Experimental and Computational Analysis on Aeroacoustic Mechanisms of Sibilant Fricative Production

Tsukasa Yoshinaga

March 2018

Experimental and Computational Analysis on Aeroacoustic Mechanisms of Sibilant Fricative Production

A dissertation submitted to

The Graduate School of Engineering Science

Osaka University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Engineering

by

Tsukasa Yoshinaga

March 2018

Abstract

The sibilant fricatives are speech sounds that are produced by a turbulent jet flow in the vocal tract. The aeroacoustic mechanisms of sibilant fricative production were studied in the context of three domains: mechanical replicas, theoretical modeling, and numerical simulations. To study the relationship between the vocal tract geometry and acoustic properties of generated sound, realistic and simplified replicas were constructed using medical images of a Japanese speaker pronouncing the sibilant fricatives. With the realistic replica of sibilant /s/, effects of lip horn on distribution patterns of the generated sound were investigated. Results showed that the lip horn enhances the pressure amplitude at the center of the lips in both transverse and sagittal plane.

The essential geometric factors which generate the acoustic properties of sibilant fricatives were investigated by constructing the simplified vocal tract replicas. Cross-sectional areas and vertical heights at five positions in the vocal tract were used to form a rectangular flow channel. In addition, dimensions of a simplified tongue replica were modified in accordance with the medical images. The experimental measurements revealed that the spectral difference between the Japanese fricatives /s/ and /ʃ/ was reproduced by changing the position, tip shape, and constriction width of the tongue replica.

The multimodal acoustic theory was applied to the simplified vocal tract geometry to investigate the acoustic properties of the sibilant fricatives. In the modeling, a sound source was placed either at the inlet of the vocal tract or downstream from the tongue constriction. The modeled pressure amplitude was validated experimentally using a compression driver or airflow supply at the vocal tract inlet. Results showed that the predicted spectrum captured the spectrum measured with both the acoustic driver and the flow supply. By positioning the source near the upper teeth wall, higher frequency peaks observed for the Japanese speaker were 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 vocal tract when the source was placed downstream from the constriction.

The large-eddy simulations were applied to both realistic and simplified replicas in order to investigate the relationship between the flow configuration and generated sound. The numerical simulations revealed that the jet flow generated from the constriction impinged on the upper teeth wall and caused the main sound source upstream and downstream from the gap between teeth. While magnitudes of the sound source decreased with increments of the frequency, amplitudes of the pressure downstream from the constriction increased at the peak frequencies of the corresponding tongue position. These results indicate that the sound pressures at the peak frequencies increased by acoustic resonance in the channel downstream from the constriction, and the different frequency characteristics between /s/ and /ʃ/ were produced by changing the constriction and the acoustic node positions inside the vocal tract.

The investigation in three domains enabled to analyze the aeroacoustic mechanisms, *i.e.* the relationship among the vocal tract geometry, the source generation from the jet flow, and the acoustic resonance properties. Further understanding of speech production can be expected by applying these methodologies on other subjects using other languages as well as the co-articulation of fricative consonants in word contexts.

Contents

List of Figures

List of Tables

List of Symbols

c Velocity of sound (m/s)
f Frequency (Hz)
δ End correction of the rectangular channel (mm)
L_{CA} Length between the lower teeth and the constriction (mm)
L_f Length between the constriction and lip outlet (mm)
p Sound pressure (Pa)
$\boldsymbol{v}$ Particle velocity vector (m/s)
ω Angular frequency (rad/s)
j Imaginary number
ρ Density of air (kg/m^3)
$k = \omega/c$ Free field wave number (m^{-1})
ϕ Velocity potential
n Number of modes in *x*-direction
m Number of modes in *y*-direction
a_{mn} Forward propagation amplitudes
b_{mn} Backward propagation amplitudes
ψ_{mn} Propagation modes
$\gamma_{z,mn}$ Propagation constant
L_x Length of cross-section in *x*-direction (mm)
L_y Length of cross-section in *y*-direction (mm)
σ_m, σ_n = 1 (*m*,*n* = 0) or 1/2 (*m*,*n* ≥ 1)
$f_{c,mn}$ Cutoff frequency (Hz)
v_z Particle velocity along *z*-axis (m/s)
$\boldsymbol{\psi}$ Column vector composed of ψ_{mn}
$\boldsymbol{a}$, $\boldsymbol{b}$ Column vector composed of $j\omega\rho a_{mn}$, $j\omega\rho b_{mn}$
$\boldsymbol{Z}_C$ Characteristic impedance matrix
$\boldsymbol{D}$ Propagation constant matrix
T Transpose operator
$\boldsymbol{P}$ Modal sound pressure
$\boldsymbol{V}$ Modal particle velocity
$\boldsymbol{\Psi}_{i,i+1}$ Mode-coupling matrix
S_i Area of section *i* (mm^2)
$\boldsymbol{Z}_{rad}$ Radiation impedance matrix
$\boldsymbol{P_{out}}, \boldsymbol{V_{out}}, \boldsymbol{\psi_{out}}$ Modal pressure, modal velocity, and propagation mode at outlet
$\boldsymbol{V_{in}}, \boldsymbol{\psi_{in}}$ Modal velocity, and propagation mode at inlet
v_0 Particle velocity of sound source (m/s)
Ω_0 Area of vibrating surface (mm^2)
Q Volume flow rate (mm^3/s)
$\boldsymbol{P}^+, \boldsymbol{V}^+, \boldsymbol{P}^-, \boldsymbol{V}^-$ Variables downstream and upstream of source section

G_{lip} Pressure-pressure transfer function at lip cavity (dB)
t Time (s)
u_i $(i = 1, 2, 3)$ Flow velocity components (m/s)
φ_i Momentum flux, $= \rho u_i$ (kg/m^{2}s)
e Internal energy (J)
$E = 1/2\,|u_i|^2 + e$ Total energy (J)
T Temperature (K)
$\bar{\cdot}$ Grid scale value
$\tilde{\cdot}$ Favre mean filtered value
σ_{ij} Viscous stress tensor
ν Kinematic viscosity of air, = 0.15 cm^2/s
ν_{SGS} Subgrid-scale viscosity (cm^2/s)
s_{ij} Strain rate tensor
q_j Heat flux (W/m^2)
α Thermal diffusivity (m^2/s)
α_{SGS} Subgrid-scale thermal diffusivity (m^2/s)
k_{SGS} Subgrid-scale turbulent energy (m^2/s^2)
τ_{ij} Turbulence stress tensor
C_{s} Constant for Smagorinsky turbulence model
C_k, C_ϵ Constants for One-equation type turbulence model
P_{rt} Turbulence Prandtl number
Δ Scale filter length
L_{C} Length between the tongue tip and lower teeth (mm)
L_{T} Length between the tongue tip and constriction (mm)
h Height of constriction (mm)
Re Reynolds number, $= uh/\nu$
M Mach number, $= u/c$
T_{ij} Lighthill tensor
ψ Source term in Lighthill's analogy
ψ' Source term in frequency domain
SPL Sound pressure level (dB)
ψ_Ω, SPL_Ω Surface integral values (dB)
S_Ω Plane surface for surface integral (mm^2)
ψ'_k, SPL_k, S_k Discretized values for surface integral
q Second invariant of velocity gradient tensor

