Journal of Sound and Vibration 411 (2017) 247–259

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Multimodal modeling and validation of simplified vocal tract acoustics for sibilant /s/

T. Yoshinaga [a], A. Van Hirtum [a,b,*], S. Wada [a]

[a] Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, 560-8531, Osaka, Japan
[b] GIPSA-Lab, UMR CNRS 5216, Grenoble Alpes University, 11 rue des Mathématiques (BP46), 38402, Grenoble, France

ARTICLE INFO

Article history:
Received 17 October 2016
Revised 6 September 2017
Accepted 6 September 2017
Available online XXX

Keywords:
Speech production
Multimodal theory
Simplified vocal tract geometry
Sibilant fricatives

ABSTRACT

To investigate the acoustic characteristics of sibilant /s/, multimodal theory is applied to a simplified vocal tract geometry derived from a CT scan of a single speaker for whom the sound spectrum was gathered. The vocal tract was represented by a concatenation of waveguides with rectangular cross-sections and constant width, and a sound source was placed either at the inlet of the vocal tract or downstream from the constriction representing the sibilant groove. The modeled pressure amplitude was validated experimentally using an acoustic driver or airflow supply at the vocal tract inlet. Results showed that the spectrum predicted with the source at the inlet and including higher-order modes matched the spectrum measured with the acoustic driver at the inlet. Spectra modeled with the source downstream from the constriction captured the first characteristic peak observed for the speaker at 4 kHz. By positioning the source near the upper teeth wall, the higher frequency peak observed for the speaker at 8 kHz was predicted with the inclusion of higher-order modes. At the frequencies of the characteristic peaks, nodes and antinodes of the pressure amplitude were observed in the simplified vocal tract when the source was placed downstream from the constriction. These results indicate that the multimodal approach enables to capture the amplitude and frequency of the peaks in the spectrum as well as the nodes and antinodes of the pressure distribution due to /s/ inside the vocal tract.

1. Introduction

The production mechanisms of sibilant fricatives have been discussed using simplified vocal tract geometries and aeroacoustic theory [1–3]. Stevens [1] argued that the sibilant sound is generated by turbulent flow generated by the constriction in the vocal tract, and modeled the sound generation with a vibrating spoiler in a tube. Shadle [2] assumed that the frequency characteristics of sibilants are determined by the position of the constriction formed by the tongue and the alveolar ridge, and constructed simplified vocal tract replicas to test this hypothesis. The sound generated by the replicas was measured experimentally, and two constriction positions, 15 mm and 25 mm from the lip outlet, reproduced the characteristic spectrum peak of sibilant /s/ (4 kHz), and /ʃ/ (2.5 kHz), respectively. An example of a spectrum of sustained /s/ pronounced by a Japanese male speaker is shown in Fig. 1. In the spectrum, first characteristic peak appeared at 4 kHz and the maximum amplitude was observed at 8 kHz. In previous studies, the first peak of /s/ varied from 4 to 7 kHz, and the high frequency peak can be broad or more peaky depending on the speaker's language or gender [4,5] whereas frequency values depend mainly on the constriction

[*] Corresponding author. GIPSA-Lab, UMR CNRS 5216, Grenoble Alpes University, 11 rue des Mathématiques (BP46), 38402, Grenoble, France.
E-mail address: annemie.vanhirtum@gipsa-lab.grenoble-inp.fr (A. Van Hirtum).

http://dx.doi.org/10.1016/j.jsv.2017.09.004
0022-460X/© 2017 Elsevier Ltd. All rights reserved.

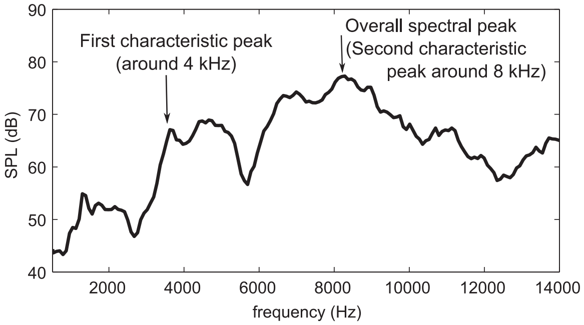

Fig. 1. Example of a spectrum of sustained /s/ pronounced by a male Japanese speaker. There was no vowel context for /s/ and the sound was measured at 30 cm from the lip of the speaker. The simplified geometry is derived from geometrical data measured on this subject (see section 2.2).

position and volume of the front cavity [2].

In addition to the mechanical experiment, Shadle [2] theoretically modeled the sound generation using a plane-wave model for 30 mm front cavity and a dipole source at the upstream surface of the obstacle, and discussed about the frequency characteristics of the generated sound. Howe and McGowan [3] assumed that a monopole sound source is generated at the space between the upper and lower teeth, and constructed a one-dimensional theoretical model. The modeled spectrum and overall sound pressure level (SPL) of /s/ were validated against measurements of the same sound made by human speakers. Previous studies have shown that simplified geometries can be used to reproduce spectral features of sibilant fricatives. However, in the studies described above, the potential impact of higher-order acoustic resonance modes related to the three-dimensional (3D) geometry of the vocal tract on the spectral features was not studied.

Blandin et al. [6] applied multimodal theory, which considers propagation of 3D modes, to vocal tract geometries for vowels. The modeled pressure fields and transfer functions were validated against experimental observations and finite-element simulations. The influence of higher-order modes on spectral amplitudes above 4.5 kHz was observed. Because fricatives are characterized by the acoustic energy in this frequency range, this influence indicates that higher-order modes affect perceptually-significant spectral features of fricatives more than those of vowels. Therefore, influence of higher-order modes on pressure fields inside the vocal tract, as well as radiated pressure outside of the vocal tract for sibilant /s/ will be investigated in this paper.

Motoki et al. [7] and Motoki [8] applied multimodal theory to a simplified vocal tract geometry of /ʃ/ to investigate the influence of geometrical changes on the transfer function of the vocal tract. However, the characteristic spectrum peak of /ʃ/ was lacking from the modeled transfer function because the sound source was located at the inlet of the vocal tract. Indeed, since unvoiced sibilant fricatives are generated mainly by the impingement of jet flow on the teeth and lips, the main sound source is located downstream from the constriction [1–3].

Therefore, in this study, multimodal theory is applied to a simplified vocal tract geometry of sibilant /s/ with two different source positions, at the vocal tract inlet and downstream from the constriction. Since the excitation and propagation of higher order acoustic modes is known to depend on the center line curvature of the geometry, the used simplified geometry accounts for the curvature of the vocal tract near the teeth which was omitted in the cited studies [1–3] although some recent studies do propose a simplified curved geometry to study the flow field [9,10]. The outcome of the model is compared with experimental data obtained by imposing a known acoustic source and by supplying airflow at the inlet. In addition, the position of the source downstream from the constriction was varied in order to investigate the effect of the source position on the acoustic characteristics of sibilant /s/. By comparing theoretically modeled spectra with experimental measurements with flow, we examined whether multimodal theory can predict the characteristic peak of sibilant /s/ or other acoustic characteristics, such as antinodes in the vocal tract. Frequencies of up to 14 kHz are measured and compared with multimodal modeling in this paper.

2. Method

2.1. Multimodal theory

Multimodal theory has been developed and implemented by several researchers, *e.g.* Refs. [6–8,11–15]. Vocal tract geometry is simplified as a concatenation of waveguides with constant cross-sections. In the following, z indicates the main propagation direction; cross-sections are situated in the (x, y)-plane, with the x-axis from left to right and the y-axis from inferior to superior.

