$$

where $\dot{p}$ is the time derivative of $p(x,t)$, U' is the spatial derivative of $U(x,t)$, and $\dot{A}$ is the time derivative of $A(x,t)$.

Without the second term, Eq. (2) is the acoustic continuity equation in a lossless medium because the spatial derivative of the particle velocity u' is equal to U'/A. Thus, the second term expresses the pressure fluctuation caused by the wall vibration. To approximate the visco–thermal losses on the vocal tract wall, if another term related to compressibility attenuation in the medium [6] is incorporated, then Eq. (2) is transformed as follows:

$$\dot{p} = -\rho c^2 \frac{U'}{A} - \rho c^2 \frac{\dot{A}}{A} - \alpha\rho c^2 p, \tag{3}$$

where α is the attenuation coefficient.

2.2. Equation of Motion

The impulse J in the x-axis direction of the truncated cone, shown in the left panel of Fig. 1, is calculated as follows:

$$\begin{aligned} J = &\int_{t}^{t+\Delta t} p(x,s)A(x,s)ds \\ &- \int_{t}^{t+\Delta t} p(x+\Delta x,s)A(x+\Delta x,s)ds \\ &+ \int_{t}^{t+\Delta t}\int_{x}^{x+\Delta x} p(y,s)A'(y,s)dyds. \end{aligned} \tag{4}$$

The third term on the right side of Eq. (4) is the impulse that the air of the truncated cone receives from the wall owing to the sound pressure. Because Δt is sufficiently small, by performing a Taylor expansion and numerical integration on the right side of Eq. (4), the first-order terms of Δt are extracted as follows:

$$\begin{aligned} J &= p(x,t)A(x,t)\Delta t \\ &\quad - p(x+\Delta x,t)A(x+\Delta x,t)\Delta t \\ &\quad + p(x,t)(A(x+\Delta x,t) - A(x,t))\Delta t \\ &= -A(x+\Delta x,t)p'\Delta x\Delta t. \end{aligned} \tag{4$'$}$$

Because the impulse J is equal to the change in momentum, $\rho\Delta x\dot{U}\Delta t$ and $\Delta xA(x+\Delta x,t)$ can be approximated by $\Delta xA(x,t)$, and Eq. (4′) is simplified as:

$$\rho\dot{U} = -Ap'. \tag{5}$$

Equation (5) is the acoustic equation of motion in a lossless medium because $\dot{u} = \dot{U}/A$. Note that wall vibration effects appear in the continuity equation but not in the equation of motion.

2.3. Wall Vibration

The displacement of the wall surface r (Fig. 1) is obtained by solving the equation of motion for a mass-spring-damper system representing the vocal tract wall. The mass, resistance, and spring constant per unit area of the wall surface are M, B, and K, respectively. The equation of motion is:

$$M\ddot{r} + B\dot{r} + Kr = p, \tag{6}$$

where $\ddot{r} = d^2r/dt^2$ and $\dot{r} = dr/dt$.

2.4. Discretization of Governing Equations

Figure 2 shows the staggered grid [7] in the cascaded acoustic tubes with an equal length, Δl, representing the vocal tract. Using this staggered grid and leapfrog algorithm [7], the governing equations can be discretized. Here, $p(n,m)$ denotes the pressure at the center of the n-th tube at time step m, and $U(n,m)$ indicates the volume velocity at the labial end of the n-th tube. The number of tubes is N, with the first tube ($n = 1$) corresponding to the glottal end and the last tube ($n = N$) to the labial end.

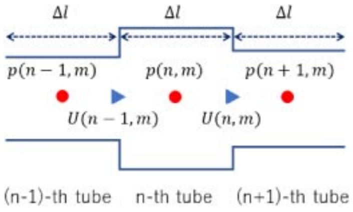

Fig. 2 Staggered grid in the cascaded acoustic tubes.

