Journal of Sound and Vibration 476 (2020) 115337

Contents lists available at ScienceDirect

Journal of Sound and Vibration

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

Acoustic modeling of fricative /s/ for an oral tract with rectangular cross-sections

T. Yoshinaga [a], A. Van Hirtum [b,*], K. Nozaki [c], S. Wada [d]

[a] Toyohashi University of Technology, 1-1 Hibarigaoka, Tenpaku, Toyohashi, 441-8580, Aichi, Japan
[b] LEGI, UMR CNRS 5519, Grenoble Alpes University, 11 rue des Mathématiques (BP46), 38058, Grenoble, France
[c] Osaka University Dental Hospital, 1-8 Yamadaoka, Suita, 565-0871, Osaka, Japan
[d] Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, 560-8531, Osaka, Japan

ARTICLE INFO

Article history:
Received 23 April 2019
Revised 13 March 2020
Accepted 17 March 2020
Available online 18 March 2020
Handling Editor: O Guasch

Keywords:
Speech production
Multimodal theory
Oral tract simplification
Sibilant fricative

ABSTRACT

Fricative /s/ is known to be pronounced by jet generation and subsequent impact on walls of the oral cavity. The prediction of acoustic characteristics of /s/ is an ongoing research topic due to the aeroacoustic nature of the sound source. In this study, acoustic characteristics are modeled using the multimodal theory with monopole and dipole sources positioned in the oral tract waveguide. The oral tract geometry of fricative /s/ was simplified by concatenating rectangular channels whose cross-sectional areas and vertical height are derived from medical imaging. To validate the model accuracy, transverse and sagittal directivity patterns (49 cm every 15°) were measured for flow supplied to a realistic oral tract replica. Comparison between measured and modeled spectra showed that the modeling with the dipole source predicted the pressure amplitude within a discrepancy of ±5 dB up to 14 kHz. Modeled acoustic directivity patterns using a dipole source reflected main tendencies observed on measured directivity patterns in both the transverse and sagittal planes. The proposed modeling approach enables a systematic analysis of the high frequency (>5 kHz) acoustic characteristics as a function of geometrical details for speech production due to an aeroacoustic sound source.

1. Introduction

The cause of acoustic characteristics for speech sounds has been widely investigated by modeling the phenomena occurring in the vocal tract. Fant [1] pioneered acoustic modeling for vowels using vocal tract waveguide geometries derived from X-ray images. To model vowel production, a sound source was situated at the larynx where vocal folds are vibrating. Stevens [2] modeled sound generation for fricative consonants assuming a sound source of turbulent flow at the constricted part of the oral tract. Later on, Shadle [3] proposed a geometrical approach to study fricatives consisting of the combination of a constricted channel with an obstacle inserted downstream from the constriction to investigate the relationship between geometrical parameters and acoustic characteristics of fricatives.

Howe and McGowan [4] argued that the main sound source of /s/ is generated at the gap between incisors and proposed a one-dimensional aeroacoustic model. In this model, the one-dimensional vocal tract area function was simplified using only cross-sectional areas at four positions, and the first characteristic spectral peak of /s/ around 4 kHz [5] was predicted. In addition,

* Corresponding author.
E-mail address: annemie.vanhirtum@univ-grenoble-alpes.fr (A. Van Hirtum).

https://doi.org/10.1016/j.jsv.2020.115337
0022-460X/

Krane [6] proposed an aeroacoutic theory for a constricted duct and predicted the acoustic characteristics of unvoiced speech sound. However, these theories consider the one-dimensional plane wave mode, so that such theories can be applicable only for sounds below 5 kHz.

More recently, by using cross-sectional shapes at five positions, differences of the acoustic characteristics between /s/ and /ʃ/ were reproduced experimentally [7]. Flow simulations for the vocal tract geometries of /s/ showed that the sound source is mainly generated near the surfaces of upper and lower incisors [8,9]. However, the relationship between the generated source and acoustic characteristics in far-field is still unclear.

To investigate the relationship between source location and the acoustic characteristics, multimodal acoustic modeling was applied to the simplified vocal tract model of /s/ [10]. The multimodal modeling considers higher-order modes of acoustic propagation and enables to estimate the acoustic amplitude in higher frequency ranges compared with the plane wave analysis (*i.e.* above 5 kHz) [11,12]. Therefore, this modeling can be used to advance the understanding of speech sounds where the radiated acoustic energy peaks above 5 kHz. By applying the multimodal modeling to the simplified vocal tract of /s/ proposed in Ref. [7], two spectral peaks characterizing /s/ were predicted in the frequency range up to 9 kHz [10]. However, the discrepancy between the measured and the modeled spectrum exceeds 9.9 dB, which at least for vowels is shown to be perceptually significant [13]. This is caused by the use of a monopole source in the modeling, and it is hypothesized that the use of a dipole source, which represents the characteristics of sound source generated by flow impingement on a wall [2–4,6], might improve the model accuracy. In addition, the use of dipole source is a strong motivation for the multimodal modeling which can properly express the sound propagation for the frequency range of flow noises.

Meanwhile, the multimodal acoustic model approach is also used to investigate the directivity pattern of speech sounds up to 15 kHz. Blandin et al. [14,15] used vowel vocal tract geometries and found that the low-frequency symmetrical directivity pattern is affected mainly in the frequency range above 7 kHz. For the fricative /s/ uttered by English speakers, measurements with a microphone antenna showed that amplitudes at the center, *i.e.* in the speaker's direction, increases with up to 20 dB [16]. Measurements with a realistic vocal tract replica of /s/ with and without lips revealed that the directivity pattern is largely affected by the lips and sound source location [17]. However, the multimodal modeling has not been applied to the vocal tract geometry of /s/ to study its directivity pattern.

Therefore, in this study, we apply the multimodal model approach (Section 2.1) to a simplified oral tract geometry of /s/ in order to predict both the far-field sound spectrum and directivity pattern of /s/. In the previous study [17], the directivity pattern of /s/ was investigated by comparing the experiments with the flow source and artificial acoustic source at the glottis. Hence, there was no intention to model the directivity pattern of /s/ with the multimodal theory. Meanwhile, in Ref. [10], the multimodal modeling was developed to predict the acoustic characteristics of /s/ using the monopole source. However, the dipole source was not used, and the directivity pattern was not modeled in the paper. Therefore, in this study, we develop the multimodal modeling of /s/ with the dipole source which is more physically accurate to represent the unsteady aerodynamic forces in the vicinity of the incisors, and the directivity pattern is modeled and compared with the experimental results.