Chapter 1. Introduction and Literature Review

1.1 Introduction

Speech production is one of the most important roles of human species, and only human can frame a sentence and communicate through the speech production. In the speech production, generated sounds are classified as vowels and consonants. For the vowel sound, the vocal tract is in relatively open configuration and the sound source is mainly generated by the vibration of vocal folds in a glottis. In contrast to the vowel sounds, the vocal tract of consonants is relatively closed and the sound source is generated by turbulent flow in the constricted channel of the vocal tract (Stevens, 1998). The consonants are classified by the manner of sound production (*e.g.*, nasal, stop, and fricative consonants) and place of articulation (*e.g.*, labial, sibilant, glottal fricatives). The fricative consonants are sounds generated by the turbulent flow at a narrow constriction downstream from the vocal fold, *e.g.* /f/, /s/, and /h/. The sibilant fricatives are sounds generated by forming the constriction between the tongue and hard palate, and mainly divided into two kinds of phonemes: /s/ with the constriction at the alveolar ridge; and /ʃ/ with the constriction at posterior position to /s/ (Fig. 1.1), in English, Japanese and other languages. The typical frequency spectra of sustained Japanese sibilant fricatives /s/ and /ʃ/ are shown in Fig. 1.2. The sibilant fricatives are generally characterized as broadband noise in the frequency range above the characteristic peak, about 4-7 kHz for /s/ and 2-3 kHz for /ʃ/ (see the Appendix A for more examples). In addition, above consonants are classified to voiced and voiceless consonants. In the voiced consonants, the sound source is additionally generated at the vocal fold as well as the turbulent flow at the constriction, *e.g.* /z/, /ʒ/.

The speech production mechanisms have been investigated by using acoustic theories. Fant (1960), Heinz and Stevens (1961) firstly described the frequency characteristics of vowels and consonants by using the acoustic source-filter theory. The filter function was derived from the vocal tract geometry measured by X-ray images, and frequency characteristics of vowels were obtained by multiplying the filter to the source spectrum of the vocal folds or flow noise. Moreover, the acoustic properties of fricative consonants were estimated by modeling the vibrating spoiler in a tube (Stevens, 1971). The detailed acoustics of fricative consonants were described by Shadle (1985) conducting the mechanical experiment of the simplified vocal tract replicas. She simplified the vocal tract geometry by using a constricted channel and obstacle in a cylinder, and explained the difference of acoustic characteristics of fricatives by changing the position and size of the constriction and obstacle. From the knowledge of the acoustic mechanisms for the fricatives, classification of the fricatives based on the sound measurement was obtained (Jesus and Shadle, 2002), and this classification is used for the recent technology of speech recognition (*e.g.* Patgiri *et al*., 2013).

The acoustic properties of the sibilant fricatives have been also widely discussed in the field of phonetics. The differences of the tongue movement and acoustic characteristics during the word pronunciation have been discussed (Shadle and Scully, 1995; Iskarous, *et al*., 2011). In addition, acoustic differences of sibilant fricatives among different languages

have been investigated by several researches. Badin (1991) reported the spectra of sustained fricative consonants pronounced by male French and female English speaker. Reidy (2016) investigated the acoustic difference of English and Japanese sibilant fricatives /s/ and /ʃ/ in word pronunciation. Moreover, in the field of neurology and phonetics, Perkell *et al*., (1979) argued that the acoustic distinctness between /s/ and /ʃ/ is related to the speaker's auditory discrimination. Their group has also investigated the relationship among speech motor control, somatosensory function, auditory feedback, and brain model using the sibilant fricative production (Perkell, 2012).

From the view point of physics and fluid dynamics, the sibilant fricatives have complicated phenomena in the vocal tract. Since the fricative sounds are generated by the turbulent flow in the vocal tract, it is important to examine how the sound source is generated and how the sound propagates from the source. In addition, to understand the sibilant fricative production in the word context, we need to consider the movement of tongue and jaws. In order to predict the flow configuration and the source distribution in the vocal tract geometry, Nozaki *et al*. (2005) applied the turbulent flow simulation to the vocal tract geometry obtained by the medical images. In addition, several flow simulations have been applied to the simplified vocal tract geometry (Ramsay and Shadle, 2006; Van Hirtum *et al*., 2010). In contrast to the numerical simulation, Howe and McGowan (2005) applied the aeroacoustic theory to a quasi-one-dimensional simplified vocal tract model and predicted the far-field sound spectrum of /s/.

As described above, many scientific fields including acoustics, flow dynamics, phonetics, and neurolinguistics have been focused on the sibilant fricative production. However, there are still unclear points on the aeroacoustic mechanisms of the sibilant fricatives. Therefore, the target is focused on the sibilant fricatives in this thesis. Before addressing the problem, the recent studies are reviewed in more detail below.

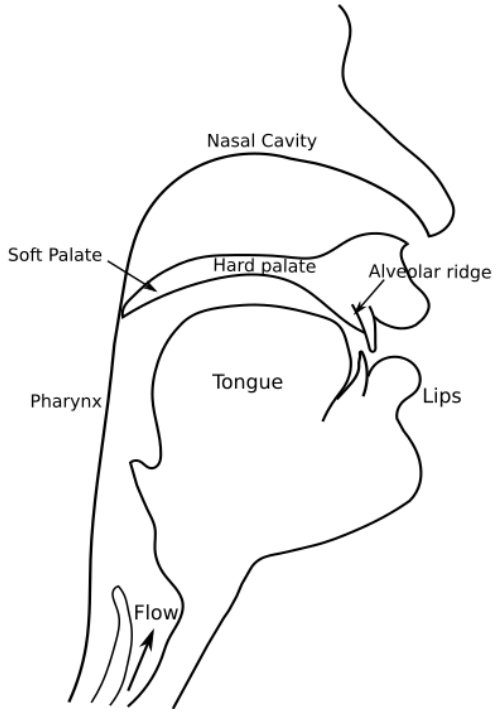


Fig. 1.1 Mid-sagittal plane of vocal tract geometry pronouncing /ʃ/.

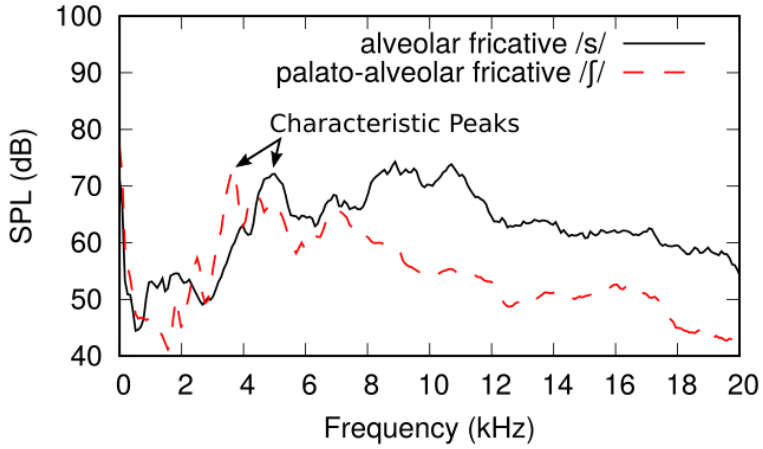


Fig. 1.2 Typical frequency spectra of voiceless sibilant fricatives /s/ and /ʃ/.