Although the second and third terms on the right side of Eq. (3) are necessary for the pressure correction, these terms require p. Thus, at each simulation step, p is first calculated using the following equation.

$$\dot{p} = -\rho c^2 \frac{U'}{A}. \tag{3$'$}$$

This is the continuity equation for lossless media. Although the wall vibration can change the area of each tube, the tube shape can be considered almost static because the amount of change is sufficiently small. Thus, the area of the n-th tube, $A(n)$, is time-invariant in Eqs. (3′) and (5). Therefore, Eq. (3′) can be discretized with second-order accuracy in space and time as follows.

$$\begin{aligned} p(n,m+1/2) &= p(n,m-1/2) \\ &\quad - \frac{\Delta t}{\Delta l}\frac{\rho c^2}{A(n)}\{U(n,m) - U(n-1,m)\}, \end{aligned} \tag{7}$$

where Δt is the simulation time step. Here, the pressure calculation is $\Delta t/2$ behind the volume velocity calculation according to the leapfrog algorithm because at each simulation step, the pressure calculation is executed first, followed by the volume velocity calculation. Equation (7) indicates that p at the new time step can be calculated from p and U at the previous time steps.

Using this p, the second and third terms of Eq. (3) can be calculated. For the second term, r can be obtained by solving Eq. (6). Let r_p be r at time $t - \Delta t$ and r_{pp} be r at time $t - 2\Delta t$, then the derivative is approximated by the difference; r at time t can be obtained as follows:

$$r = \frac{p + M(2r_p - r_{pp})/\Delta t^2 + Br_p/\Delta t}{M/\Delta t^2 + B/\Delta t + K}. \tag{8}$$

Then, the second term of Eq. (3) is approximated and linearized as follows:

$$\begin{aligned} -\rho c^2 \frac{\dot{A}}{A}\Delta t &= -\rho c^2 \frac{A(n,m+1/2) - A(n,m-1/2)}{\Delta t A(n)}\Delta t \\ &= -\rho c^2 \frac{\pi(r_0 + r)^2 - \pi(r_0 + r_p)^2}{\pi r_0^2} \\ &\approx -\rho c^2 \frac{2(r - r_p)}{r_0}. \end{aligned} \tag{9}$$

Note that Δt is derived from the left side of Eq. (3): $\dot{p} \approx \{p(n,m+1/2) - p(n,m-1/2)\}/\Delta t$. In addition, the wall vibration effects can be controlled by multiplying

Eq. (9) by the adjustment factor β ($0 \leq \beta \leq 1$). Therefore, the second term of Eq. (3) approximating by Eq. (9) can be discretized as follows:

$$-\beta\rho c^2 \frac{2\{r(n,m+1/2)-r(n,m-1/2)\}}{r_0(n)}, \qquad (9')$$

where $r(n,m+1/2)$ and $r(n,m-1/2)$ are the radii of the n-th tube at the current and previous simulation time steps, respectively, and $r_0(n)$ is the original radius of the tube.

Similarly, the equation of motion, Eq. (5), can be discretized with second-order accuracy in space and time as follows:

$$\begin{aligned} U(n,m+1) &= U(n,m) \\ &\quad -\frac{\Delta t}{\Delta l}\frac{A(n)}{\rho}\{p(n+1,m+1/2) \\ &\quad -p(n,m+1/2)\}. \qquad (10)\end{aligned}$$

Here, Eq. (10) indicates that U at the new time step can be calculated from U and p at the previous time steps, as in Eq. (7).

2.5. Input Volume Velocity

As shown in Fig. 2, the volume velocity at the glottal end of the first tube was not defined. Thus, the volume velocity should be provided externally as $U_{\mathrm{in}}(m)$. When an ideal impulse, that is, 1 in the first step and 0 in the other steps, is provided, the impulse response can be obtained as p_{out}, which is the output pressure waves at the lips. Using discrete Fourier transform, the obtained impulse response can be converted into the transfer function of $p_{\mathrm{out}}/U_{\mathrm{in}}$. When the glottal waves, such as the Rosenberg waves [8], are provided, vowel sounds can be obtained as p_{out}.