The oral tract geometry was simplified as a concatenation of rectangular cavities with constant cross-section and of length 0.5 mm each, in order to represent the changing area of the realistic oral tract geometry up to the lip edge (Section 2.2, 2.3). The accuracy of the model outcome for different sound sources was compared with respect to the realistic geometry (Section 2.4). The directivity measurements were conducted with the flow supplied to a realistic oral tract replica and a rectangular baffle set in two positions, posterior and anterior. The posterior baffle position represents a normal lip-face junction position. For the anterior baffle position, the position was shifted from the lip-face junction to the trailing edge of the lips so that the outlet geometry becomes comparable with the modeling. Directivity patterns measured with the anterior baffle were compared to the modeled ones (Section 3) and modeled and measured results are discussed (Section 4).

2. Method

2.1. Multimodal modeling

Multimodal acoustic theory has been developed by several researchers [18–20] and applied to vocal tract geometries [10–12] to clarify the relationship between vocal tract geometry and acoustic characteristics. The amplitude of sound pressure $p(\mathbf{x})$ and particle velocity vector $\mathbf{v}(\mathbf{x})$ in waveguides are calculated from velocity potential $\phi(\mathbf{x})$ as

$$p(\mathbf{x}) = \mathrm{j}\omega\rho\phi(\mathbf{x}), \tag{1}$$

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

with angular frequency ω, air density ρ and spatial position vector $\mathbf{x} = (x_1, x_2, x_3)$ where x_1 indicates the propagation direction. The velocity potential satisfies the Helmholtz equation

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

where $k = \omega/c$ denotes the free field wave number and c is the speed of sound in air. The solution of the Helmholtz equation becomes a summation of an infinite number of propagation modes $\psi_{mn}(x_2, x_3)$ weighted by forward and backward propagation

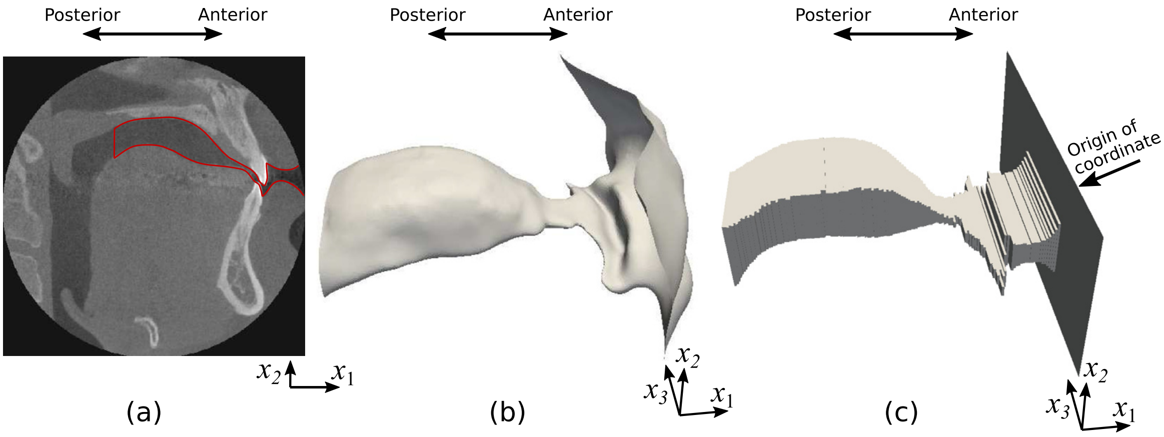

Fig. 1. Oral tract simplification. (a) Computed tomography image of subject pronouncing /s/. The red line shows the region where the brightness values changed. (b) Extracted surface of oral tract geometry. (c) Simplified oral tract geometry with rectangular cross-sections. The cross-sectional areas and vertical heights are taken for the transverse plane (x_2-x_3) every 0.5 mm along axis x_l. The origin of the coordinate system is at the center of the lip outlet. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

amplitudes a_{mn}, b_{mn}:

$$\phi(\mathbf{x}) = \sum_{m,n=0}^{\infty} \psi_{mn}(x_2, x_3)\{a_{mn}\exp(-\gamma_{x_1,mn}x_1) + b_{mn}\exp(\gamma_{x_1,mn}x_1)\}, \tag{4}$$

where m and n are the number of modes in the x_2-direction and x_3-direction, respectively. The propagation mode $\psi_{mn}(x_2, x_3)$ 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_2} and L_{x_3};

$$\psi_{mn}(x_2, x_3) = \frac{\cos(m\pi x_2/L_{x_2})}{\sqrt{L_{x_2}\sigma_m}}\frac{\cos(n\pi x_3/L_{x_3})}{\sqrt{L_{x_3}\sigma_n}}, \tag{5}$$

where σ_m, σ_n are 1 ($m, n = 0$) or 1/2 ($m, n \geq 1$). Note that this method can be applicable for arbitrary cross-sectional shapes by using other solutions of two dimensional Helmholtz equation [12] The propagation constant $\gamma_{x_1,mn}$ is given as

$$\gamma_{x_1,mn} = \sqrt{\left(\frac{m\pi}{L_{x_2}}\right)^2 + \left(\frac{n\pi}{L_{x_3}}\right)^2 - k^2}. \tag{6}$$

To estimate the pressure amplitude in the oral tract, the geometry is simplified as a concatenation of rectangular waveguides and the relationship among pressure, velocity and acoustic impedance is calculated at each cross-section. The number of modes is truncated based on the cutoff frequency of each cross-section. In this paper, the number of modes was considered up to eight. Details of the calculation procedure are reported in [10,11].

The boundary condition at the outlet of the oral tract waveguide needs to account for the presence of the face, so that the radiation impedance was computed assuming an infinite baffle [21]. The impedance at the inlet was set as the non-reflection boundary condition. The pressure amplitude outside of the oral tract waveguide was calculated with the Rayleigh-Sommerfield integral [22] for the radiation impedance and particle velocity at the outlet. These equations were implemented in MATLAB (Mathworks, Natick, USA).

2.2. Oral tract simplification

The oral tract geometry of fricative /s/ is simplified as a concatenation of rectangular channels with uniform cross-sectional area with the following procedure. Images of the oral tract during pronunciation of /s/ were obtained using computed tomography (CT) scan [23]. The subject is a 32-yr-old male native Japanese speaker with normal dentition (Angle Class I) and no speech disorder in self-report. The CT scan was taken in 9.6 s while the subject sustained /s/ in seated position without vowel context. A surface of the oral tract was extracted based on brightness values of the scan with the software itk-SNAP [24]. The CT scan and extracted surface are shown in Fig. 1(a) and (b), respectively. The surface was extracted from the lip tip up to the anterior part of pharynx.