In the 3D acoustic field, the amplitude of sound pressure $p\,(x, y, z)$ and particle velocity vector $v\,(x, y, z)$ are defined as

$$p(x, y, z) = \jmath\omega\rho\phi(x, y, z), \tag{1}$$

$$\mathbf{v}(x, y, z) = -\nabla\phi(x, y, z), \tag{2}$$

with velocity potential $\phi(x, y, z)$ omitting the time dependence exp (jωt), where ω is angular frequency and ρ is density. The velocity potential satisfies the 3D spatial wave equation *i.e.* Helmholtz equation

$$\nabla^2\phi(x,y,z) + k^2\phi(x,y,z) = 0, \tag{3}$$

where $k = \omega/c$ is free field wave number and c is speed of sound. The solution of Helmholtz equation yields a summation of an infinite number of propagation modes ψ_{mn} (x, y) weighted by forward and backward propagation amplitudes a_{mn}, b_{mn} as:

$$\phi(x,y,z) = \sum_{m,n=0}^{\infty} \psi_{mn}(x,y)\{a_{mn}\exp(-\gamma_{z,mn}z) + b_{mn}\exp(\gamma_{z,mn}z)\}, \tag{4}$$

where m and n are the number of modes in the x-direction and y-direction, respectively, and $\gamma_{z,mn}$ is the propagation constant (modal wave numbers) along the z-axis. The propagation mode ψ_{mn} (x, y) is the solution of the two-dimensional Helmholtz equation and can be obtained analytically when considering waveguides with a rectangular cross-section of dimensions L_x and L_y [11];

$$\psi_{mn}(x,y) = \frac{\cos(m\pi x/L_x)}{\sqrt{L_x\sigma_m}}\frac{\cos(n\pi y/L_y)}{\sqrt{L_y\sigma_n}}, \tag{5}$$

where σ_m, σ_n are 1 (m, n = 0, *i.e.* plane wave) or 1/2 ($m, n \geq 1$). The propagation constant is derived from the dispersion relationship as

$$\gamma_{z,mn} = \sqrt{\left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2 - k^2}. \tag{6}$$

When the propagation constant is $\gamma_{z,mn} = 0$, the cutoff frequency $f_{c,mn}$ yields

$$f_{c,mn} = \frac{c}{2\pi}\sqrt{\left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2}. \tag{7}$$

Each propagation mode is rapidly attenuated along the waveguide for frequencies below its cutoff frequency, but can propagate above that cutoff frequency. The cutoff frequency of plane wave ($m, n = 0$) is 0 Hz so that sound can propagate at any frequency.

To obtain the acoustic field in the waveguide, the infinite series in Eq. (4) is truncated to a certain value depending on the frequency range of interest. From Eqs. (1), (2) and (4), the sound pressure and particle velocity along the z-axis v_z are calculated as

$$p(x,y,z) \approx \boldsymbol{\psi}^{\mathrm{T}}(x,y)\{\mathbf{D}(-z)\mathbf{a} + \mathbf{D}(z)\mathbf{b}\}, \tag{8}$$

$$\begin{aligned} v_z(x,y,z) &= -\frac{\partial\phi(x,y,z)}{\partial z} \\ &\approx \boldsymbol{\psi}^{\mathrm{T}}(x,y)\mathbf{Z}_C^{-1}\{\mathbf{D}(-z)\mathbf{a} - \mathbf{D}(z)\mathbf{b}\}, \end{aligned} \tag{9}$$

where superscript T denotes the transpose operator, $\mathbf{D}(z)$ is propagation constant matrix $\mathbf{D}(z)$ = diag [exp $(\gamma_{z,mn}z)$], $\mathbf{Z}_C$ is characteristic impedance matrix $\mathbf{Z}_C$ = diag [j$\omega\rho/\gamma_{z,mn}$], and $\boldsymbol{\psi}$, $\mathbf{a}$, and $\mathbf{b}$ are column vectors composed of ψ_{mn}, j$\omega\rho a_{mn}$, and j$\omega\rho b_{mn}$, respectively. The modal sound pressure $\mathbf{P}$ and modal particle velocity $\mathbf{V}$ are then defined as

$$\mathbf{P} = \mathbf{D}(-z)\mathbf{a} + \mathbf{D}(z)\mathbf{b}, \tag{10}$$

$$\mathbf{V} = \mathbf{Z}_C^{-1}\{\mathbf{D}(-z)\mathbf{a} - \mathbf{D}(z)\mathbf{b}\}. \tag{11}$$

When we consider a rectangular waveguide with varying cross-section, continuity equations of pressure and volume velocity are applied at each junction. Each mode is projected through the junction by considering the mode-coupling matrix

$$\boldsymbol{\Psi}_{i,i+1} = \frac{1}{S_i}\int_{S_i}\boldsymbol{\psi}_i(x,y)\boldsymbol{\psi}_{i+1}^{\mathrm{T}}(x,y)\mathrm{d}S \tag{12}$$

between sections i and $i + 1$, where S_i is the area of section i. Note that $S_i < S_{i+1}$ and the area expands from section i to the section $i + 1$. By using the coupling matrix, modal pressure, modal velocity and impedance matrix are calculated as

$$\mathbf{P}_i = \boldsymbol{\Psi}_{i,i+1}\mathbf{P}_{i+1}, \tag{13}$$

$$\mathbf{V}_{i+1} = \boldsymbol{\Psi}_{i,i+1}^{\mathrm{T}}\mathbf{V}_i, \tag{14}$$

$$\mathbf{Z}_i = \boldsymbol{\Psi}_{i,i+1}\mathbf{Z}_{i+1}\boldsymbol{\Psi}_{i,i+1}^{\mathrm{T}}. \tag{15}$$

The outlet of the waveguide is flanged with an infinite baffle in order to approximate the baffle shape of a face. The radiation impedance matrix Z_{rad} at the outlet is obtained as

$$\begin{aligned}\mathbf{Z}_{\mathrm{rad}} &= [Z_{mn,pq}]\\ &= \left[\frac{\mathrm{J}k\rho c}{2\pi S}\int_S\int_S \psi_{mn}(x,y)\psi_{pq}(x',y')\frac{\mathrm{e}^{-\mathrm{J}kr}}{r}\mathrm{d}S'\mathrm{d}S\right],\\ r &= \sqrt{(x-x')^2+(y-y')^2},\end{aligned} \tag{16}$$

where S is the area of the outlet, and (x, y) and (x', y') are coordinates of points on S [15]. The relationship between the modal pressure, modal velocity, and radiation impedance at the outlet is written as

$$\mathbf{P}_{\mathrm{out}} = \mathbf{Z}_{\mathrm{rad}}\mathbf{V}_{\mathrm{out}}. \tag{17}$$

The radiation impedance is then propagated backward to get the impedance matrix at each section from the exit towards the position of the sound source. Then, the modal pressure and modal velocity are calculated from the sound source to the outlet of the waveguide. The far-field sound pressure at the position (x, y, z) is calculated using Rayleigh-Sommerfield integral [11]

$$\begin{aligned}p(x,y,z) &= \frac{\mathrm{J}k\rho c}{2\pi S}\int_S \mathbf{V}_{\mathrm{out}}\cdot\boldsymbol{\psi}_{\mathrm{out}}(x',y')\frac{\mathrm{e}^{\mathrm{J}kh}}{h}\mathrm{d}S',\\ h &= \sqrt{(x-x')^2+(y-y')^2+(z-z')^2}.\end{aligned} \tag{18}$$

At each frequency, the pressure distribution inside and outside of the waveguide are computed from the radiation impedance and particle velocity at the sound source. All equations are implemented in MATLAB R2013a (Mathworks, Natick, USA).