2.6. Radiation Impedance

According to Fig. 2 and Eq. (6), the volume velocity at the labial end of the N-th tube cannot be calculated. Thus, to calculate $U(N,m)$, radiation impedance should be introduced. Let A_{out} and U_{out} be the area and volume velocity of the N-th tube, respectively. According to Maeda [4], the acoustic radiation at the lips can be modeled by two elements, susceptance S_{rad} and conductance G_{rad} in parallel, which represent a circular piston in an infinite baffle. Essentially,

$$U_{\mathrm{out}} = \int_0^t dt S_{\mathrm{rad}} p_{\mathrm{out}} + G_{\mathrm{rad}} p_{\mathrm{out}}, \qquad (11)$$

where $S_{\mathrm{rad}} = 3\pi\sqrt{\pi A_{\mathrm{out}}}/8\rho$ and $G_{\mathrm{rad}} = 9\pi^2 A_{\mathrm{out}}/128\rho c$. In the present study, at time step m, $A_{\mathrm{out}} = A(N)$, $p_{\mathrm{out}} = p(N,m-1/2)$, and $U_{\mathrm{out}} = U(N,m)$. Thus, Eq. (11) can be approximated as follows:

$$\begin{aligned} U(N,m) &= \Delta t S_{\mathrm{rad}} \sum_m p(N,m-1/2) \\ &\quad + G_{\mathrm{rad}} p(N,m-1/2). \qquad (12)\end{aligned}$$

To improve the calculation accuracy, the trapezoidal rule should be applied to Eq. (11).

2.7. Stability Condition and Simulation Procedure

Prior to the simulation, Δl and Δt are determined. In the FDTD method, Δl is typically set to 1/10–1/20 of the wavelength. According to the Courant–Friedrichs–Lewy condition, Δt can be set to satisfy the stability condition $\Delta t \leq \Delta l/c$ [9].

When the input volume velocity U_{in} consists of M steps, the main part of the simulation procedure is presented in the following pseudocode:

```
$p_{\mathrm{sum}} = 0;$
for $m = 1, 2, \ldots, M$
    $p[1] = p[1] - \frac{\Delta t}{\Delta l}\frac{\rho c^2}{A[1]}(U[1] - U_{\mathrm{in}}[m]);$
    for $n = 2, 3, \ldots, N$
        $p[n] = p[n] - \frac{\Delta t}{\Delta l}\frac{\rho c^2}{A[n]}(U[n] - U[n-1]);$
        $r_{pp}[n] = r_p[n];$
        $r_p[n] = r[n];$
        $r[n] = \frac{p[n] + M(2r_p[n] - r_{pp}[n])/\Delta t^2 + Br_p[n]/\Delta t}{M/\Delta t^2 + B/\Delta t + K};$
        $p[n] = p[n] - \beta\rho c^2 \frac{2(r[n] - r_p[n])}{r_0} - \alpha\rho c^2 p[n]\Delta t;$
    end
    for $n = 1, 2, \ldots, N-1$
        $U[n] = U[n] - \frac{\Delta t}{\Delta l}\frac{A(n)}{\rho}(p[n+1] - p[n]);$
    end
    $p_{\mathrm{sum}} = p_{\mathrm{sum}} + p[N];$
    $U[N] = \Delta t S_{\mathrm{rad}} p_{\mathrm{sum}} + G_{\mathrm{rad}} p[N];$
end
```

Here, p_{sum} is the summation of p_{out} from the first to the current step, and β is the adjustment coefficient of the wall vibration effects. Note that the calculation of the wall vibration in the first tube is omitted in this pseudo code. Although the same variable appears on both sides, the time step is different. For example, $U[n]$ on the right side is one step (Δt) behind that on the left side.

3. VALIDATION

The simulation results are compared with those of TLM to examine the validity of the proposed model. Although the TLM does not always accurately calculate acoustic properties of the vocal tract in a wide frequency region, it is expected to be valid in the low-frequency region below 4 or 5 kHz [1,2]. Thus, by using the TLM and a uniform tube, the transfer function and radiation impedance are first investigated under the lossless wall condition. Subsequently, using the uniform tube with all the losses on the wall, the proper implementation of the wall vibration effects is examined. The adjustment factor β is tuned to agree with the first peak frequency calculated by the proposed model (f_{R1}) with that calculated by the TLM (f_{R1_TLM}). Finally, the transfer functions of the five Japanese vowels are calculated using the proposed model and TLM for comparison. Furthermore, the first four peak frequencies calculated by the proposed model (f_{R1}–f_{R4}) are compared with those calculated by TLM (f_{R1_TLM}–f_{R4_TLM}).