Then, oral tract's cross-sectional areas and vertical heights (L_{x_2}) along the coronal plane ($x_2 - x_3$) were measured every 0.5 mm along the posterior-anterior propagation direction (axis x_1). The transverse length of the simplified tract (L_{x_3}) was then obtained by dividing cross-sectional area with L_{x_2}. When the cross-section along the coronal plane reached far-field (no wall

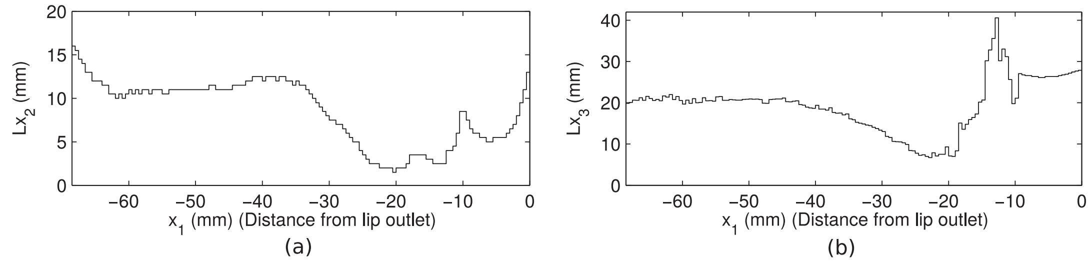

Fig. 2. Dimensions of the simplified oral tract model. Vertical height Lx_2 and transverse length Lx_3 for each rectangular cross-section are plotted for (a) and (b), respectively.

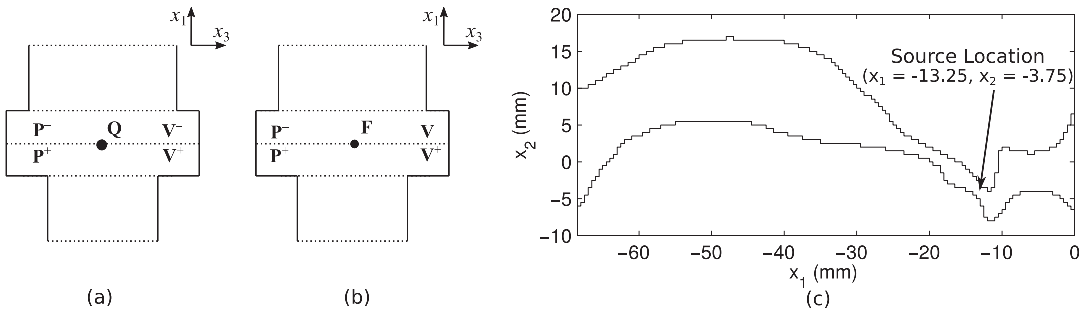

Fig. 3. Position of the sound source. Positions of flow rate fluctuation of monopole and force of dipole source are depicted in (a) and (b), respectively. Position of sound source in sagittal plane (X_1-X_2) is shown in (c).

along axis x_3), the cross-sectional area was estimated with L_{x_3} obtained at 0.5 mm posterior position, and L_{x_2} and L_{x_3} were calculated up to the tip of lip surface. Then, uniform channels of length $L_{x_1} = 0.5$ mm and with rectangular cross-section of dimensions L_{x_2} and L_{x_3} were concatenated along the centerline of the oral tract (see Ref. [25] for details). The simplified oral tract geometry is depicted in Fig. 1 (c), and dimensions L_{x_2} and L_{x_3} are plotted in Fig. 2. The origin of the coordinate system is located at the center of the lip outlet. The total number of cross-sections is 137 so that the total length from inlet to lip outlet yields 68.5 mm.

2.3. Sound source modeling

To model the sound generation of /s/, the sound source was situated near the upper and lower incisors. When the jet flow is generated from the constricted channel of the oral tract, monopole, dipole, and quadrupole sound sources are generated whereas the dipole source is argued to be dominant for the sound generation of /s/ [2,3]. Therefore, the multimodal model approach is applied using a dipole source as well as a monopole source [10] in order to consider the imprint of source characteristics on the far-field spectrum. For the monopole source, the cross-section was divided into an upstream (anterior) section and a downstream (posterior) section along axis x_1, and modal pressure **P** and velocity **V** up to its cutoff frequency are calculated at each cross-section as

$$\mathbf{P}^+ - \mathbf{P}^- = 0, \tag{7}$$

$$\mathbf{V}^+ - \mathbf{V}^- = \mathbf{Q}, \tag{8}$$

$$\mathbf{Q} = Q\boldsymbol{\psi}, \tag{9}$$

where + and − represent the variables downstream (posterior) and upstream (anterior) of the source section, and Q denotes the amplitude of the fluctuating volume flow rate. The flow source was positioned at the center in the transverse direction (axis x_3) as depicted in Fig. 3 (a). Although modeling with the monopole source is physically inaccurate for the sound generated by flow fluctuations in the static vocal tract, we used it to examine influences of the source characteristics on the radiating sound.

For the dipole source, an aerodynamic flow force F is applied to the modal pressure in the anterior-posterior direction (axis x_1)

$$\mathbf{P}^+ - \mathbf{P}^- = \mathbf{F}, \tag{10}$$

$$\mathbf{V}^+ - \mathbf{V}^- = 0, \tag{11}$$

$$\mathbf{F} = F\boldsymbol{\psi}, \tag{12}$$

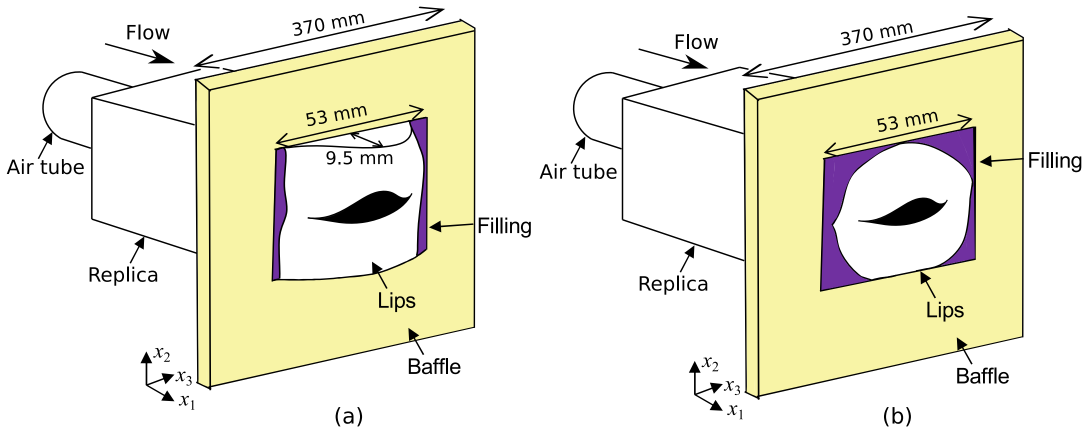

Fig. 4. Oral tract replica and baffle. The baffle is positioned at posterior position in (a) and at anterior position in (b).

as depicted in Fig. 3 (b).