2.2. Simplified vocal tract geometry

A simplified vocal tract geometry is used consisting of a concatenation of six sections, each with uniform cross-sectional area, and a rectangular cross-sectional shape. The geometry of the vocal tract pronouncing /s/ obtained by computed tomography (CT) scan [16] is depicted in Fig. 2 (a). The subject is a 32-year-old male native Japanese speaker. He has normal dentition (Angle Class I) and no speech disorder in self-report. CT scan data were taken in 9.6 s while the subject sustained /s/ in seated position without vowel context. His vocal tract geometry was simplified to a rectangular channel based on these six sections: the back cavity extending from the pharynx to the posterior part of oral cavity (section 1); the tongue constriction (section 2); the region above the anterior tongue (section 3); space in the z-direction between lower and upper teeth (section 4); space below upper teeth (section 5) and lip cavity (section 6). By using a cross-sectional area and height at the center of each section, the six rectangular cavities were constructed. The geometry of the simplified vocal tract is illustrated in Fig. 2 (b) and dimensions are given in Table 1. It is observed that the center line from resulting vocal tract geometry is curved. The total length from upstream inlet into section 1 to downstream outlet from section 6 is 172 mm. We confirmed that the simplified geometry with air flowing through it at 300 cm^{3}s^{-1} reproduces the main spectral features of sound /s/ shown in Fig. 1 (first and overall spectral peak), and the maximum discrepancy between the spectra was less than 9 dB in the frequency range 0.5–14 kHz.

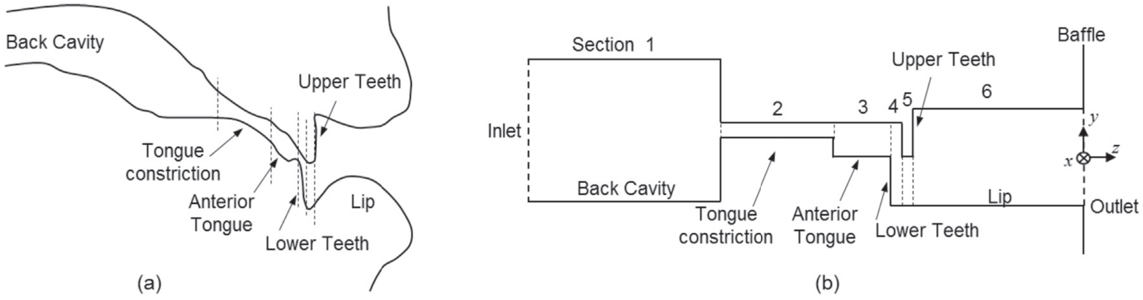

Fig. 2. (a) Mid-sagittal section of a vocal tract of male Japanese subject pronouncing /s/, (b) Simplified rectangular vocal tract geometry of sibilant /s/.

Table 1
Dimensions (mm) of each section of the sibilant /s/ vocal tract geometry.

Section	1	2	3	4	5	6
L_x	16.8	8	23	21	21	23.5
L_y	12.5	1.25	3	7.25	4.25	8.5
L_z	140	10	5	1	1	14

2.3. Source at the inlet of the back cavity

For computational model 1, the sound source was positioned at the inlet of section 1 and the pressure distribution in section 6 was modeled. The modal velocity of the sound source at the inlet is calculated as

$$\mathbf{V}_{\text{in}} = \int_{\Omega_0} v_0 \boldsymbol{\psi}_{\text{in}}(x, y)\text{d}S, \tag{19}$$

where v_0 is the particle velocity of the sound source, Ω_0 the area of the vibrating surface. We imposed a 10 × 10 mm vibrating surface with $v_0 = 1$ mms^{-1} at the center of the inlet face. According to the cutoff frequencies of each mode at the outlet (section 6), which are shown in Table 2, four modes from mode number 1 to 4 are used for section 6 since only the audible spectrum is of interest (≤ 20 kHz). There is no difference on the modeled pressure distributions when more modes are included (mode number greater than 5) at section 6. Meanwhile, by considering only one mode for all sections, the differences between the plane wave and the multimodal models were investigated.

For physical model 1, a replica of the vocal tract geometry was constructed using rapid prototyping of plaster (Zprinter, 3D systems, USA; accuracy: ±0.1 mm). The compression driver (PSD2002S-8, Eminence, USA), which produces the sound, was connected to the center of the inlet with a communication hole of diameter 11 mm. The outlet of the replica had a round edge of radius 1 mm. A rectangular baffle (350 × 350 mm) was attached to the edge of the outlet in order to mimic the flanged outlet condition used in multimodal theory. A 25-mm microphone probe (type 4182, B&K, Denmark) was positioned inside the replica (section 6) with a 3D spatial positioning system (PS35, OWIS, Germany; accuracy: ±100 μm). The schematic of the experimental setup is depicted in Fig. 3.

Measurements were taken along two planes, horizontal plane ($x - z$) and vertical plane ($y - z$), within section 6. In the horizontal plane ($y = 0$), measurements were taken in 2-mm intervals in both the x-direction and the z-direction. In the vertical plane ($x = 0$), measurements were taken in 1-mm intervals in the y-direction and 2-mm steps in the z-direction. The starting position was 0.2 mm downstream from section 5 along the center line (x, $y = 0$, $z = -13.8$), and the position was varied in -10 mm $\leq x \leq 10$ mm, $-$ 13.8 mm $\leq z \leq 0.2$ mm for the horizontal plane and in -3 mm $\leq y \leq 3$ mm, $-$ 13.8 mm $\leq z \leq 0.2$ mm for the vertical plane. In total, the microphone tip was placed at 88 positions within the horizontal plane and at 56 positions within vertical plane.

At each measurement position, the compression driver produced a linear sweep sound signal from 2 kHz up to 15 kHz with a duration of 20 s. The acoustic pressure p at the microphone probe was recorded during each sweep signal using a data

Table 2
Mode (m, n) and modeled cutoff frequency at the outlet of the sibilant /s/ geometry.

Mode number	1	2	3	4	5	6
(m, n)	(0,0)	(1,0)	(2,0)	(0,1)	(1,1)	(3,0)
f_c (kHz)	0	7.3	14.6	20.2	21.5	21.9

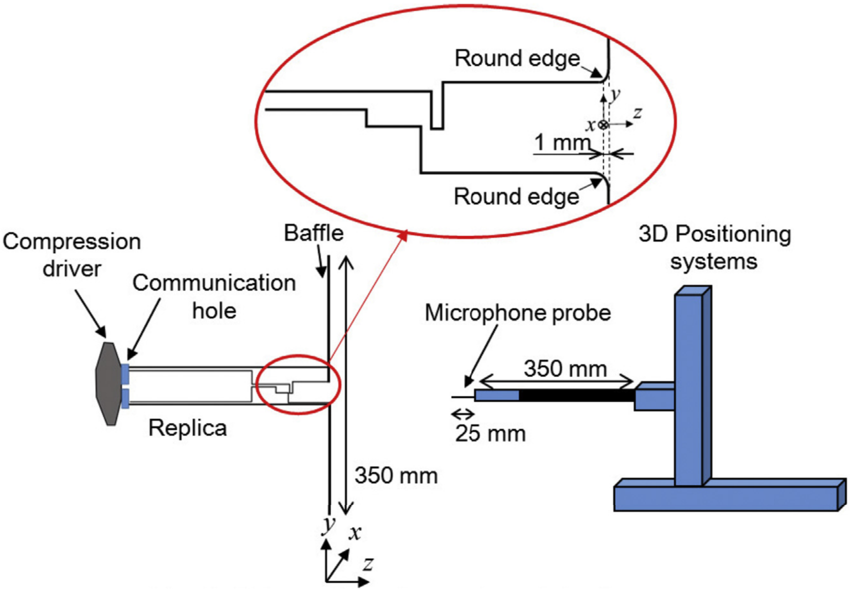

Fig. 3. Schematics of experimental setup.