1.2 Literature Review

1.2.1 Recent Study of Sibilant Fricative Production

In recent years, vocal tract geometry of sibilant fricatives has been measured by using X-ray or Magnetic Resonance Imaging (MRI), and the production mechanisms have been investigated by using those imaging techniques. Narayanan *et al.*, (1995) measured the vocal tract geometry of fricative consonants using the MRI, and reported the area functions which can be used to estimate the transfer function of acoustic modeling (Narayanan and Alwan, 2000). Toda and Honda (2003) measured the front cavity area and defined the palatalization index based on the nasal spine and tongue surface in the MRI. By using the index, they distinguished the vocal tract geometries /s/ and /ʃ/. With Wisconsin X-ray Microbeam (XRMB) database (Westbury, 1994) which contains the movement of the tongue, jaw, and lips, and simultaneously recorded acoustic output, Iskarous *et al.*, (2011) analyzed effects of the tongue and jaw movement on the production of /s/. Meanwhile, by using the X-ray cone-beam computed tomography (CT) images of sustained /s/, Nozaki *et al.*, (2014) constructed the vocal tract replica of /s/ and measured the flow velocity and sound generated by the replica with a flow supply.

In addition to the X-ray and MRI, several methods have been proposed to measure the vocal tract geometry while subjects are pronouncing the sibilant fricatives. Wood *et al.*, (2009) used an electro-palatogrphy (EPG) to measure the area of the tongue surface attaching on the hard palate. Zharkova (2016) measured the tongue shape of /s/ and /ʃ/ using the ultrasound imaging. By comparing the tongue shape and simultaneously measured sound of adults and preadolescent children, she found that the difference between the adults and children occurs in consonant-vowel boundaries.

By using X-ray CT scan, the detailed three-dimensional (3D) geometry can be obtained in short amount of time. However, usage of CT scan is limited because of exposure to the nuclear radiation. Meanwhile, detailed 3D geometry can be also obtained by MRI, although it takes longer time than CT scan, and the teeth geometry has to be additionally obtained by taking another scan for teeth geometry (Takemoto *et al.*, 2004). By using the ultrasound, the mid-sagittal tongue surface can be measured with high frequency rate.

However, it is difficult to obtain the 3D shape including the hard palate and teeth by using the ultrasound. Therefore, it is important to choose the techniques depending on the kind of information we want to obtain from the measurement.

In the field of phonetics, the fricative production mechanisms have been investigated by obtaining various spectral parameters from the recorded sounds. Jongman *et al*., (2000) measured the spectral peak location and spectral moments, and showed that the four fricatives /f/, /ɵ/, /s/, and /ʃ/ can be distinguished by using those parameters. Based on the knowledge obtained by the simplified vocal tract model, Jesus and Shadle (2002) measured the slope values of line fit to the spectrum to distinguish the fricatives /f/, /s/, and /ʃ/. By using the parameters, the recorded sounds were classified for four Portuguese speakers. Meanwhile, Haley *et al*., (2010) used the spectral mean and skewness (first and second spectral moments) to distinguish two sibilants /s/ and /ʃ/ for ten participants, and showed that only the spectral mean distinguished two phonemes in individual speaker. Reidy (2016) introduced the ERB number which is a frequency scale in psychoacoustics, and divided the sibilant fricatives /s/ and /ʃ/ for temporal variation in both English and Japanesse. To find the cause of these parametrical changes, it is necessary to focus on the physical phenomena in the vocal tract which produce the frequency characteristics of the fricatives. Although the simple acoustic analysis has been done by Toda and Maeda (2006) to identify the vocal tract geometry which causes the characteristic spectral peak, it is still unclear how the overall frequency characteristics are produced by the turbulent flow source in the vocal tract.

In the field of acoustics, the influence of higher order modes, which considers the propagation of 3D modes, has been recently discussed. Blandin *et al*. (2015) applied multimodal theory to vocal tract geometries of vowels, and showed the influence of higher-order modes on spectral amplitudes above 4.5 kHz. Motoki *et al*. (2000) and Motoki (2013) 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/ lacked from the modeled transfer function because the sound source was located at the inlet of the vocal tract. Meanwhile, Blandin *et al*., (2016) used the multimodal theory to investigate the cause of speech directivity pattern reported by the measurement conducted on the normal speaker and singer (Cabrera *et al.*, 2011; Monson *et al*., 2012). Because the sibilant fricatives are characterized by the acoustic energy in higher frequency range than those of vowels, the higher-order modes may affect perceptually-significant spectral features of fricatives compared to the vowels. Therefore, it is desirable to investigate the influence of higher-order modes on the spectral characteristics as well as the radiated pressure pattern outside of the vocal tract for the sibilant fricatives.

In the field of neurolinguistics, effects of acoustic feedbacks (*i.e*. speech recognition) on the tongue movement, constriction formation, and generated sound characteristics have been investigated. Perkell *et al*., (2004) conducted experiments with a contact sensor on the posterior surface of the lower teeth, and measured the speaker's tongue position, generated sound and ability of auditory discrimination between /s/ and /ʃ/. Their results indicate that differences in degree of acoustic contrast between /s/ and /ʃ/ were related to differences in

their use of tongue contact and in their discriminative performance. Shiller *et al*., (2009) conducted the experiment with the auditory feedback manipulation synthesizing the fricative sound from /s/ to /ʃ/, and found that the synthesized auditory feedback alter the speaker's motor control and acoustic boundaries between /s/ to /ʃ/. In addition, Ghosh *et al.*, (2010) investigated the relationship between the auditory acuity, somatosensory acuity, and magnitudes of the acoustic contrast between sibilant fricatives /s/ to /ʃ/ by using plastic dome sensors. Their results also showed that the auditory and somatosensory goals are related to the sibilant fricative production. Moreover, in order to investigate the tongue movement mechanisms, biomechanical models of tongue muscle fibers have been proposed (Buchaillard *et al*., 2009; Stavness *et al*., 2012). The relationship between the auditory feedback and tongue movement was investigated by constructing a Bayesian model using the biomechanical model and acoustic analysis (Patri *et al*., 2015).

1.2.2 Vocal Tract Models for Sibilant Fricatives

The production mechanisms of sibilant fricatives have been studied by many researchers using simplified and realistic vocal tract models. The vocal tract models are summarized in Fig. 1.3. Firstly, simplified models have been proposed by Stevens (1971) and Shadle (1985) with a constricted channel and obstacle in a tube. With their models, the frequency characteristics of sibilant fricatives, which consist of broad band noise with a characteristic peak in the frequency range 2 to 5 kHz, were reproduced by assuming the broad band noise source at the constriction in the simple tube. In addition, Shadle (1985) experimentally demonstrated that the position of the constriction affects the frequency characteristics of the sound, which form the differences between sibilant fricatives /s/ and /ʃ/. Motoki *et al*., (2000) proposed a 36-section acoustic model using rectangular channels and cross-sectional areas measured by MRI of a subject pronouncing /ʃ/, and investigated effects of the vocal tract geometry on the pressure distribution and transfer function. Howe and McGowan (2005) pointed out the importance of the source position and assumed that the monopole source is generated by the turbulent boundary layer at the gap between upper and lower teeth. They proposed a quasi-one-dimensional model considering the capacity of a back cavity, a constriction, a lower mouth cavity, and a lip cavity, and theoretically showed that the broadband noise of English /s/ with a characteristic peak at 4 kHz can be produced by assuming the sound source at the gap between the teeth. Meanwhile, by using one-dimensional (1D) acoustic model, Toda and Maeda (2006) investigated effects of constriction length and front cavity length on the characteristic peak frequency of the sibilant fricatives. Moreover, several simplified vocal tract models of sibilant fricatives have been proposed to investigate the flow and acoustic properties inside the vocal tract by using the experimental measurement and 1D flow theory (Shadle *et al*., 2008; Van Hirtum *et al*., 2011).