Firstly, the amplitude of the source was expressed by a white noise with constant amplitude in the frequency range for both monopole and dipole sources. The amplitude of the sources was adjusted to reproduce sound pressure level (SPL) of the first characteristic peak observed in the experiment. Then, the diffractive source spectrum [4], for which the amplitude decreases with frequency (approximately −0.5 dB/kHz), was implemented for the dipole source to express the frequency characteristics of the flow source. Based on previous findings [10,26] both sources were positioned at the gap upstream from the upper incisor ($x_1 = -13.5$ mm, $x_2 = -3.75$ mm) as depicted in Fig. 3 (c).

2.4. Experimental setups

To validate the model accuracy, experimental measurements were conducted with a realistic oral tract replica and flow supply. The realistic oral tract shown in Fig. 1 (b) was constructed using rapid prototyping of acrylic resin (Objet30Pro, Stratasys, USA; accuracy: ±0.1 mm). A rectangular baffle (370 × 370 mm) was attached at the anterior and posterior positions of the replica outlet as depicted in Fig. 4. When the baffle is at the posterior position (Fig. 4 (a)), the baffle is aligned with the lip junctions and the tip of the lip protrudes 9.5 mm with respect of the baffle. When the baffle is at the anterior position (Fig. 4 (b)), the baffle is aligned with the tip of the lip surface, and the gap between the lip region and the baffle was filled with modeling clay to obtain a flat baffle at the outlet. The protruding lip surfaces with the posterior baffle position are more realistic compared to that with the anterior position. However, the radiation impedance in the current method cannot express the protruding surface from the infinite baffle. Therefore, we changed the lip configuration with the anterior position of the baffle.

Steady airflow was provided to the replica by a compressor (GA7, Atlas Copco), a pressure regulator (type 11-818-987, Norgren), a manual valve, a mass-flow meter (model 4045, TSI), and a settling chamber ($40 \times 40 \times 50$ cm^3) all connected with air tubes of inner diameter 10 mm. The settling chamber was used to dissipate turbulence and sound generated upstream from the replica inlet. The flow rate was fixed with the subject's flow rate for /s/ production at 400 cm^{3}s^{-1} [7].

Sound generated by the replica was measured with a single 1/2-inch microphone with flat frequency response up to 20 kHz (type 4192 and pre-conditioner type 5935L, B&K) in an anechoic chamber [27] as depicted in Fig. 5. The replica was set at 82 cm from the floor and the microphone was positioned at the same height as the replica. The distance between the lip outlet and the microphone was 49 cm. The position of the microphone was shifted from 0° to 180° in steps of 15° as shown in Fig. 5 (b) to measure the sound in a transverse plane ($x_1 - x_3$). Then, the replica was turned on its side, and the sound was measured in sagittal plane ($x_1 - x_2$) from 0° to 180° in steps of 15° as shown in Fig. 5 (c). The position 90° donates the direction of axis x_1 for both planes. If the wavelength of measured sound exceeds the distance between the sound source and measurement position, we cannot accurately obtain the acoustic nodes and anti-nodes in the directivity measurement. Therefore, we estimated the minimum reliable frequency of directivity measurement 705 Hz as the ratio between the sound speed (345 m/s) and the measurement distance (0.49 m).

At each spatial position, the sound was recorded during 3 s using a data acquisition system (PXI0MIO 16XE, National Instruments, USA) with sampling frequency 44.1 kHz. Sound spectra were computed using Welch's method using 60 time segments of 1024 samples, Hanning window and 30% overlap between segments. Acoustic pressure levels (dB SPL) were calculated based on a reference level of 20×10^{-6} Pa.

3. Results

Measured sound spectra at 49 cm along axis x_1 (90° in the transverse plane) are plotted in Fig. 6 for both baffle positions. Modeled spectra with monopole and dipole sources (white noise characteristics) are also plotted in Fig. 6. The sound amplitude increased with frequency from 4 kHz (20 dB) and reached a maximum of 46 dB at 6 kHz. For frequencies greater than 6 kHz

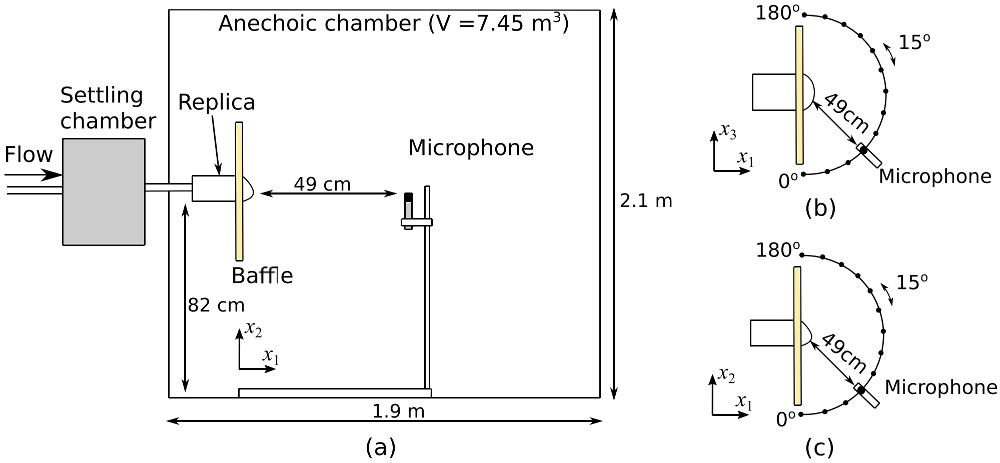

Fig. 5. Experimental setups. (a) The oral tract replica with supply and measurement setups in the anechoic chamber. The measurements were conducted in transverse plane (x_1 - x_3) (b) and in sagittal plane (x_1 - x_2) (c) every 15° from 0° to 180°.