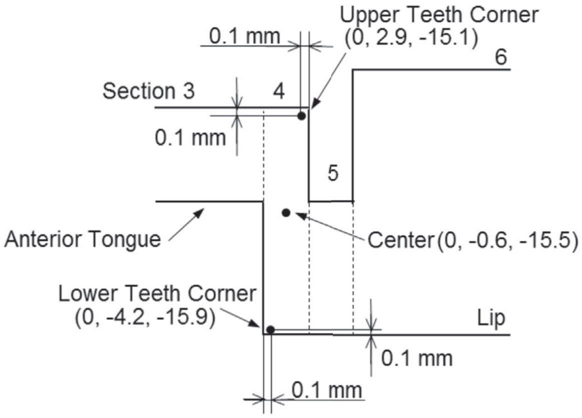

Fig. 4. Position of the sound source in section 4.

acquisition system (PXI0MIO 16XE, National Instruments, USA) with sampling frequency 44.1 kHz. The measured signal was Fourier transformed with 1024 sample points and averaged over 863 time segments (total 20 s). The following pressure-pressure transfer function is used to compare the measured and modeled pressure distribution along the centerline of section 6;

$$G_{\mathrm{lip}}(f) = 20\log_{10}(p(0,0,-1.8,f)/p(0,0,-13.8,f)). \tag{20}$$

Two pressure positions were chosen from the measurement positions closest to the teeth ($z = -13.8$ mm) and near the outlet ($z = -1.8$ mm).

2.4. Source downstream from the constriction

When the turbulent flow is generated in the vocal tract, monopole, dipole, and quadrupole sound sources of flow fluctuation are assumed to be produced [1–3]. For computational model 2, to simulate the source generation, the pressure distribution and far-field sound were calculated using multimodal theory with a simple monopole source downstream from the constriction. Then, the source position was changed inside the vocal tract to explore the main source position in the vocal tract. The same vocal tract geometry shown in Fig. 2 (b) was used. At the sound source position, the section was divided into an upstream section and a downstream section. The modal pressure and velocity are calculated at each section as

$$\mathbf{P}^{+} - \mathbf{P}^{-} = 0, \tag{21}$$

$$\mathbf{V}^{+} - \mathbf{V}^{-} = \mathbf{Q}, \tag{22}$$

$$\mathbf{Q} = Q\boldsymbol{\psi}(x,y), \tag{23}$$

where + and − represent the variables downstream and upstream of the source section, and Q is the volume flow rate supplied at the inlet [14]. We imposed a fluctuating volume flow rate $Q = 190$ mm^{3}s^{-1} as a monopole sound source to reproduce the SPL of the flow source [3].

The position of the sound source was varied from the center in x, y and z direction of the particular section from 3 to 6 (*i.e.* sections downstream from the constriction). In addition, the source was shifted away from the center position in section 4 to explore the influence of source position in the space between lower and upper teeth in more detail since findings in literature [3,17] suggest that the sound source is situated here. Concretely, we located the sound source near the upper teeth corner ($y = 2.9, z = -15.1$), center of section 4 ($y = 0, z = -15.5$), and near the lower teeth corner ($y = -4.15, z = -15.9$). The positions of the sources are depicted in Fig. 4. By changing the source position in this model, effects of the source position on the spectral characteristics as well as internal multimodal pressure patterns were investigated.

Moreover, in order to examine the differences between the plane wave and the multimodal models, the number of modes was decreased to one as in the Computational Model 1. The inlet impedance was set as a non-reflective boundary condition, *i.e.* the characteristic impedance Z_C was calculated with the same area of inlet used for the acoustic driver.

The modeled far-field sound spectra with the source downstream from the constriction were compared with the spectra measured when airflow was supplied to the replica. For physical model 2, the same mechanical replica of Fig. 2 (b) constructed using rapid plaster prototyping was used in this flow experiment. Steady airflow was provided using a compressor (YC-4RS, Yaezaki, Japan) equipped with a mass-flow controller (MQV0050, Azbil, Japan) and a 3 m air tube of inner diameter 8 mm connected to the inlet of the replica. The length of the air tube is long enough to dissipate the sound generated upstream from the inlet of section 1 (upstream noise due to flow was less than 1 dB). The flow rate was fixed at an average flow rate for sibilant /s/ at 300 cm^{3}s^{-1} [18]. A rectangular baffle (350 × 350 mm) was attached at the outlet of the replica to be consistent with the infinite baffle of the model.

The sound generated by the replica was measured with a 1/4 inch omnidirectional microphone (Type 4939, Bruel & Kjaer, Denmark) placed 30 cm downstream from the outlet of the mechanical replica along the centerline (z-axis) of section 6. The sound was recorded for 2 s using a data acquisition system (PXIe-4492, National Instruments, USA) with sampling frequency 44.1 kHz. The measured acoustic pressure was Fourier transformed with 512 sample points multiplied by a Hanning window and averaged for 200 time segments with 30% overlap. The SPL of the modeled and measured pressures were calculated based on the reference level 20×10^{-6} Pa.

3. Results and discussion

3.1. Acoustic characteristics with source at the inlet

For the case where the compression driver is providing an acoustic source at the inlet, the pressure amplitudes were measured and modeled along the horizontal and vertical planes in section 6 for different frequencies (5050 Hz, 10,050 Hz, and 13,050 Hz). The horizontal plane comparisons are shown in Fig. 5. Note that as in Fig. 2 (b), $z = 0$ mm corresponds to the outlet. The pressure amplitude was normalized by its maximum value observed within the plane for each frequency. As the frequency increased, the maximum pressure shifted from $z = -12$ mm, near the teeth, to $z = -4$ mm, near the outlet, in both measured and modeled pressure fields. In addition, small troughs appeared at $z = -13$ mm for 10,050 Hz and at $z = -9$ mm for 13,050 Hz. Note that small 3D effects were observed in both the measured and modeled pressure distribution below the cutoff frequency of the