In addition to the acoustic models, complex flow dynamics in the sound generation for the vocal tract of the sibilant fricatives have been investigated using incompressible flow simulation. Ramsay and Shadle (2006) applied the flow simulation on the simplified vocal tract model and investigated influence of the constriction position on the flow

configuration inside the vocal tract. Cisonni *et al*., (2011) and Cisonni *et al*., (2013) examined the effect of the vocal tract geometry on the magnitude of the sound source generated around the teeth-shaped obstacle by calculating Powell sound source (Powell, 1964). Nozaki *et al*., (2012) investigated the effect of expiratory flow rate on the magnitude of the sound source generated in Shadle's simplified model. Although the magnitude of the sound source was calculated in their flow simulation, the acoustic characteristics of sound propagating from the predicted sound source have not been examined yet. Therefore, it is desirable to investigate the effect of the source distribution on properties of the propagating sound.

In contrast to the simplified models, Nozaki *et al*., (2005) and Nozaki (2010) constructed a realistic vocal tract geometry using CT images of a Japanese subject pronouncing /s/, and applied the incompressible flow simulation. Their simulation showed that the sound source mainly arises near the lower teeth surface. Meanwhile, Toda *et al*., (2003) conducted the finite element acoustic simulation on the realistic vocal tract geometry of sibilant fricatives with dipole sound source at the gap between teeth. Nozaki *et al*., (2014) constructed a realistic replica of the subject's vocal tract by using the CT images and a 3D printer. The separately printed parts of the throat, tongue, upper and lower jaws, and lips were assembled to construct the replica of /s/. The velocity at five positions from throat, constriction, space behind the upper teeth, gap between teeth, and lips was measured by inserting the tip of anemometer in a hole of the replica. The experimental measurements using the anemometer and a microphone demonstrated that the replica reproduced the frequency characteristics of /s/ articulated by the subject up to 15 kHz accompanied by large fluctuations of the flow at points behind the upper front teeth and the gap between the upper and lower teeth. These results indicated that the jet flow needs to spread and disturb in the space downstream from the constriction before impingement on the teeth in order to create the sound close to the actual fricative /s/ produced by the subject. However, since these phenomena were observed only with this subject's /s/, it is needed to examine the effects of geometrical changes, *e.g.* /s/ and /ʃ/, on generated flow and sound, and explore the essential cause of the frequency characteristics of sibilant fricative production.

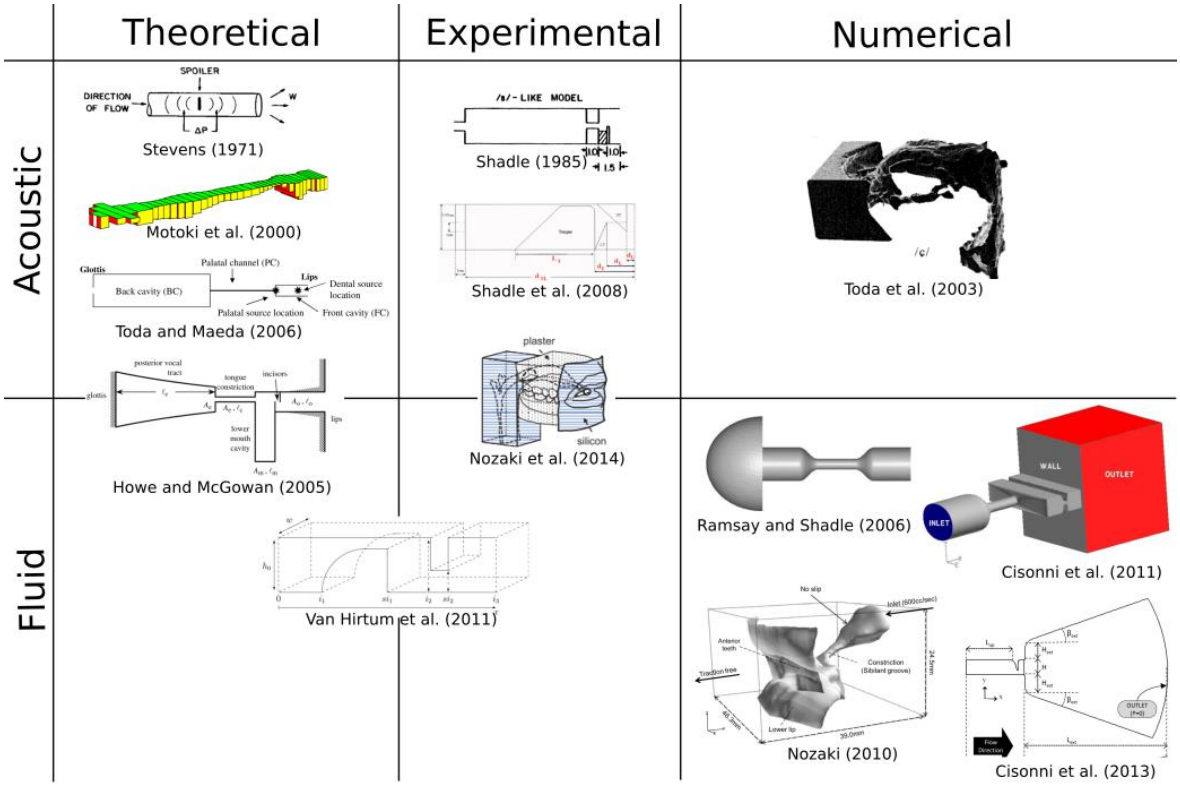


Fig 1.3 Examples of vocal tract models of sibilant fricatives.

1.2.3 Aeroacoustic Sound Generation

The sound source of fricative sounds is generated by the turbulent jet flow issued by the constriction in the vocal tract. In order to understand the aeroacoustic mechanisms of sibilant fricative production, it is necessary to examine how the sound source is generated by velocity fluctuations in the flow field of the vocal tract.

The theory of aeroacoustic sound was firstly introduced by Lighthill (1952). He described the sound source with three types of configuration: a monopole source; a dipole source; and a quadrupole source depending on the source generation mechanisms. In Lighthill's analogy, the fluctuation of mass flux at a point produces the monopole source, the fluctuation of external force vector produces the dipole source, and the fluctuation of stress tensor produces the quadrupole source which is known as Lighthill's sound source. With dimensional analysis, he showed that the total sound power generated by the monopole source increases with the fourth power of the flow velocity u^4, the dipole source increases with sixth power u^6, and the quadrupole increases with eighth power u^8. From the Lighthill's theory, Powell (1964) introduced the sound source using vorticity of the vortexes generated in the flow field. Then, Howe (2003) applied the theoretical analysis on various problems of aeroacoustic sound including sibilant /s/ (Howe and McGowan, 2005).