The TLM proposed by Adachi and Yamada [10] is used in these examinations. This model considers wall vibration effects, as well as the thermal exchange and viscous resistance on the vocal tract wall. In the following examinations, the TLM and the proposed model used the same values for the air density and speed of sound, $\rho = 1.17\,\mathrm{kg/m^3}$ and $c = 346\,\mathrm{m/s}$, respectively. The other simulation constants of the TLM are the same as those described in [10]. The simulation constants specific to the proposed model are as follows: attenuation coefficient, $\alpha = 2.0 \times 10^{-3}$; per unit area mass, $M = 15\,\mathrm{kg/m^2}$; per unit area resistance, $B = 6{,}170\,\mathrm{kg/m^2s}$; and per unit stiffness, $K = 1.34 \times 10^5\,\mathrm{kg/m^2s^2}$.

All the area functions in the examinations are measured at equal intervals (0.25 cm, to ensure that the tube length Δl was 0.25 cm), and the simulation time step Δt is set to 5.0×10^{-6} s, which satisfies the stability condition described in Sect. 2.7. In addition, the impulse response with a duration of 1 s is calculated using the proposed model such that the frequency resolution of the transfer function is 1 Hz.

3.1. Comparison of Transfer Functions and Radiation Impedances for a Uniform Tube with a Lossless Wall

Prior to implementing all the wall losses, the transfer function of $p_{\mathrm{out}}/U_{\mathrm{in}}$, and the radiation impedance for a uniform tube, which has no losses on the wall, are compared between the proposed model and TLM. The tube length and radius are 17 cm and 1 cm, respectively. For these calculations, in the proposed model, the wall vibration effects are eliminated by setting β as 0 and the pseudo visco–thermal losses are omitted by setting α as 0. In the TLM, the terms corresponding to the wall losses [10] are removed. Figure 3 shows these results. The transfer function and radiation impedance calculated by the proposed model agreed with those by the TLM. Note that the radiation impedance in the proposed model is calculated by $p_{\mathrm{out}}(\omega)/U_{\mathrm{out}}(\omega)$, where $p_{\mathrm{out}}(\omega)$ and $U_{\mathrm{out}}(\omega)$ are obtained by Fourier transforming the sound pressure and volume velocity at the lips when an impulse is input. On the other hand, the radiation impedance in the TLM is defined in the frequency domain [10]. The agreement of the radiation impedances between the two models indicates that Eq. (11) gives frequency dependent characteristics to the radiation impedance, although S_{rad} and G_{rad} have no frequency dependent characteristics in themselves.

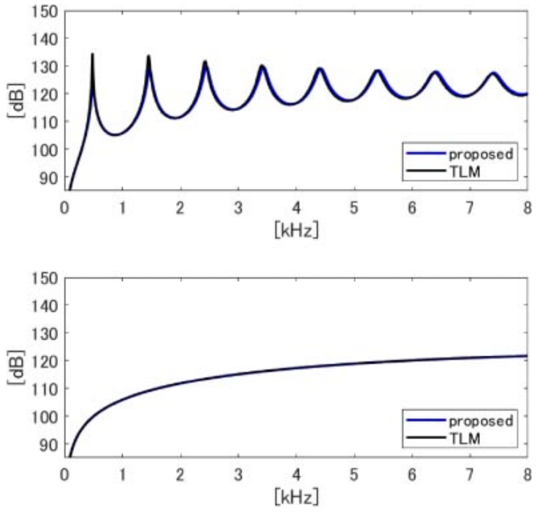

Fig. 3 Transfer function (upper) and radiation impedance (lower) of a uniform tube with no losses on the wall.