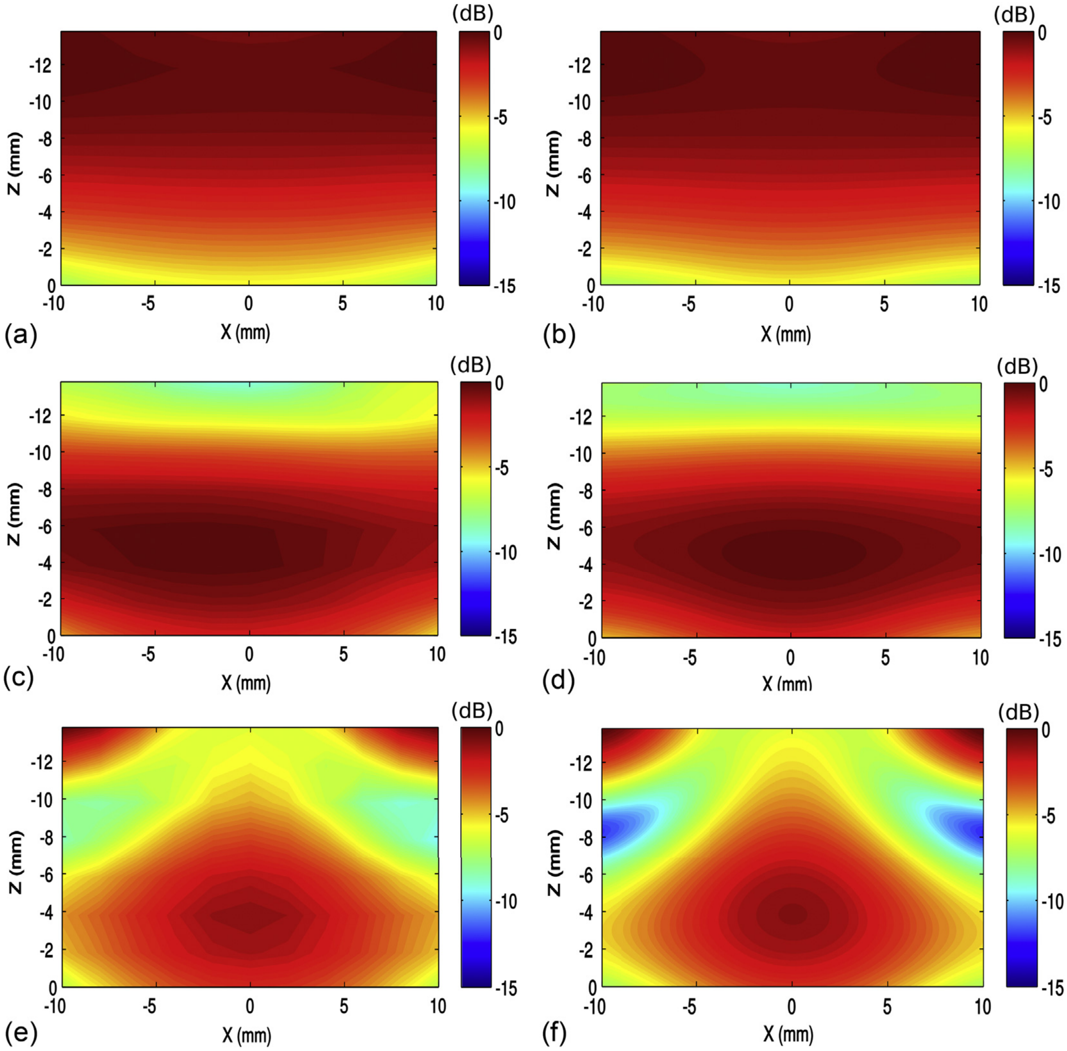

Fig. 5. Measured (Physical Model 1, left column) and Modeled (Computational Model 1, right column) pressure distribution along the horizontal plane at y = 0 in lip section (section 6 in Fig. 1) for frequency 5050 Hz (a–b), 10050 Hz (c–d), and 13050 Hz (e–f). Amplitudes were normalized by the maximal value on the plane. The z-axis corresponds to the main propagation direction with the outlet at z = 0.

second-order mode (7.3 kHz in Table 2). This indicates that the cutoff frequency is slightly changed from the estimated value due to the complex gemetry, and as a result, higher-order modes slightly affect the plane wave distribution (<1 dB) through the junction of the teeth for frequencies below the cutoff frequency.

Fig. 6 shows the comparisons of measured and modeled pressure amplitudes for the vertical plane ($y - z$ plane) at $x = 0$ in section 6 for the same frequencies as in Fig. 5 (5050 Hz, 10,050 Hz and 13,050 Hz). In the horizontal plane, the maximum pressure at $z = -12$ mm, near the upper teeth, shifted to $z = -4$ mm, near the outlet, as the frequency increased. The onset of 3D effects was also observed in the vertical plane for the frequency below the cutoff frequency of second-order mode. A strong asymmetry appeared in the z-direction above the cutoff frequency (10,050 Hz and 13,050 Hz). The dip observed at $z = -13$ mm, near the upper teeth, for 10,050 Hz became smaller for 13,050 Hz.

Measured and modeled pressure distributions along the horizontal (Fig. 5) and vertical (Fig. 6) plane are in overall agreement. This suggests that the applied modeling approach captures the acoustic pressure field inside the simplified rectangular vocal tract geometry when an acoustic sound source is placed at the inlet. The comparison between measured and modeled pressures is further quantified by considering the measured and modeled pressure-pressure transfer function defined in Eq. (20).

Measured and modeled transfer functions G_{lip} calculated with the pressure amplitude at two positions, near the outlet $(x, y = 0, z = -1.8)$ and 0.2 mm downstream from the teeth $(x, y = 0, z = -13.8)$, are shown in Fig. 7. The effect of the higher-order modes on the model outcome is assessed by comparing results with those of the plane wave model. The difference between the modeled plane wave transfer function and the measured transfer function increases with frequency and becomes noticeable (>0.6 dB) for frequencies above 4 kHz and significant (>1.6 dB) above 8 kHz. The difference in spectral amplitudes appeared

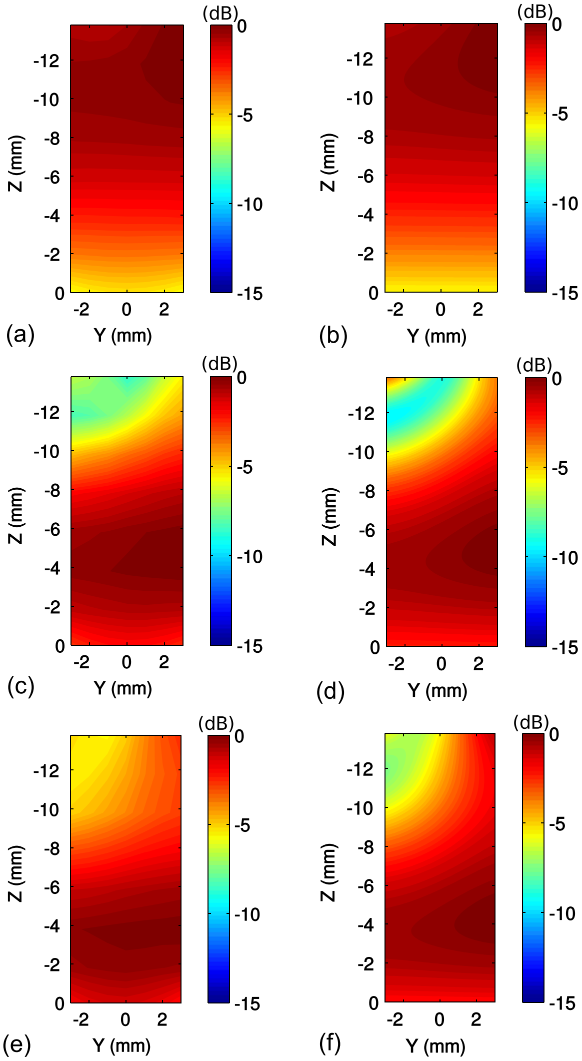

Fig. 6. Measured (Physical Model 1, left column) and modeled (Computational Model 1, right column) pressure distribution along the vertical plane at x = 0 in lip section (section 6 in Fig. 1) for frequency 5050 Hz (a–b), at 10050 Hz (c–d), and 13050 Hz (e–f). Amplitudes were normalized by the maximal value on the plane. The z-axis corresponds to the main propagation direction with the outlet at z = 0.