The aeroacoustic sound generation mechanisms have been also investigated by using computational fluid dynamics (CFD). Since the sibilant fricatives are known to be generated in the flow field of high Reynolds number ($Re = uh/v$, where u is flow velocity, h is the characteristic length, and v is kinematic viscosity) and low Mach number ($Ma = u/c$,

where *c* is speed of sound), the methodology is divided mainly in two ways: direct method and coupling method. In direct method, both flow and sound are directly calculated by solving the three-dimensional compressible Navier-Stokes Equations. In addition, the turbulent flow field is estimated by the Large-eddy simulation (LES) when the computational grid size is not sufficient to capture the smallest vortexes. The LES considers variables only on the computational grids, and turbulent vortexes in a scale smaller than the grid scale are modeled by subgrid-scale (SGS) model. Gloerfelt and Lafon (2008) conducted the direct numerical simulation of turbulent flow and sound passing through a diaphragm in a tube. Yokoyama *et al*., (2015) analyzed the fluid-acoustic interactions in a recorder by using the direct method and computational grids constructed from an actual recorder. In their simulation, the turbulent flow was calculated by explicitly filtering the flow field with a spatial filter, and artificial viscosity was added as the SGS model and numerical stabilizer. With the direct method, the sound generated by the flow can be observed including the acoustic resonance in a flow channel and the acoustic feedback on the flow field. However, the computational cost is higher than the coupling method because the time step of the simulation becomes relatively small in order to calculate the sound propagation with the speed of sound. To solve the problem of the computational cost, the coupling method has been proposed. In the coupling method, the flow and acoustic fields are calculated separately by using incompressible Navier-Stokes Equations and wave equation. Kato *et al*., (2007) applied the coupling method on several engineering product in order to predict the noise level of sound generated by low Mach number flow. They numerically solved Lighthill-Curle's equation (Curle, 1955) or tailored Green's function (Howe, 2003) from the sound source obtained by the flow fields of the incompressible flow simulation. Oberai *et al.*, (2000) proposed a methodology to solve the structural-acoustic interaction (*i.e.* sound propagation around the ridged or elastic walls) by solving the Lighthill's equation using the sound source obtained from the flow simulation. However, the coupling method cannot be used to deal with the fluid-acoustic interaction of the flow induced sound.

In those simulations, the numerical errors often occur from several causes including quality of computational grids, discretization accuracy, and choice of the SGS model. Therefore it is important to confirm that the estimated flow and acoustic fields have enough accuracy in the frequency range of speech sound, by conducting the experimental validation.

1.3 Aim and Outline of the Thesis

The literature reviewed above show that experimental and computational (including theoretical analysis and numerical simulation) analyses have been applied to the sibilant fricative production. Each of these methodologies offers a different type of explanation and insight for the production mechanisms and has a different type of limitation. Firstly, since the theoretical analysis has a merit in low computational costs, it is easier to investigate the effects of geometrical changes on the generated sound. Meanwhile, since it has a difficulty to apply the theoretical analysis on the complicated geometry, only the simplified vocal tract geometries have been used on the analysis. Secondly, the experimental measurements enable to investigate the flow and acoustic properties in the complicated geometry as well as to examine the effects of the geometrical changes in the simplified models. However, there is a limitation for the amount of information available in the experiment, especially information of the flow fields in the vocal tract. Although the hot-wire anemometer enables to measure the velocity fluctuation at a point, it is difficult to measure the velocity distribution in the entire vocal tract geometry. Meanwhile, particle image velocimetry (PIV) enables to measure the velocity distribution in the vocal tract (Geoghegan *et al*., 2012), although it is difficult to obtain the velocity fluctuation in the high frequency rage which causes the frequency characteristics of the sibilant fricatives. Finally, the numerical simulation enables to observe the flow and acoustic fields which cannot be obtained in the experimental measurement. In addition, the simulation can be applied on both simplified and realistic vocal tract geometry. Meanwhile, it takes huge computational costs to simulate a short amount of time in the simulation. In addition, for both theoretical analysis and numerical simulation, it is necessary to validate the computational accuracy with the experiment.

Following the merits and limitations of those methodologies, we conducted all three methods to clarify the aeroacoustic mechanisms of sibilant fricative production. Firstly, in Ch. 2, we conducted the experimental measurement on the realistic and simplified vocal tract geometries. Based on the measurement of the realistic vocal tract replica, we proposed a novel simplified geometry to describe the cause of the characteristic spectral peaks of /s/ and /ʃ/ observed in the subject of medical images. In addition, we conducted the directivity measurement on the realistic replica with and without lips, and investigated the effects of lip horn on the propagation pattern of the sibilant fricatives.

By using the proposed simplified vocal tract geometry, we conducted the theoretical analysis using multimodal acoustic theory in Ch. 3. The multimodal modeling enables to capture the 3D acoustic modes emerging in the vocal tract for the high frequency pressure distribution. Thus, we examined the effects of the higher order modes on the far-field sound spectrum of the sibilant fricatives. In addition, effects of the source positions on the frequency characteristics of generated sounds were investigated in the theoretical modeling.

In Ch. 4, the numerical simulations of both direct and coupling method were applied on the simplified and realistic vocal tract geometry. In the coupling method, LES of incompressible flow and acoustic analysis were conducted on the simplified geometry. In

the direct method, LES of compressible flow was applied for both simplified and realistic vocal tract geometry. The computational accuracy was validated by the experimental measurement. Then, using the validated results, the source generation mechanisms inside the vocal tract as well as the cause of the frequency characteristics of sibilant fricatives were examined.

Through the three kinds of methodologies on the realistic and simplified vocal tract geometries, we will be in a position to answer the final question. What are the aeroacoustic mechanisms of sibilant fricative production? We can find the answer from the comparison of the results in the experimental, theoretical and numerical analysis on the simplified and realistic vocal tract geometries. The conclusion of results, discussion of the question, and perspectives for future works are presented in Ch. 5.

Figure 1.4 shows the nomenclature list unifying the different names used in different chapters for anatomical parts, main planes, and their orientation on the vocal tract of /s/.

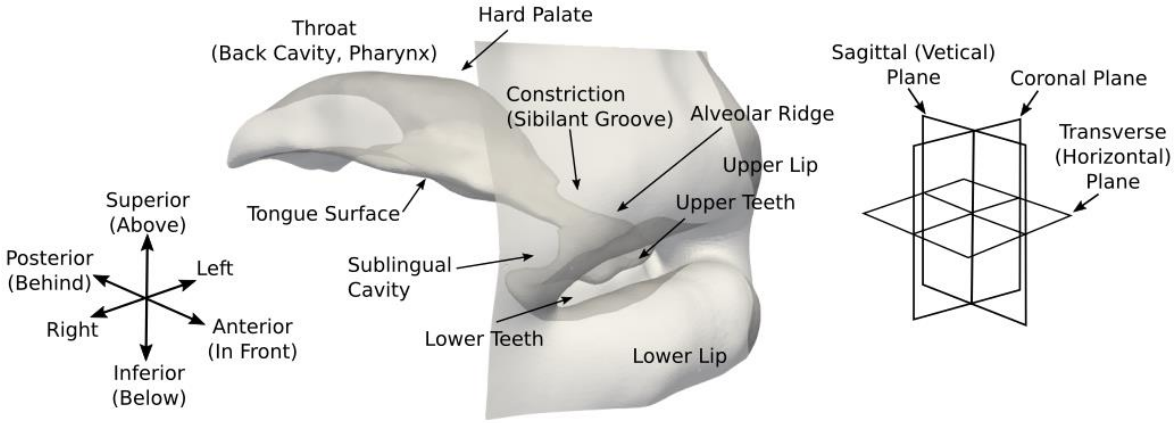


Fig. 1.4 Nomenclature list for anatomical parts, main planes, and their orientation in this thesis.