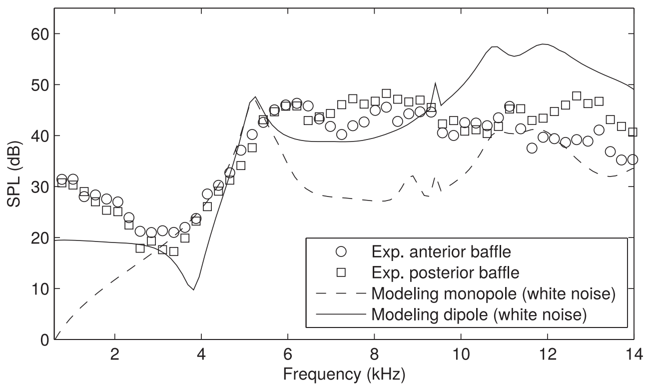

Fig. 6. Sound spectra measured at 90° in the transverse plane (x_1 - x_3) with baffle at anterior (○) and posterior (□) positions and modeled spectra for monopole (- - -) and dipole source (white noise) (——). The distance between the replica and the measurement position is 49 cm.

a broadband spectrum is observed whose amplitude remained greater than 40 dB for the posterior baffle position. Shifting the baffle to the anterior position decreased the amplitude (<10 dB) for frequencies greater than 7 kHz. Modeled pressure amplitudes increased with frequency up to a first spectral peak of 47 dB at 5.3 kHz for both sound sources. Although the slope of the first peak (between 4 and 5.3 kHz) of the dipole source was steeper than those of the monopole and experiments, amplitudes below the first peak (< 3 kHz) of the dipole were closer to those of the experiments compared to the monopole source. For frequencies greater than the first peak (5.3 kHz), amplitudes of dipole agreed with those of experiments up to 10 kHz, whereas amplitudes of monopole were 15 dB smaller than those of experiments. In contrast, the dipole source overestimated amplitudes above 10 kHz, whereas amplitudes of the monopole agreed with those of experiments above 11 kHz.

The modeled spectra with dipole source of white noise and diffractive source characteristics are plotted in Fig. 7. By implementing the diffractive source, the amplitudes above the first peak (5.3 kHz) were decreased so that amplitudes above 10 kHz are more in agreement with measured data than when the dipole with white noise characteristics was used. The maximum discrepancy of amplitudes between the experiments and dipole source with diffractive source was approximately 5 dB in the frequency range above 5 kHz.

To visualize the directivity pattern at each frequency, normalized pressure amplitudes measured in the transverse plane ($x_1 - x_3$) and in the sagittal plane ($x_1 - x_2$) are depicted in Fig. 8 for the case with the baffle at posterior (a), (b) and anterior (c), (d) positions, respectively. The pressure amplitude was normalized by the spatial mean at each frequency. The pressure amplitude was centered at 90° above approximately 4 kHz in the transverse plane, whereas the amplitude was centered at 90° above 6 kHz in the sagittal plane for the posterior baffle position. When the baffle was at the anterior position, the large amplitude shifted to 120° in the frequency range from 9 up to 12 kHz in the transverse plane, whereas no significant directivity pattern appeared in the sagittal plane.

Normalized modeled pressure amplitudes in the transverse plane ($x_1 - x_3$) and in the sagittal plane ($x_1 - x_2$) with the monopole source and dipole source (diffractive source) are shown in Fig. 9. The modeled pressure amplitudes are also plotted in every 15°. In the transverse plane, the largest amplitude appeared at 90° above 10 kHz for both sources, whereas in the range from 8 up to 10 kHz the amplitude was more pronounced towards the sides (0° and 180°) enveloping a trough from 40° to 140° for monopole source. For the dipole source, the amplitude shift towards 180° around 9 kHz was smaller than that of the

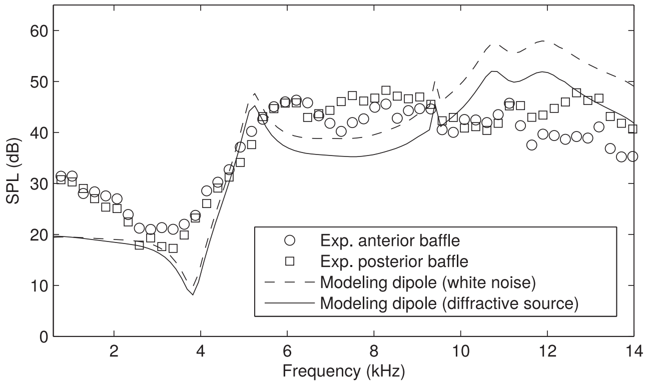

Fig. 7. Sound spectra measured at 90° in the transverse plane (x_1 - x_3) with baffle at anterior (○) and posterior (□) positions and modeled spectra for dipoles with white noise (- - -) and difractive source (——). The distance between the replica and the measurement position is 49 cm.

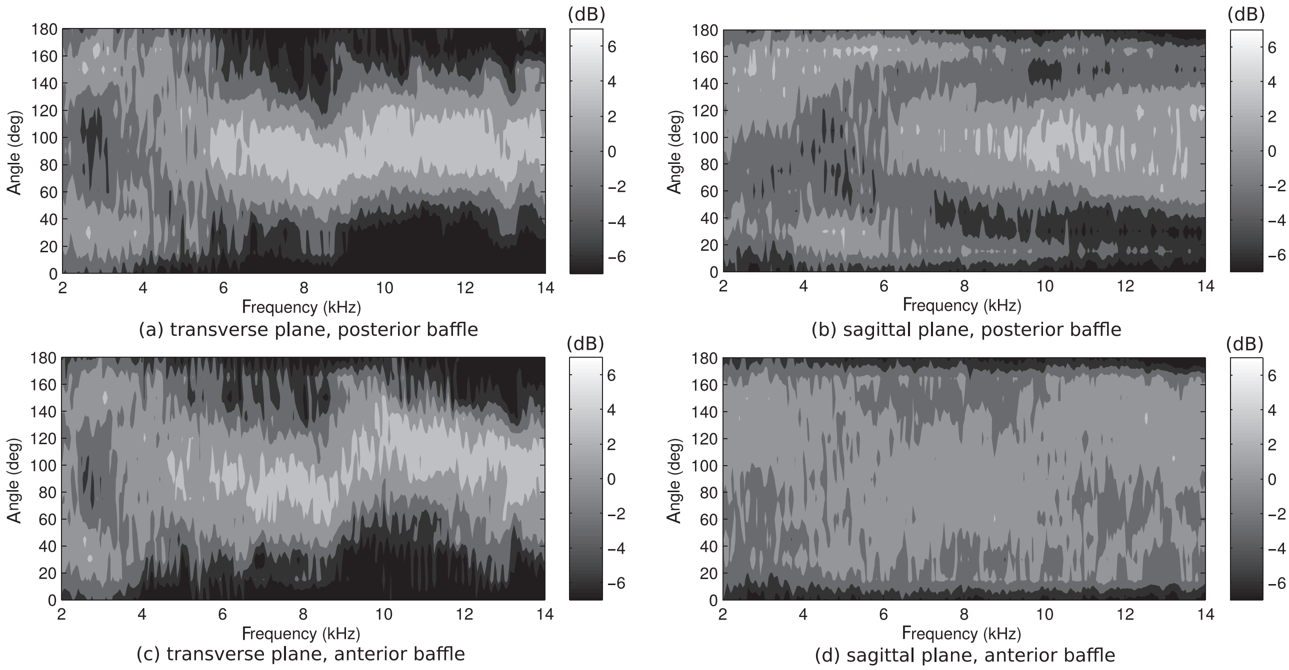

Fig. 8. Normalized pressure amplitudes measured in the experiments. The transverse plane (x_l - x_3) and the sagittal plane (x_1 - x_2) with the baffle at posterior position are plotted in (a) and (b), whereas these with the baffle at anterior position are plotted in (c) and (d), respectively.