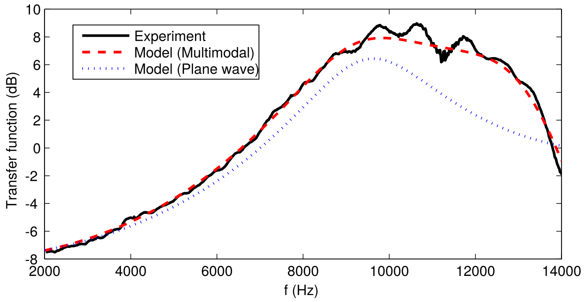

Fig. 7. Experimentally measured (Physical Model 1, ----) and theoretically modeled (Computational Model 1, plane wave: ....., multimodal: - - -) pressure-pressure transfer function G_{lip} as a function of frequency between positions z = -13.8 mm (near teeth) and z = −1.8 mm (near outlet). Experiment was conducted with the compression driver at the inlet of the mechanical replica.

for frequencies lower than it did in the case of vowels (4.5 kHz) [6]. This indicates that the smaller cross-sectional distances of sibilants have a greater effect on the pressure distribution than those of vowels. The transfer function obtained using the multimodal model approach matches with the measured transfer below 9 kHz and above 12 kHz. Between 9 and 12 kHz, although the general tendency of the measured transfer function is predicted, a difference (max. 1.4 dB) between modeled and measured transfer functions was observed. This difference is probably caused by experimental factors that were not considered in the model, such as plaster roughness (± 0.1 mm) or wall impedance [8].

3.2. Acoustic characteristics with source downstream from the constriction

When airflow is supplied to the inlet of the replica in the experiment, the sound source is assumed to be downstream from the constriction (from section 3–6 in Fig. 2 (b)) [1–3]. The sound spectrum measured at 30 cm from the outlet along the centerline of the replica is plotted in Fig. 8 (rectangle symbols). The replica produced a sound similar to subject's sibilant /s/ (shown in Fig. 1), which is characterized as broadband noise above 4 kHz with the first characteristic peak at 4 kHz and overall spectral peak at 8 kHz.

To assess the effect of the source position in the multimodal model (Eq. (21)–(23)), far-field pressure was calculated with Eq. (18) for four different sound source positions: the centers on all three axes for the dimensions of sections, *i.e.*, section 3 (0, 1.5, −18.5); section 4 (0, −0.6, −15.5); section 5 (0, −2.1, −14.5); section 6 (0, 0, −7). The modeled pressure spectra at 30 cm from the outlet along the centerline (z-axis) are plotted in Fig. 8. The first characteristic peak occurred around 4 kHz for all source positions. This peak frequency is in agreement with the frequency of the first characteristic peak observed in the spectrum measured with airflow. The amplitude of the first peak increased from 59 dB when the source is located at the center of section 6 (outlet section) to 63 dB when the source is located at section 3 (nearest section to the constriction). A second characteristic peak, associated with a spectral maximum for frequencies higher than the first characteristic peak, is observed in both modeled and measured spectra. The frequencies of the predicted second peaks are 7.7 kHz, 8.2 kHz, 8.3 kHz, and 8.5 kHz when the sources are positioned in sections 3, 4, 5, and 6, respectively. The peak amplitude varied from 50 dB to 70 dB. Meanwhile, the frequency of the second peak in the measured sound spectrum occurred at 8.1 kHz and yields 70 dB.

To explore the cause of the higher frequency peaks observed in the experiment, the position of the source was changed within section 4, *i.e.* the section center as well as positions shifted away from the section center are assessed in the model. In

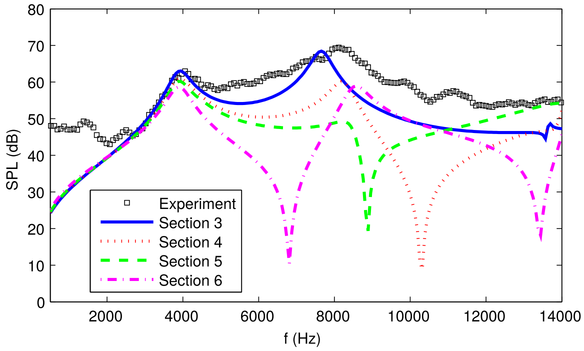

Fig. 8. Experimentally measured spectrum (Physical Model 2, □) and theoretically predicted spectra (Computational Model 2) with multimodal model and monopole source at the center of sections 3-6 (the center of section 3: —— , section 4: ....., section 5: - - -, and section 6: - · - ·) at 30 cm from the outlet along the centerline (z-axis). The experiment was conducted by supplying air to the mechanical replica. SPL is based on the reference level 20×10^{-6} Pa.

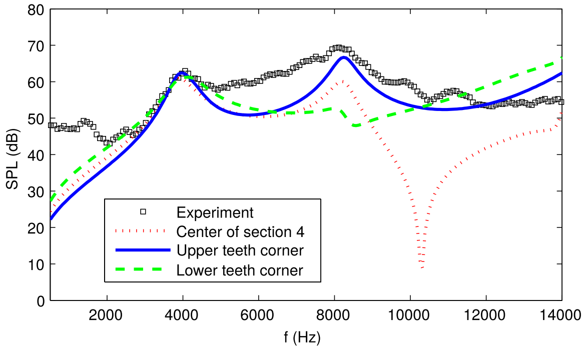

Fig. 9. Experimentally measured spectrum (Physical Model 2, □) and theoretically predicted spectra (Computational Model 2) with multimodal model at 30 cm from the outlet along the centerline (z-axis). A monopole source was located within section 4: near the upper teeth corner (Y = 2.9, Z = −15.1, ——— ; at the center (Y = 0, Z = −15.5, ...); and near the lower teeth corner (Y = −4.15, Z = −15.9, – – –). The experiment was conducted by supplying air to the mechanical replica. SPL is based on the reference level 20×10^{-6} Pa.

particular, positions near the wall of teeth (corner positions) are considered as illustrated in Fig. 4. The experimentally measured and theoretically modeled spectra for the sound source positioned near the upper teeth ($y = 2.9$, $z = -15.1$), center of section 4 ($y = 0$, $z = -15.5$), and near the lower teeth ($y = -4.15$, $z = -15.9$) are shown in Fig. 9. The source near the upper and lower teeth was located at 0.1 mm from the corner of the wall in section 4. By shifting the source from lower teeth corner (y = −4.2, z = −15.9) to upper teeth corner (y = −2.9, z = −15.1), the amplitude of the maximum peak at 8 kHz was increased from 52 to 67 dB. Best spectral match between experimentally measured and theoretically modeled spectra was obtained when the source was positioned near the upper teeth corner. In addition, the overall spectral peak observed at 8 kHz matches the spectra measured with European Portuguese speakers' /s/ [5] and Shadle's /s/-like model [2]. This indicates that the studied geometrical approximation with a source near the wall of upper teeth generates main spectral features (two spectral peaks) of sibilant /s/ for this speaker (see Fig. 1), and indeed confirms previous findings in literature [2,3,17].