Chapter 2. Experimental Analysis using Simplified and Realistic Replicas

In this chapter, effects of the vocal tract geometry on the acoustic properties of the sibilant fricatives were investigated by the experimental measurement using the realistic and simplified vocal tract replicas. The both realistic and simplified replicas were constructed based on the medical images of a male Japanese subject pronouncing /s/ or /ʃ/. The realistic replica was constructed in the same way as the sibilant /s/ replica proposed by Nozaki *et al.,* (2014). Then, effects of the lip horn on the far-field sound spectrum as well as the directivity pattern of sibilant /s/ were examined. Moreover, in order to explore the essential vocal tract geometry to pronounce the sibilant fricatives, the simplified vocal tract replicas were constructed based on the knowledge obtained by the realistic replica. Through the construction of the simplified replica, we clarify the fundamental factors to produce the sibilant fricatives /s/ and /ʃ/. In addition, further investigation on the acoustic and flow phenomena is expected for the computational analysis.

2.1 Method

2.1.1 Realistic Replica

The vocal tract geometry was reconstructed from the CT images of a 32-year-old Japanese male subject with normal dentition (Angle Class I) without any speech disorder (self-report). During CT imaging, 512 sagittal slices of 512 × 512 pixels (isotropic 0.1 × 0.1 × 0.1 mm voxels) were obtained while the subject sustained the sibilant /s/ for 9.6 s in a seated position. There is no vowel context for the sustained /s/.

The replica consisted of a pharynx, upper and lower jaws, a tongue, and lips. The upper and lower jaws were produced using rapid prototyping (Zprinter, 3D systems; accuracy: ±0.1 mm). The pharynx, tongue, and lips were made of silicone resin (Wave Silicone, Wave Corp.). The lips were firmly attached to the upper and lower jaws so that there was no gap between jaws and lips. The inlet of the pharynx in the replica was positioned just above the base of the subject's epiglottis. The dimensions of the maximum constriction in the coronal plane are 8 mm by 1.25 mm. The distance between the constriction and tip of the upper teeth is 8 mm. It has been confirmed that this replica reproduced the subject's sibilant /s/ in the frequency range from 3 to 15 kHz (Nozaki *et al.*, 2014).

A rectangular baffle was placed on the jaws or lips, as shown in Fig. 2.1 to mimic the face. For the replica with lips, the baffle was positioned at the point where the lower and upper lips joined. For the replica without lips, the baffle was positioned at the same position as when the lips were present. The space between the baffle and canine teeth or lips was filled with modeling clay. The size of the baffle was changed depending on the measurement distance. The maximum aperture along the sagittal z-direction between the upper and lower lips was 5 mm. The largest aperture along the transverse y-dimension of the lip opening was 27 mm. For the case without lips, the largest aperture of the front teeth along the y-dimension was also 27 mm. The vocal tract length along the tongue surface (109 mm) was increased by 13 mm with the lower lip and by 12.5 mm with the upper lip.

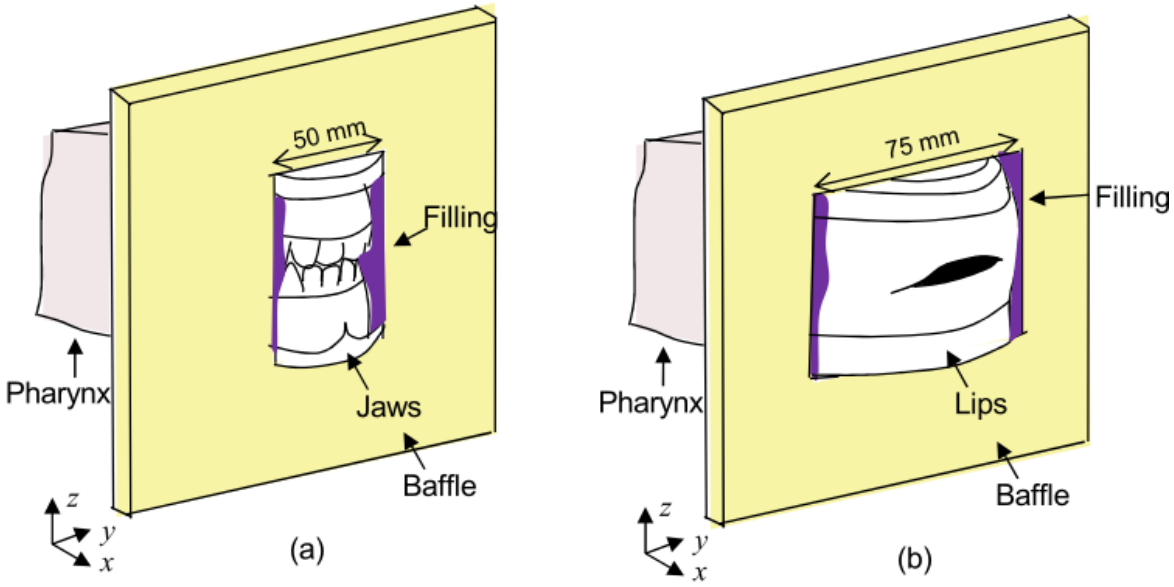


Fig. 2.1 Realistic replica with a rectangular baffle without lips (a) and with lips (b).

2.1.2 Simplified Replica

In this study, we explore the *subject-specific* simplified vocal tract geometry for Japanese sibilant fricatives /s/ and /ʃ/. The simplified vocal tract proposed by Howe and McGowan (2005) produced the similar frequency characteristics of English sibilant /s/. Based on the Howe's model, a novel simplified geometry was constructed in accordance with the knowledge obtained from the realistic replica (Nozaki *et al.*, 2014). Howe's model consisted of four cavities: throat, constriction, lower mouth cavity, and lip cavity. Thus, we proposed a model consists of a tongue, alveolar ridge, upper and lower teeth in a rectangular duct to form the four cavities. Schematic of the simplified model is depicted in Fig. 2.2.

The dimensions of the main duct were determined from the vertical height and transverse length of the subject's lip cavity (8 × 25 mm). Widths of the upper and lower teeth were set to 1 mm and the gap between the teeth was fixed to 1 mm. The heights of both teeth were 4 mm, while the upper teeth do not overlap the lower teeth. The alveolar ridge plate with a thickness of 1 mm was placed at the posterior position of the upper teeth. In order to form the constriction and cavity behind the upper teeth, the tongue model, which has a groove on the top, and tip at the front, was positioned between the back cavity and the teeth. Although the lower mouth cavity between the tongue blade and lower teeth was considered in Howe's model for the production of /s/, the lower mouth cavity was not observed for this subject. Therefore, the tongue tip was designed to fill up the cavity when the tongue model was positioned at the forefront in the vocal tract for /s/. The size of the groove and tip was changed in order to explore the difference between /s/ and /ʃ/. Since the main duct is straight from the back cavity to lip cavity, the tongue replica is movable so that we can examine the effect of tongue position on the acoustic properties of sound generated by the replica. At the outlet of the replica, a baffle (350 × 350 mm) was set to mimic a face.