3.2. Effects of Wall Vibration and Adjustment Factor β

Figure 4 shows the pressure impulse response up to 16 ms for the uniform tube used in the previous section calculated using the proposed model. In this calculation,

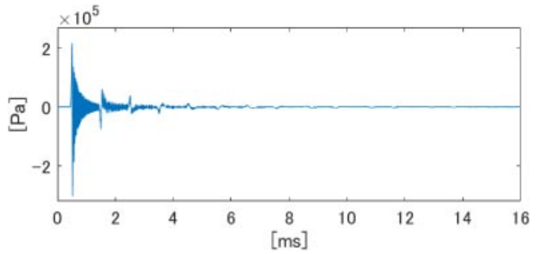

Fig. 4 Impulse response of a uniform tube calculated by the proposed model without considering the wall vibration effects ($\beta = 0$).

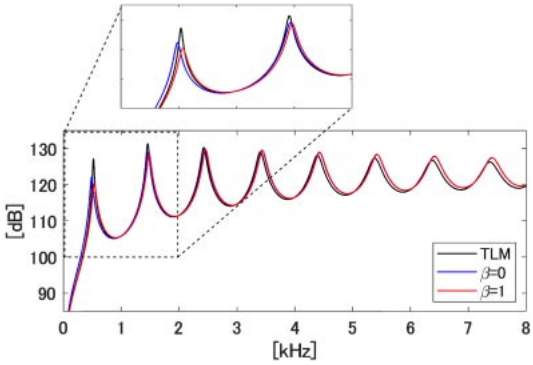

Fig. 5 Transfer functions of a uniform tube calculated by the TLM and proposed model ($\beta = 0$ and $\beta = 1$).

the pseudo visco–thermal losses are considered by setting α as 2.0×10^{-3}, while the wall vibration effects are eliminated by setting β as 0. The first peak of the impulse response is observed at 0.495 ms, which corresponds to the time at which the sound wave passed through the uniform tube.

Figure 5 shows three transfer functions of $p_{\mathrm{out}}/U_{\mathrm{in}}$ for the uniform tube. The line denoted by TLM presents the transfer function calculated by the TLM while considering all the wall losses. The other two lines indicate the transfer functions calculated by the proposed model with ($\beta = 1$) and without ($\beta = 0$) considering the wall vibration effects, respectively. The peaks calculated with the effects tended to be higher in frequency than those calculated without the effects in the lower peaks. The ratio of the frequency increase relative to the peak frequency without the effects is 9.63% for the first peak, 1.09% for the second peak, and less than 0.5% for the higher peaks. This result agreed with that of a previous study [1]; the wall vibration effects selectively increase the first peak frequency.

The first peak frequency calculated using the TLM, $f_{\mathrm{R1\_TLM}}$, is 519 Hz. For the proposed model, when the wall vibration effects are fully considered, that is, $\beta = 1$, the first peak frequency f_{R1} is 535 Hz. On the other hand, when the effects are ignored, that is, $\beta = 0$, f_{R1} is 488 Hz. Thus, to coincide the first peak frequencies calculated by the proposed model and TLM, β should have a value between 0 and 1. The value of β at which f_{R1} becomes 519 Hz is 0.64. Thus, β was set as 0.64 hereafter.

Regardless of the wall vibration effects, the peak levels of the transfer function calculated by the proposed model are lower than those calculated by the TLM below the third peak and vice versa above the peak. This could be due to the third term in Eq. (3). This term is not a frequency independent loss and uniformly reduces the peak level. On the other hand, the losses caused by the thermal exchange and viscous resistance are frequency dependent [10]. These differences could cause peak level differences.

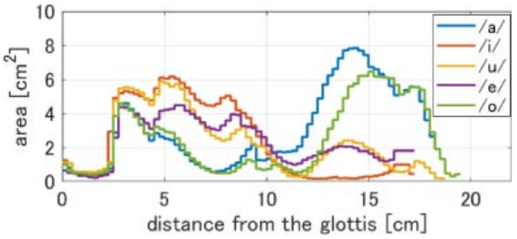

Fig. 6 Area functions of the vocal tract for the five Japanese vowels.