General tendencies of the measured sound spectra are captured by applying the multimodal model when the sound source is located near the upper teeth in section 4. In particular, the first and second characteristic peaks of sibilants, which were not observed in Motoki's model [8], were observed by placing the sound source near the upper teeth wall. However, the amplitude in the frequency range 4.5–12 kHz was lower than that measured in the flow experiment (approximately 5–10 dB). This might be due to the characteristics of the flow sound source. Impingement of oscillating jet flow on a wall generates not only a monopole source but also dipole and quadrupole sources, owing to the airflow velocity fluctuations [19]. Therefore, further agreement might be achieved by accounting for dipole or quadrupole source distributions in multimodal theory. In addition, the sharp edges in the replica potentially generated spurious sound in the flow experiment, which is not considered in the modeling, and it is desirable to improve for future study. Moreover, further agreement on the spectrum might also be expected by considering resonances upstream from the air tube or by considering the rounding of the teeth [17].

The spectrum resulting from the multimodal model with the source positioned near the upper teeth in section 4 is compared with that resulting from a plane wave model in Fig. 10. The discrepancy (>5 dB) between multimodal and plane wave spectra appeared above 6 kHz when the source was located near the upper teeth in section 4 ($y = 2.9$, $z = -15.1$). This discrepancy over 5 dB is significant for listeners with normal hearing [20]. Nevertheless, the frequency of the first characteristic peak was predicted with the plane wave model in the same way as Howe and McGowan's one-dimensional model [3]. In addition, the second characteristic peak was predicted by the multimodal model for this source location. By increasing the number of modes

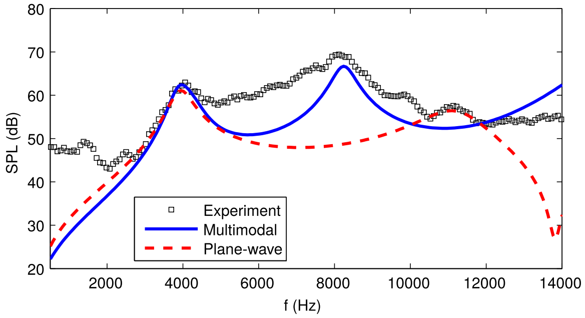

Fig. 10. Experimentally measured (Physical Model 2, □) and theoretically modeled spectra (Computational Model 2, plane wave mode: – – – , multimodal: ———) for the source near the upper teeth corner (Y = 2.9, Z = −15.1), observed at 30 cm from the outlet along the centerline (z-axis). The experiment was conducted by supplying air to the mechanical replica. SPL is based on the reference level 20×10^{-6} Pa.

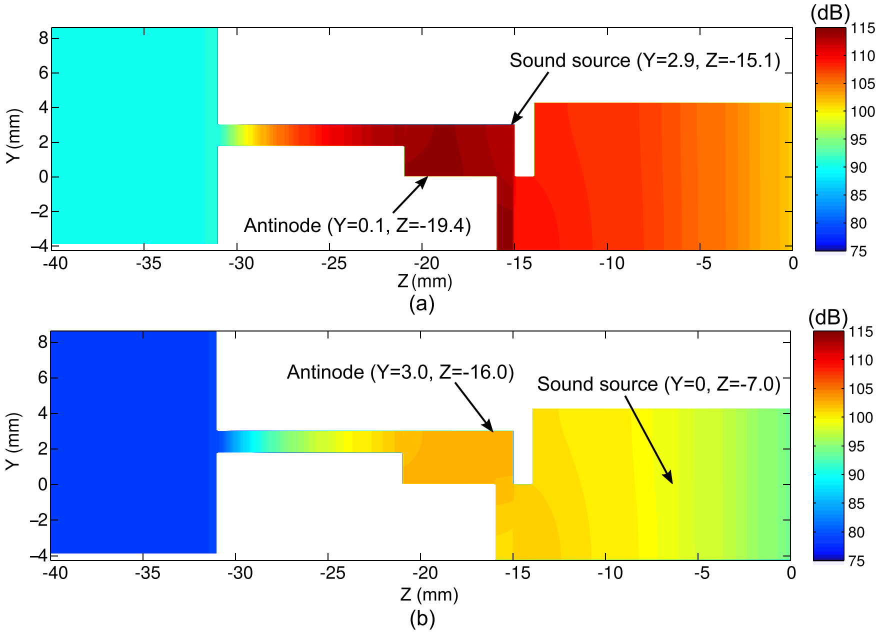

Fig. 11. Pressure distribution predicted by the multimodal model (Computational Model 2) along the vertical center plane (x = 0) of a portion of the vocal tract geometry for 4 kHz. The monopole source was positioned near the upper teeth corner in section 4 (a), and at the center of section 6 (b).

in multimodal modeling, the peak observed at 11 kHz for plane wave model was shifted to the overall spectral peak at 8 kHz. This suggests that the main source in the simplified vocal tract approximation of /s/ is generated near the wall of upper teeth, and higher-order modes are needed to capture this high frequency behavior.

The modeled pressure distributions along the vertical center plane ($x = 0$) of the vocal tract geometry for 4 kHz (first characteristic peak) are depicted in Fig. 11 for the source positioned near the upper teeth corner and the center of section 6. The maximum pressure occurs between the constriction and the upper teeth for both source positions. This suggests that the antinode of the first characteristic peak (4 kHz in Figs. 8–10) appeared within the cavity between the constriction and the upper teeth.

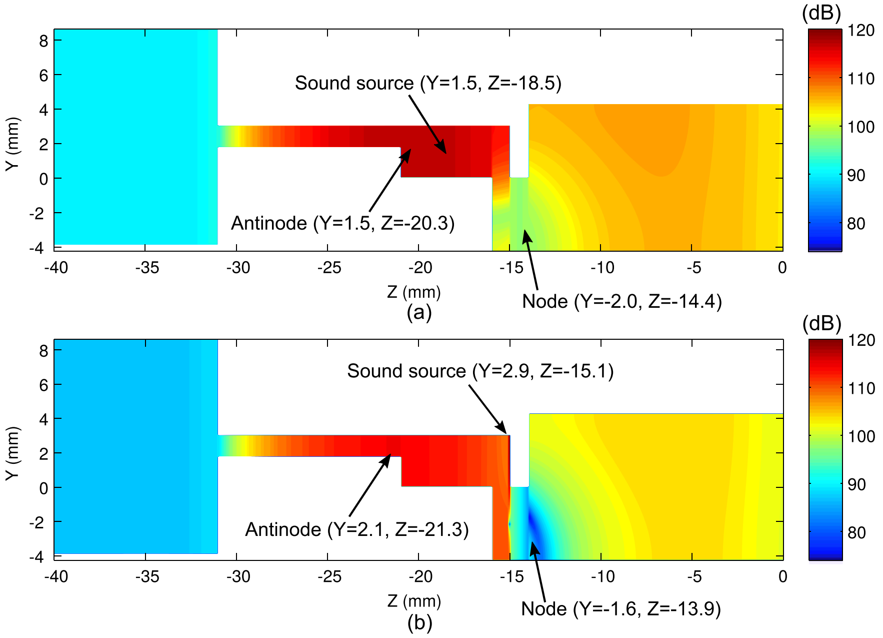

Fig. 12. Pressure distribution predicted by the multimodal model (Computational Model 2) along the vertical center plane (X = 0) for the frequency of the second peak 7.6 kHz (a) and 8.2 kHz (b). The monopole source was positioned at the center of section 3 for (a) and near the upper teeth corner for (b).