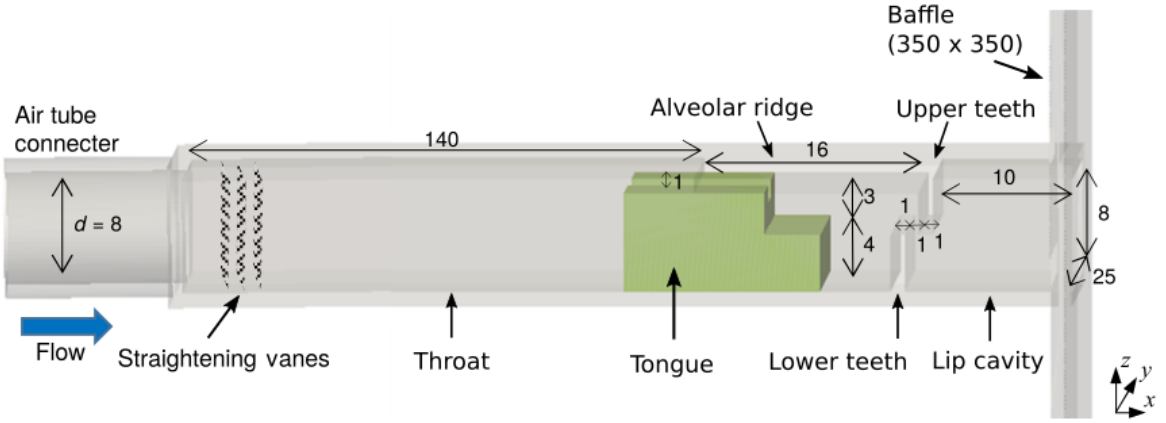


Fig. 2.2 Simplified vocal tract replica. Unit of the dimensions is mm.

In order to compare the vocal tract geometry of /s/ and /ʃ/, mid-sagittal images and coronal images of a 42-year-old male Japanese subject's vocal tract while pronouncing /s/ and /ʃ/ were measured by a 320-row Area Detector CT (20 fps; 320 slices of 512 × 512 pixels; 0.488 × 0.488 × 0.5 mm voxels, Tohshiba medical systems corporation, Ibaraki, Japan). The measured images are shown in Fig. 2.3 (a). The position of the coronal image is in the middle of the constriction (dotted line in mid-sagittal plane). The subject has normal dentition (Angle Class I) without any speech disorder (self-reported). The subject pronounced the Japanese word /mʲisoʃirɯ/ (misoshiru) for 0.6 s while the subject was sitting down. The images were collected during the middle of the articulation of both /s/ and /ʃ/.

In this replica, we assume that the acoustic difference between /s/ and /ʃ/ is formed by the changes of the constriction width and length, position of the constriction, and the amount of sublingual space (Perkell *et al*. 1979; Shadle, 1985). From the images, dimensions of the vocal tract were roughly estimated as depicted in Fig. 2.3 (a). Note that the length, width, and height below correspond to the dimensions in the *x* axis, *y* axis, and *z* axis, respectively. When the subject pronounced /s/, the constriction of height 1 mm, width 8 mm, and length 10 mm was formed at 4 mm posterior to the lower teeth. Meanwhile, when the subject pronounced /ʃ/, the constriction of height 1 mm, width 15 mm, and length 8 mm was formed at 6 mm posterior to the lower teeth. Four kinds of tongue models were proposed based on the estimated dimensions in Fig. 2.3 (b):

(1) /s/ constriction width 8 mm and /s/ tip 4 × 4 mm [Tongue model 1, Fig. 2.3 (b-1)],
(2) /ʃ/ constriction width 15 mm and /s/ tip 4 × 4 mm [Tongue model 2, Fig. 2.3 (b-2)],
(3) /s/ constriction width 8 mm without tip [Tongue model 3, Fig. 2.3 (b-3)],
(4) /ʃ/ constriction width 15 mm without tip [Tongue model 4, Fig. 2.3 (b-4)].

The total length and height of the tongue models were fixed to 14 and 7 mm, respectively. All parts were made of acrylic board.

The mid-sagittal plane (*x*-*z* plane at the center of the *y* axis) of the replica is depicted in Fig. 2.4. The variables corresponding to the geometrical factors are also depicted in Fig. 2.4. L_{CA} is the length between the lower teeth and the constriction. L_f is the length between

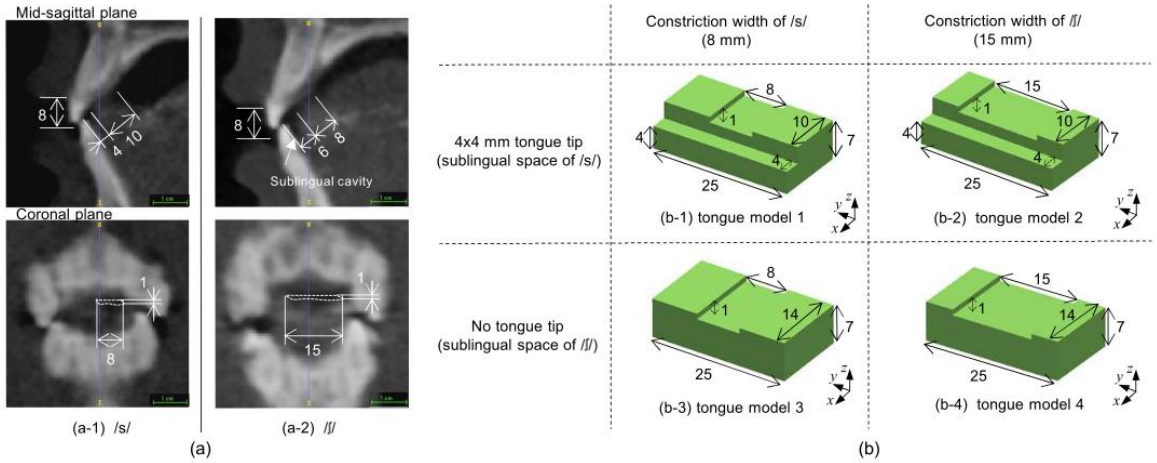


Fig. 2.3 (a) Mid-sagittal plane and coronal plane of the computed tomographic images of the subject pronouncing /s/ (a-1) and /ʃ/ (a-2). The estimated dimensions of the vocal tract are depicted on the images. (b) Four kinds of proposed tongue models. Unit of the dimensions is mm.

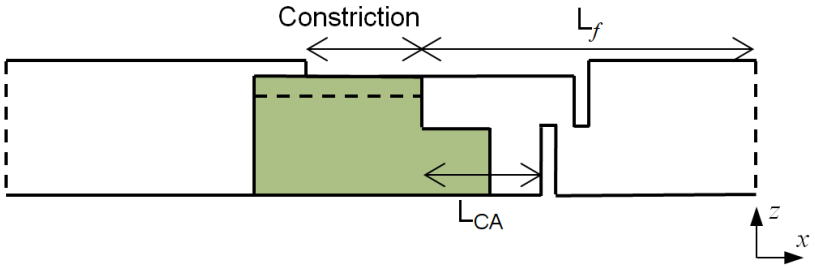


Fig. 2.4 Medium plane of the simplified vocal tract model.

the constriction and outlet of the replica. The tongue model was moved in 1 mm increments along the x axis from L_{CA} = 4 to 11 mm for tongue models 1 and 2, and L_{CA} = 0 to 6 mm for tongue models 3 and 4.

2.1.3 Experimental Setups for Directivity Measurement

The sound generated from a compression driver (SP-DYN-PRO2, Sphynx) positioned at the inlet of the pharynx was measured along two semi-circles of radius 4 cm (near-field) and radius 48 cm (far-field) from the lips. A compression driver is used instead of a loud speaker in order to ensure sufficient acoustic energy in the frequency range of interest. The compression driver was connected to the inlet of the replica using a communication hole of 2 mm diameter as shown in Fig. 2.5. The replica was mounted near the inlet of a quasi-anechoic chamber (internal volume 7.45 m^3) (Van Hirtum and Fujiso, 2012) as schematized in Fig. 2.6 (a) for the near-field measurement and Fig. 2.6 (b) for the far-field measurement. The size of the baffle was changed from 36 × 36 cm for the near field measurement to 160 × 130 cm (y × z) for the far-field measurement.