monopole source. In the sagittal plane, troughs appeared in the frequency range from 8.5 up to 9.4 kHz for the monopole source. The amplitude at 90° increased for frequencies greater than 10 kHz for both sources.

To compare the directivity pattern between the experiments and modeling, normalized pressure amplitudes near the troughs (8.4 and 9.4 kHz) are plotted as a function of angle in Fig. 10. The amplitudes were normalized by the maximum value at each frequency.

In the transverse plane at 8.4 kHz, amplitudes at 0° and 180° decreased by 15 dB compared to that of the center (90°) in the experiment. Meanwhile, a trough of amplitudes appeared at 110° for the monopole modeling, whereas the amplitudes at 0° and 180° decreased by 7–12 dB for the dipole modeling. At 9.4 kHz in the transverse plane, the amplitude at 0° decreased by 15 dB for the experiment, whereas the amplitude decreased by 25 dB for the dipole modeling. For the monopole modeling, a trough appeared at 75°.

In the sagittal plane at 8.4 kHz, the amplitudes were almost uniform from 0° to 180° for both the experiment and the dipole modeling. In contrast, the monopole modeling gradually departed from the experimental levels between 90° and 180° up until 7 dB. At 9.4 kHz in the sagittal plane, the amplitudes were almost uniform from 0° to 180° for both the experiment and the dipole modeling, whereas the amplitudes between 90° and 180° decreased by 10 dB for the monopole modeling.

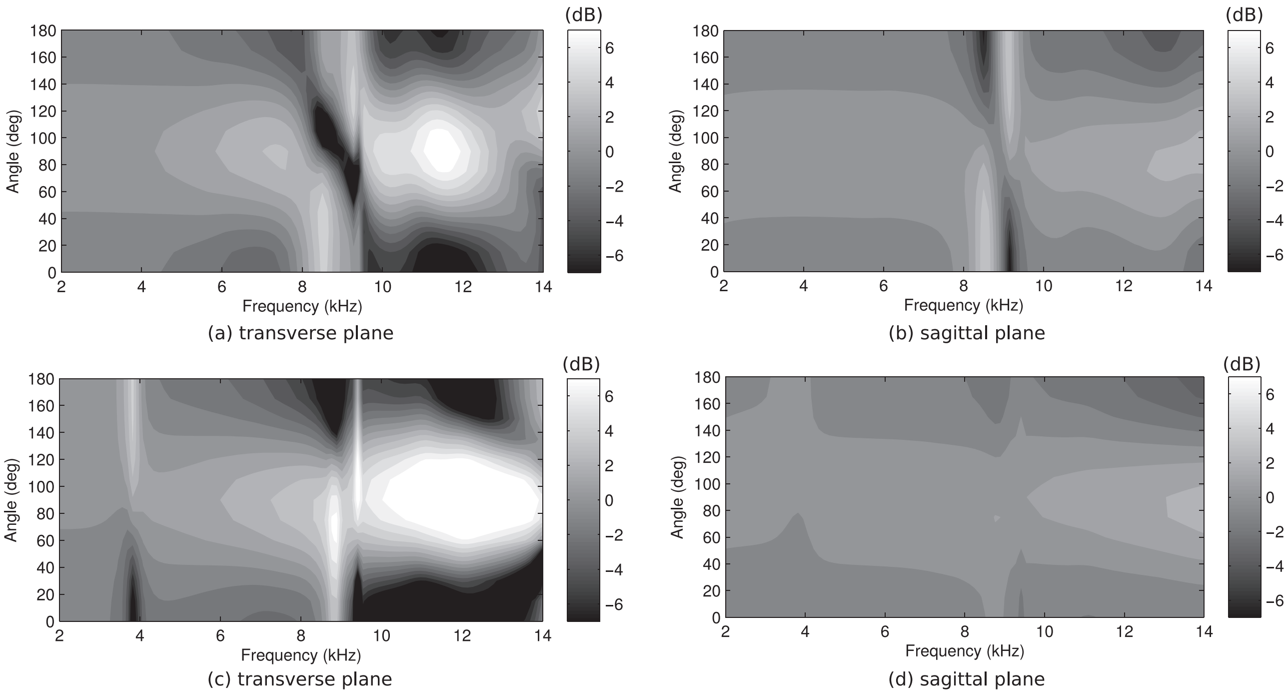

Fig. 9. Modeled normalized pressure amplitude. The transverse plane (x_1 - x_3) and the sagittal plane (x_1 - x_2) with monopole source are plotted in (a) and (b), whereas these with dipole source (diffracive source) are plotted in (c) and (d), respectively.