3.3. Comparison of Transfer Functions and the Lower Four Peak Frequencies for the Five Japanese Vowels

Figure 6 shows the vocal tract area functions of a native Japanese male for the five vowels /a/, /i/, /u/, /e/, and /o/. The area functions were extracted at equal intervals of 0.25 cm from MRI (Magnetic Resonance Imaging) data included in the "ATR MRI database of Japanese vowel production" [11], according to the algorithm described in Takemoto *et al.* [12]. The vocal tract lengths of /a/, /i/, /u/, /e/, and /o/ are 18.25, 17.25, 18.75, 17.25, and 19.50 cm, respectively. Note that the MRI data were obtained only during phonation, excluding the inhalation phase [13], and supplemented with teeth [14] in post-processing to ensure that the vocal tract shape can be accurately measured. Detailed information, such as scanning parameters, is described by Kitamura *et al.* [11].

Figure 7 shows the transfer functions of the five Japanese vowels calculated by the proposed model and TLM. Table 1 lists the frequencies of the first four peaks extracted from the transfer functions and percent difference. Note that when introducing the wall vibration effects using the adjusting factor ($\beta = 0.64$), the first peak frequency for all the vowels increases from 523 to 557 Hz for /a/ (6.5%), from 177 to 244 Hz for /i/ (37.9%), from 250 to 305 Hz for /u/ (22.0%), from 408 to 445 Hz for /e/ (9.1%), and from 361 to 408 Hz for /o/ (13.0%).

In addition, to examine the effects of these peak shifts on vowel perception, a preliminary discrimination test was performed using six subjects (two females and four males in the age range of 22–24 years) with no hearing problems. Five pairs of the vowels were synthesized using the proposed model, with and without considering the wall vibration effects when the same Rosenberg waves [8] were input. All the synthesized vowels were downsampled to 16 kHz. Using a forced choice method, each subject discriminated between the pairs randomly presented through headphones eight times. The correct ratios are 33% for /a/, 40% for /i/, 73% for /u/, 23% for /e/, and 75% for /o/. There was no clear relationship between the correct ratios and the amount or ratio of the peak shifts.

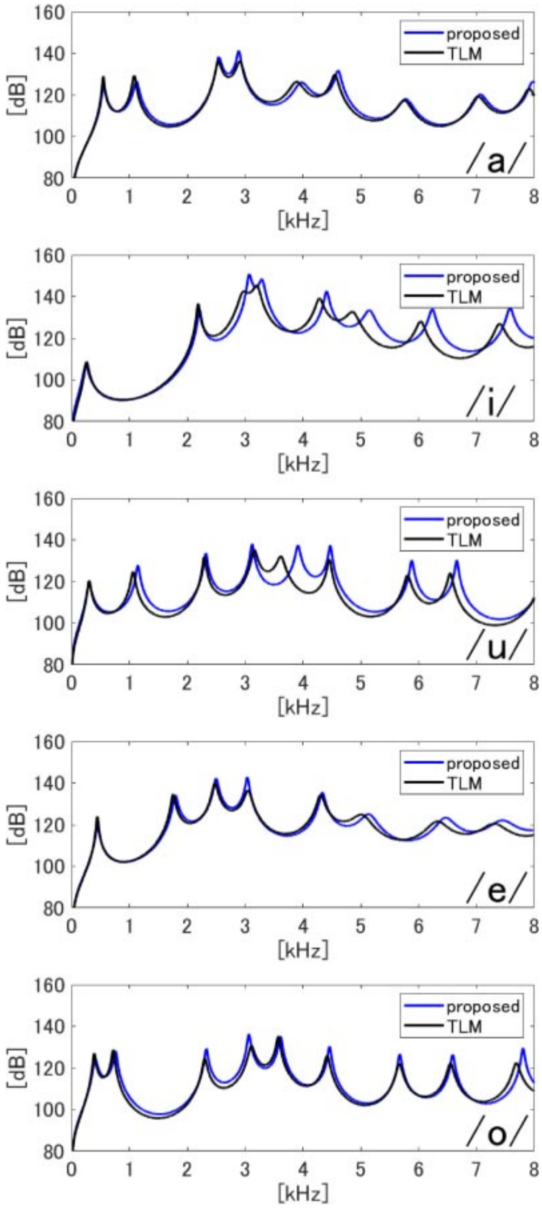

Fig. 7 Transfer functions of the vocal tract for the five Japanese vowels calculated by the proposed model and TLM.