In other words, the main resonance frequency is determined by the distance between the outlet and the exit from the constriction (the upstream end of section 3). This finding is consistent with Shadle's simplified model [2] that showed the dependence of the main resonance frequency on the position of the constriction. Note that, positioning the source within section 3 amplified the sound pressure in the entire vocal tract. Therefore, the amplitude of the first peak for the source within section 3 was larger than the amplitude for the source downstream from section 6.

The pressure distributions along the vertical center plane are shown in Fig. 12 for the frequency of the second peak 7.6 and 8.2 kHz which appeared when the sound source was located at the center of section 3 and near the upper teeth in section 4, respectively. At the frequency of the second peak, node and antinode appeared near the constriction exit and downstream of the lower teeth, respectively, for both source positions. Meanwhile, by changing the source location from section 3–4, the amplitude was decreased and position of the node and antinode was shifted towards the downstream and upstream, respectively. These results indicate that the main source in the simplified vocal tract of /s/ appears near the upper teeth wall, and show that multimodal approach allows us to capture the behavior of nodes and antinodes in the pressure distribution inside the vocal tract for the sibilant /s/. The relationship between these nodes or antinodes and the peak frequencies will be the subject of further study. In addition, investigation for the effect of source type (*i.e.* dipole or quadrupole) on the node and antinode positions is needed.

4. Conclusion

An acoustic driver and flow supply were applied to the inlet of a simplified geometrical approximation of a vocal tract for sibilant /s/. Then, multimodal theory was applied to the simplified geometry for two different source positions, at the vocal tract inlet and downstream from the constriction representing the sibilant groove. The predicted and measured pressure distributions for the source at the inlet agreed well when acoustic higher-order modes were taken into account. The first characteristic peak of sibilant /s/ measured for airflow supply was reproduced by placing the source downstream from the constriction (centers of sections 3–6). Moreover, general tendencies of the measured spectra are obtained with the source near the upper teeth wall. This result indicates that the main source in the simplified vocal tract of /s/ appears near the upper teeth wall. Note that the first characteristic peak was captured by the plane wave model in the same way as with the one-dimensional model [3]. However, the comparison with flow experiment suggests that higher-order modes have to be taken into account to be able to capture the higher mode peak. Indeed, the second peak at 8 kHz was also observed in a spectrum of European Portuguese speakers' /s/ [5] and previous simplified model [2], and it is desirable to study the mechanisms of the second peak in future study.

For the frequency of the first peak, the maximum value in the pressure distribution appears within the cavity between the constriction and the upper teeth. The maximum value remained in the same cavity when the position of the source was varied from section 3 to section 6. This result shows that the antinode of the first characteristic peak appears within the cavity between the constriction and the upper teeth. For the frequency of the second peak, node and antinode appeared near the constriction exit and downstream of the lower teeth, and position of the node and antinode was shifted downstream and upstream, respectively, by changing the source location. These results indicate that the multimodal approach allows us to capture the nodes and antinodes in the pressure distribution inside the vocal tract as well as the amplitude and frequency of the peaks observed in the subject's /s/.

This is the first detailed description of the underlying mechanism for more than one characteristic peak as observed for sibilant /s/ pronounced by the speaker. In future work, it is necessary to study the nodes and antinodes in the vocal tract geometry while airflow is supplied, to validate current findings. In addition, further investigation on higher frequency peaks can be achieved by modeling dipole or quadrupole source distributions in multimodal theory.

Acknowledgements

This work was supported by the EU-FET grant [EUNISON grant number 308874]; the JSPS Grant-in-Aid for JSPS Research Fellows [grant number JP15J00413]; MEXT as Priority Issue on Post-K computer [Project ID: hp160218]; and the Program for Leading Graduate Schools of MEXT [Humanware Innovation Program].

References

[1] K.N. Stevens, Airflow and turbulence noise for fricative and stop consonants: static considerations, J. Acoust. Soc. Am. 50 (1971) 1180–1192.
[2] C.H. Shadle, The Acoustics of Fricative Consonants, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 1985.
[3] M.S. Howe, R.S. McGowan, Aeroacoustics of [s], Proc. R. Soc. A 461 (2005) 1005–1028.
[4] P. Badin, Fricative consonants: acoustic and X-ray measurements, J. Phon. 19 (1991) 397–408.
[5] L.M.T. Jesus, C.H. Shadle, A parametric study of the spectral characteristics of european Portuguese fricatives, J. Phon. 30 (2002) 437–464.
[6] R. Blandin, M. Arnela, R. Laboissiere, X. Pelorson, O. Guasch, A. Van Hirtum, X. Laval, Effects of higher order propagation modes in vocal tract like geometries, J. Acoust. Soc. Am. 137 (2015) 832–843.
[7] K. Motoki, P. Badin, X. Pelorson, H. Matsuzaki, A modal parametric method for computing acoustic characteristics of three-dimensional vocal tract models, in: Proc. of 5th Seminar on Speech Production: Models and Data, 2000, pp. 325–328.
[8] K. Motoki, A parametric method of computing acoustic characteristics of simplified three-dimensional vocal-tract model with wall impedance, Acoust. Sci. Tech. 34 (2013) 113–122.
[9] A. Van Hirtum, X. Pelorson, O. Estienne, H. Bailliet, Experimental validation of flow models for a rigid vocal tract replica, J. Acoust. Soc. Am. 130 (2011) 2128–2138.
[10] J. Cisonni, K. Nozaki, A. Van Hirtum, S. Wada, A parameterized geometric model of the oral tract for aero acoustic simulation of fricatives, Int. J. Inf. Elec. Eng. 1 (2011) 223–228.

[11] A.D. Pierce, Acoustics: an Introduction to its Physical Principles and Applications, Acoust. Soc. Am., New York, 1989.
[12] J. Kergomard, A. Garcia, G. Tagui, J. Dalmont, Analysis of higher order mode effects in an expansion chamber using modal theory and equivalent electrical circuits, J. Sound. Vib. 129 (1989) 457–475.
[13] V. Pagneux, N. Amir, J. Kergomard, A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation, J. Acoust. Soc. Am. 100 (1996) 2034–2048.
[14] N. Amir, V. Pagneux, J. Kergomard, A study of wave propagation in varying cross-section waveguides by modal decomposition. Part II. Results, J. Acoust. Soc. Am. 101 (1997) 2504–2517.
[15] R.T. Muehleisen, Reflection, Radiation, and Coupling of Higher Order Modes at Discontinuities in Finite Length Rigid Walled Rectangular Ducts, Ph.D. thesis, Pennsylvania State Univ., State College, 1996.
[16] K. Nozaki, T. Yoshinaga, S. Wada, Sibilant /s/ simulator based on computed tomography images and dental casts, J. Dent. Res. 93 (2014) 207–211.
[17] R.S. McGowan, M.S. Howe, Compact Green's functions extend the acoustic theory of speech production, J. Phon. 35 (2007) 259–270.
[18] Y. Fujiso, K. Nozaki, A. Van Hirtum, Estimation of minimum oral tract constriction area in sibilant fricatives from aerodynamic data, J. Acoust. Soc. Am. 138 (2015) EL20–EL25.
[19] M.J. Lighthill, On sound generated aerodynamically I, general theory, Proc. R. Soc. A 1952 (1952) 564–587.
[20] B.B. Monson, A. Lotto, B.H. Story, Detection of high-frequency energy level changes in speech and singing, J. Acoust. Soc. Am. 135 (2014) 400–406.