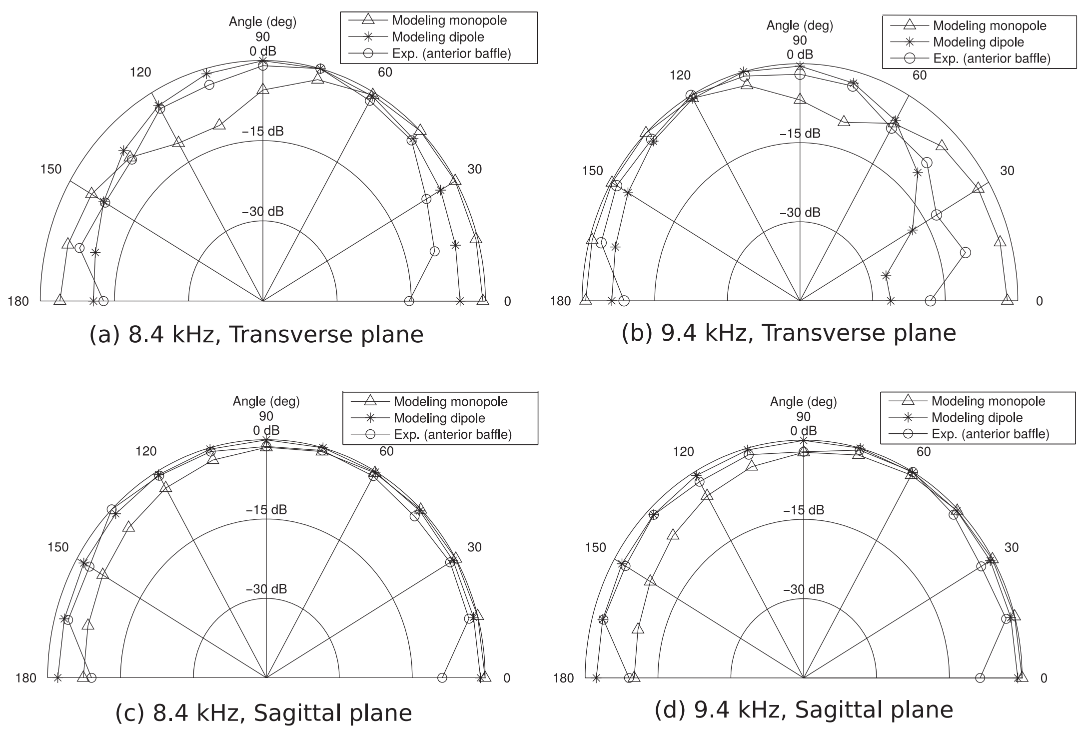

Fig. 10. Normalized acoustic amplitudes as a function of angle for the modeling with monopole and dipole sources and experiments with anterior baffle position. Amplitudes at transverse plane are plotted in (a) for 8,4 kHz and in (b) for 9.4 kHz. Amplitudes at sagittal plane are plotted in (c) for 8.4 kHz and in (d) for 9.4 kHz.

4. Discussion

The comparison of measured sound spectra (Fig. 6) with the baffle in two positions showed that amplitudes above 7 kHz were increased (<10 dB) by shifting the baffle downstream, *i.e.* from the anterior to the posterior position. This suggests that geometrical changes to the lips and speaker's face affect the amplitude of /s/ only in the high frequency range (>7 kHz). Note that changes can occur between speakers or intra-speaker, *e.g.* due to changes in facial expression, aging or wearable dental

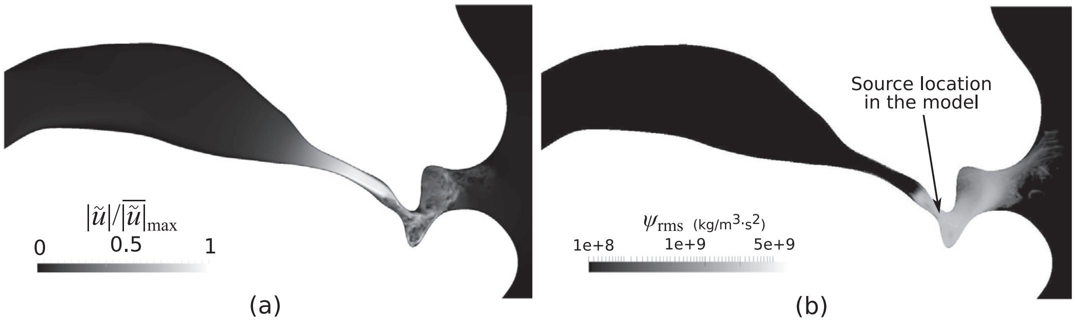

Fig. 11. Flow fields obtained by numerical simulation for the realistic oral tract geometry of /s/ shown in Fig. 1 (b) (Yoshinaga, 2018). Instantaneous normalized velocity magnitude and amplitude of sound source are plotted in (a) and (b), respectively.

apparatuses such as braces or retainers. Although it might not be significant for listeners of normal hearing [28], this finding encourages further perceptual studies focusing on the high-frequency range for /s/ and fricatives in general as well as detailed studies on the potential impact of the baffle position and lips end termination.

Modeled sound spectra with the monopole source (Fig. 6) exhibited the first peak at 5.3 kHz and the second peak at 11 kHz whose amplitudes coincide with measured amplitudes at these peak frequencies. Nevertheless, modeled amplitudes underestimate measured values for frequencies in between these peak values. This is consistent with previous findings for a five-cross-section geometrical oral tract simplification [10]. Modeled spectra with the dipole source of white noise characteristics reduced this discrepancy between the modeled and measured spectra in the frequency range from 5 kHz up to 10 kHz. In addition, by using the diffractive source characteristics for the dipole source, the amplitude above 10 kHz also agreed with the measured spectra. This result indicates that the flow source in the oral tract for /s/ was better approximated by the dipole source than by the monopole source. Consequently, using the dipole source with diffractive characteristics improved the multimodal model outcome for the prediction of /s/. Remaining differences for peak frequencies and amplitudes between experiment and modeling were probably caused by the geometrical differences between the realistic oral tract (Fig. 1 (b)) and the rectangular simplification (Fig. 1 (c)) used for the measurements and modeling, respectively. To improve the results, we need to decrease the channel length L_{x_1} until the geometrical differences become negligible.

To confirm whether the flow source in the realistic geometry of /s/ is generated in the expected place of the modeling, flow simulation conducted for the realistic geometry used for the measurements is shown in Fig. 11 [26]. The figure showed that the jet impinged on the upper incisors so that the sound source started to appear near the surface of the incisors. The position of the dipole source was within the sound source region observed in the flow simulation. Nevertheless, the flow source in the simulation covers a wider region which extends between incisors and lip surfaces. Therefore, modeling might be further improved by accounting for the source distribution in order to further clarify the relationship between characteristics of the sound source region and the generated sound.

In the directivity pattern, pressure amplitudes along the center of propagation direction (90°) increased when the baffle was at the posterior position (normal lip-face junction position). This is consistent with directivity observations for an English speaker [16] and an oral tract replica of /s/ [17]. Meanwhile, by shifting the baffle from the posterior to the anterior position, the high-frequency (10–12 kHz) directivity in the transverse plane was shifted from 90° to 160°, whereas no pattern appeared in the sagittal plane. This indicates that lip protrusion orients the pressure amplitude to the center 90°, and hence, the sound is directed towards the speaker's forward direction. Thus, as discussed for spectra shown in Fig. 6 at the beginning of this section, observed directivity patterns emphasis the need to investigate geometrical lip features in more detail in order to account for the impact of the end termination on the sound properties of /s/ and other fricatives.