Table 1 Lower four peak frequencies of the transfer functions for the five Japanese vowels calculated by the proposed model (f_{R1}–f_{R4}), those by the transmission line model ($f_{\mathrm{R1\_TLM}}$–$f_{\mathrm{R4\_TLM}}$), and the percent difference for the former relative to the latter ($f_{\mathrm{R1\_diff}}$–$f_{\mathrm{R4\_diff}}$). Bold type has an absolute percent difference over 5%.

	/a/	/i/	/u/	/e/	/o/
f_{R1}	557	244	305	445	408
f_{R2}	1,120	2,209	1,146	1,793	757
f_{R3}	2,546	3,072	2,319	2,498	2,329
f_{R4}	2,889	3,288	3,121	3,038	3,066
$f_{\mathrm{R1\_TLM}}$	547	261	304	444	391
$f_{\mathrm{R2\_TLM}}$	1,082	2,189	1,063	1,756	718
$f_{\mathrm{R3\_TLM}}$	2,539	2,980	2,293	2,479	2,302
$f_{\mathrm{R4\_TLM}}$	2,908	3,198	3,160	3,051	3,110
$f_{\mathrm{R1\_diff}}$	1.83	**−6.51**	0.33	0.23	4.35
$f_{\mathrm{R2\_diff}}$	3.51	0.91	**7.81**	2.11	**5.43**
$f_{\mathrm{R3\_diff}}$	0.28	3.09	1.13	0.77	1.17
$f_{\mathrm{R4\_diff}}$	−0.65	2.81	−1.23	−0.43	−1.41

Below the fourth peak for all the vowels, the transfer functions calculated by the proposed model agree well with those calculated by the TLM. The mean absolute percent difference for these peaks is 2.30%, whereas $f_{\mathrm{R1\_diff}}$ for /i/ and $f_{\mathrm{R2\_diff}}$ for /u/ and /o/ are larger than 5%. However, above the fifth peak, the two transfer functions for the vowels /a/, /e/, and /o/ are in good agreement with each other, while those for the vowels /i/ and /u/ are not.

For the first peak, a large difference is noted in the vowel /i/. This vowel has the lowest f_{R1} among all the vowels. As mentioned above, the first peak frequency increases when the adjustment factor of β increases. Therefore, for this vowel, because f_{R1} is smaller than $f_{\mathrm{R1\_TLM}}$, the value of β should be larger than 0.64. This result indicates that it is difficult to obtain the value of β, which adjusts the wall vibration effects of the proposed model to the same extent as those of the TLM for all the vowels.

For the second and higher peaks, large differences are noted in the vowels /i/, /u/, and /o/. These vowels have commonly strong constrictions in the vocal tract and/or a small area at the lips (Fig. 5). These facts indicate that the tubes with small cross-sectional areas could cause these differences. Near the constrictions, the pressure and volume velocity drastically change between adjacent tubes around the high resonance frequencies, indicating that finer temporal and spatial resolutions are necessary in the difference approximation. In fact, we confirmed that those differences reduced when the area function was resampled at 0.1 cm intervals. A possible cause for these differences is the FDTD scheme: the pressure and volume velocity are calculated at positions that are half-shifted at spatial discrete intervals and at times that are half-shifted at time discrete intervals.

4. CONCLUSION

In this paper, we proposed a 1D model of the static vocal tract that simulates wave propagation by considering wall vibration effects in the time domain. The formulation process of the governing equations showed that the term representing the wall vibration effects appear only in the continuity equation but not in the equation of motion. The governing equations were discretized using the FDTD scheme, that is, the staggered grid and leapfrog algorithm [7], as described in Sect. 2.7. Briefly, this model is categorized as a 1D FDTD model with a second-order accuracy in space and time.