The modeled directivity pattern showed that the large amplitude region was shifted to the region 100° to 180° at 9 kHz in the transverse plane, whereas the directivity pattern was flattened for the sagittal plane. These tendencies reflect the tendencies observed on the measured directivity pattern with the baffle at the anterior position so that it is demonstrated that the overall tendencies of directivity pattern for /s/ can be estimated using the outlined multimodal model approach.

Meanwhile, details such as small peaks and troughs characterizing modeled directivity pattern in the frequency range from 8 up to 10 kHz did not appear in the measurement patterns. However, the antenna plot (Fig. 10) showed that the measured amplitudes at 8.4 and 9.4 kHz agreed well with those of the dipole source compared to the monopole source for both planes. This result also indicates that the modeling with dipole source better expresses the flow source during the production of /s/. To improve the accuracy of the modeled directivity pattern, it is necessary to improve the model by expressing the flow source distributions observed in the flow simulation of Fig. 11. In addition, in order to be able to account for a posterior baffle (more realistic) position in the model, it is necessary to develop a methodology for calculation of the radiation impedance for an end termination representing a protruded lip outlet.

5. Conclusion

Multimodal acoustic modeling for a rectangular simplification of the oral tract geometry of fricative /s/ was developed using monopole and dipole sound sources in order to predict the acoustic characteristics of /s/. Modeled spectra and directivity patterns were compared to measured sounds with the flow supplied to the realistic oral tract replica of /s/. The discrepancy between measured and modeled spectra was less than 5 dB up to 14 kHz when a dipole sound source was positioned at the gap upstream from the upper incisor. In addition, main tendencies of the measured directivity pattern with a baffle at the anterior (lip tip) position were predicted using a dipole source. These results indicate that the proposed model approach enables to predict both the spectrum and the directivity pattern of fricative /s/ with higher accuracy compared to the previous models. The proposed approach is applicable to further analyzing oral tract geometries of other fricatives when aiming a systematic investigation of the impact of geometrical features and simplifications on the acoustic outcome.

Author contribution statement

Conceptualization: T.Y. and A.V.H., Vocal tract geometries: T.Y., A.V.H. and K.N., Data acquisition: T.Y., K.N. and A.V.H., Modeling T.Y. and A.V.H., Funding acquisition: S.W.and A.V.H., Original drafting, editing and review: T.Y., A.V.H., K.N. and S.W.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partly supported byMEXT as Priority Issue on Post-K computer [Project ID: hp190187] and by ArtSpeech project [Project ID: ANR-15-CE23-0024]. We thank the anonymous reviewers for their helpful comments and suggestions.

References

[1] G. Fant, Acoustic Theory of Speech Production, Mouton, The Hague, Paris, 1960, p. 1960.
[2] K.N. Stevens, Airflow and turbulence noise for fricative and stop consonants: static considerations, J. Acoust. Soc. Am. 50 (1971) 1180–1192 1971.
[3] C.H. Shadle, The Acoustics of Fricative Consonants, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, 1985.
[4] M.S. Howe, R.S. McGowan, Aeroacoustics of [s], Proc. R. Soc. A 461 (2005) 1005–1028 2005.
[5] L.M.T. Jesus, C.H. Shadle, A parametric study of the spectral characteristics of european Portuguese fricatives, J. Phonetics 30 (2002) 437–464 2002.
[6] M.H. Krane, Aeroacoustic production of low-frequency unvoiced speech sounds, J. Acoust. Soc. Am. 118 (1) (2005) 410–427 2005.
[7] T. Yoshinaga, K. Nozaki, S. Wada, Effects of tongue position in the simplified vocal tract model of Japanese sibilant fricatives /s/ and/, J. Acoust. Soc. Am. 141 (2017) EL314–EL318 2017.
[8] K. Nozaki, Numerical simulation of sibilant [s] using the real geometry of a human vocal tract, in: High Performance Computing on Vector Systems 2010, Springer, 2010, pp. 137–148. 2010.
[9] T. Yoshinaga, K. Nozaki, S. Wada, Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract model, Phys. Fluid. 30 (2018) 035104 2018.
[10] T. Yoshinaga, A. Van Hirtum, S. Wada, Multimodal modeling and validation of simplified vocal tract acoustics for sibilant /s/, J. Sound Vib. 411 (2017) 247–259 2017.
[11] K. Motoki, A parametric method of computing acoustic characteristics of simplified three-dimensional vocal-tract model with wall impedance, Acoust Sci. Technol. 34 (2013) 113–122 2013.
[12] 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 2015.
[13] B.B. Monson, A.J. Lotto, S. Ternstrm, Detection of high-frequency energy changes in sustained vowels produced by singers, J. Acoust. Soc. Am. 129 (4) (2011) 2263–2268 2011.
[14] R. Blandin, A. Van Hirtum, X. Pelorson, R. Laboissire, Influence of higher order acoustical propagation modes on variable section waveguide directivity: application to vowel [a], Acta Acust. United Ac. 102 (5) (2016) 918–929 2016.
[15] R. Blandin, A. Van Hirtum, X. Pelorson, R. Laboissire, The effect on vowel directivity patterns of higher order propagation modes, J. Sound Vib. 432 (2018) 621–632 2018.
[16] B.B. Monson, E.J. Hunter, B.H. Story, Horizontal directivity of low- and high-frequency energy in speech and singing, J. Acoust. Soc. Am. 132 (2012) 433–441 2012.
[17] T. Yoshinaga, A. Van Hirtum, K. Nozaki, S. Wada, Influence of the lip horn on acoustic pressure distribution pattern of sibilant /s/, Acta Acust. united Ac. 104 (1) (2018) 145–152 2018.
[18] 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 1989.
[19] 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 1996.
[20] 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 1997.
[21] 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.
[22] A.D. Pierce, Acoustics: an Introduction to its Physical Principles and Applications, Acoust. Soc. Am., New York, 1989, p. 1989.
[23] K. Nozaki, T. Yoshinaga, S. Wada, Sibilant /s/ simulator based on computed tomography images and dental casts, J. Dent. Res. 93 (2014) 207–211 2014.
[24] P.A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, Neuroimage 31 (3) (2006) 1116–1128 2006.
[25] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, A. Van Hirtum, X. Pelorson, Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds, J. Acoust. Soc. Am. 140 (3) (2016) 1707–1718 2016.

[26] T. Yoshinaga, Experimental and Computational Analysis on Aeroacoustic Mechanisms of Sibilant Fricative Production, Ph.D. thesis, Osaka University, Osaka, 2018.
[27] A. Van Hirtum, Y. Fujiso, Insulation room for aero-acoustic experiments at moderate Reynolds and low Mach numbers, Appl. Acoust. 73 (2012) 72–77 2012.
[28] 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 2014.