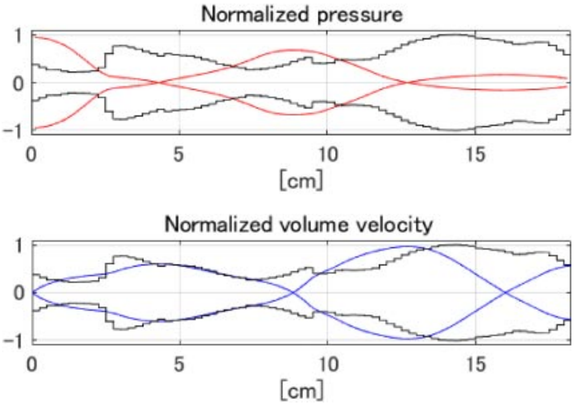

Fig. 8 Resonance mode of pressure (upper) and volume velocity (lower) at f_{R3} for /a/.

In the proposed model, the wall vibration effects were considered using a mass-spring-damper system. As described in Sect. 3.1, this model tended to slightly overestimate the effects compared with those of the TLM. When the effects were fully loaded, the first peak frequency of the uniform tube calculated by the proposed model was slightly higher than that calculated by the TLM. To suppress the effects as much as those in the TLM, the adjustment factor β was introduced and its value was determined to be 0.64. To mitigate the overestimation, the wall properties M, B, and K need to be reexamined.

The transfer functions for the five Japanese vowels calculated by the proposed model agreed with those calculated by the TLM in the lower frequency region below the fourth peak, as described in Sect. 3.3. However, relatively large discrepancies have been observed in several cases. Although the causal factors could not be identified, small cross-sectional areas at the lips and constrictions possibly decreased the accuracy of the difference approximation, resulting in the discrepancies.

The proposed model cannot always calculate the same transfer function as the TLM and the difference approximation may decrease the calculation accuracy; nevertheless, there are a few advantages. First, the proposed model is easy to implement; it can be implemented within approximately 50 lines of code using MATLAB. Second, the proposed model can easily calculate and visualize the resonance mode of the pressure and volume velocity in the vocal tract when the sinusoidal waves with the peak frequency are input (Fig. 8). Third, because the pressure just above the glottis can be obtained at every simulation time step, the proposed model can be easily combined with the vocal fold model, such as the two-mass model [15], to examine the acoustic interactions between the vocal fold and tract. Hence, despite its disadvantages pertaining to the relatively low calculation accuracy, the proposed model is simple, useful, and extendable, furthering the understanding of vocal tract acoustics in students.

ACKNOWLEDGMENT

This research was partly supported by JSPS KAKENHI Grant Number 19K12048.

REFERENCES

[1] K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, MA, 2000).

[2] J. L. Flanagan, *Speech Analysis, Synthesis, and Perception* (Springer-Verlag, New York, 1972).

[3] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, Paris, 1970).

[4] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Commun.*, **1**, 199–229 (1982).

[5] B. H. Story, *Physiologically-based Speech Simulation using an Enhanced Wave-reflection Model of the Vocal Tract* (Ph.D. Dissertation, University of Iowa, 1995).

[6] X. Yuan and M. Berggren, "Formulation and validation of Berenger's PML absorbing boundary for the FDTD simulation of acoustic scattering," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **44**, 816–822 (1997).

[7] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas Propag.*, **AP-14**, 302–307 (1966).

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

[9] S. Sakamoto, "Phase-error analysis of high-order finite difference time domain schema and its influence on calculation results of impulse response in closed sound field," *Acoust. Sci. & Tech.*, **28**, 295–309 (2007).

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

[11] T. Kitamura, H. Takemoto, S. Adachi and K. Honda, "Transfer functions of solid vocal-tract models constructed from ATR MRI database of Japanese vowel production," *Acoust. Sci. & Tech.*, **30**, 288–296 (2009).

[12] 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).

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

[14] 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).

[15] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.*, **51**, 1233–1268 (1972).