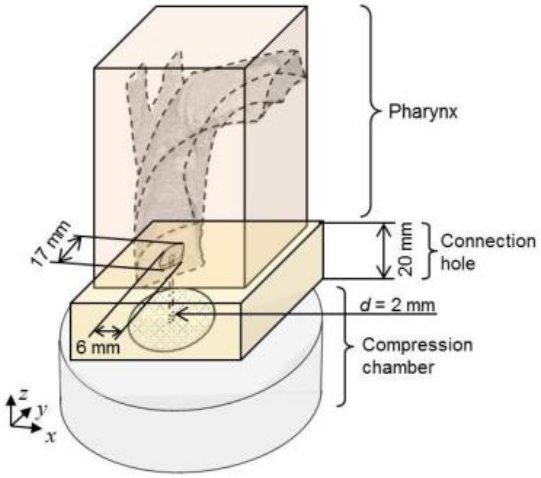


Fig. 2.5 Inlet geometry for the case with acoustic source.

In the near-field, a single microphone probe with a probe length 25 mm (flat frequency response up to 14 kHz; type 4182, B&K) was mounted on a spatial 3D positioning system (PS35, OWIS; accuracy: ±100 µm). The tip of the probe was positioned along the transverse plane and sagittal plane at positions from 0° to 180° in steps of 2° as shown in Fig. 2.7 (a) and (b). In the far-field, a single 1/2-inch microphone with flat frequency response up to 20 kHz (type 4192 and pre-conditioner type 5935L, B&K) was positioned at the same height as the replica with a rectangular bar (length: 1 m). The position of the microphone was shifted along the transverse plane (*x*-*y* plane) from 0° to 180° in steps of 15°. When the sound along the sagittal plane (*x*-*z* plane) of the replica was measured, the replica was tipped on its side so that the semi-circular transit of the microphone was in the replica's sagittal plane. Considering the wavelength of the generated sound, the minimum frequency of the far-field measurement at 48 cm is approximately 720 Hz.

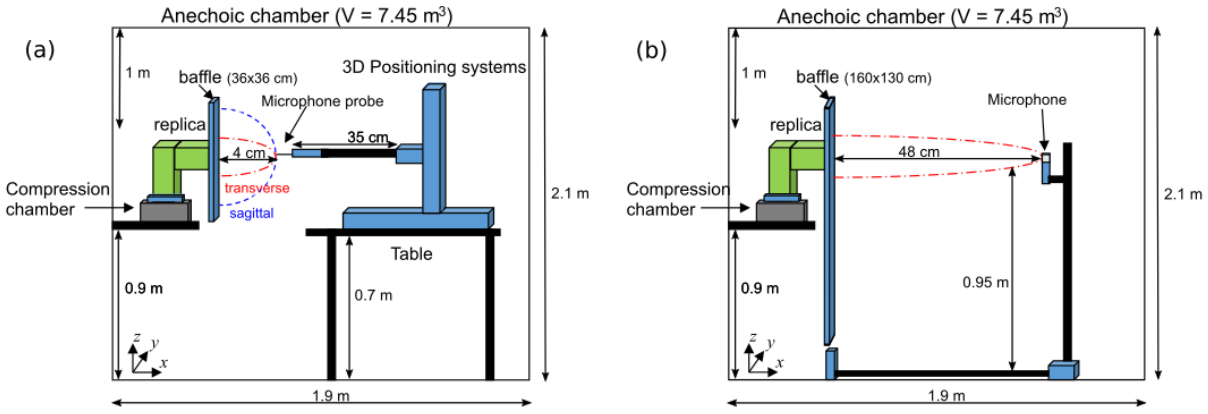


Fig. 2.6 Illustration of the experimental setup used for pressure distribution measurements with acoustic source in near-field (a), in far-field (b).

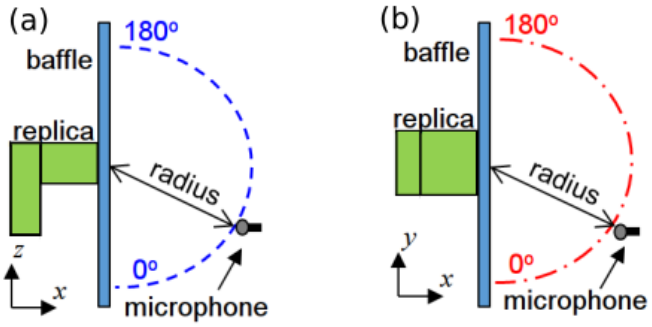


Fig. 2.7 Measurement positions in the sagittal (a) and transverse (b) planes.

At each microphone position, a linear sweep signal was amplified (A-807, Onkyo) and emitted by the compression driver in the frequency range of 2-15 kHz for 20 s. At each position, the microphone signal was recorded with a sampling frequency of 44.1 kHz using a data acquisition system with multiple input and output channels (NI PXI0MIO 16 XE, National Instruments). The temperature in the quasi-anechoic room during the measurements ranged from 20.4 to 21.2 °C yielding a sound speed of 344.0±0.3 m/s during the measurements.

The directivity was also measured with the airflow supply. Air was supplied using a flow facility consisting of 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 × 40 × 50 cm^3). The settling chamber was tapered with acoustic foam (SE50-AL-ML, Elastomeres Solutions) and equipped with flow straighteners in order to avoid acoustic resonances (due to the flow facility setup or settling chamber) and to homogenize the flow. Air issuing from the settling chamber was introduced to the replica through a silicone tube with a diameter of 1 cm and a length of 59 cm. The inlet geometry of the replica is depicted in Fig. 2.8. The turbulence level at the inlet of the replica was less than 3% (Van Hirtum, *et al.*, 2014). Experiments were conducted at a flow rate of 40 $L \cdot min^{-1}$, which corresponds to the subject's medium flow rate of /s/ in three effort levels, soft, medium and loud (Fujiso *et al.*, 2015). The flow rate was varied from 35 to 45 $L \cdot min^{-1}$ while the subject sustained /s/ at the medium level. The corresponding Reynolds number is 5636 based on the characteristic dimension of the sibilant groove (the height of the maximum constriction 1.25 mm) and flow velocity at the groove (Nozaki *et al.*, 2014). The size of the baffle is 36 × 36 cm.

A single 1/2-inch microphone (type 4192, B&K) was positioned along a semi-circle of radius 10 cm using the 3D positioning system (PS35, OWIS), as schematized in Fig. 2.9. The radius was increased compared to the near-field measurement with the acoustic sound source in order to prevent the impingement of the flow on the microphone. The spatial positioning system was placed on the floor of the acoustic chamber and covered with acoustic foam to minimize disturbances of the acoustic field. The microphone was attached to the positioning system with a bar (length: 1 m) and positioned along the transverse plane (*x*-*y* plane) and sagittal plane (*x*-*z* plane) from 0° to 180° in steps of 2° as shown in Fig. 2.7 (a) and (b). Considering the wavelength, the minimum frequency of the far-field measurement at 10 cm is approximately 3440 Hz. At each position, the microphone signal

was recorded for 2 s with a sampling frequency of 44.1 kHz using the data acquisition system. With the recorded signals, power spectral densities (PSDs) were calculated by time-averaging the Fourier transforms of 60 Hanning-windowed time segments of 1024 sample points with 30% overlap. To allow comparison between pressure patterns measured with and without flow, measured values were again normalized by the spatial mean or maximum amplitude for each frequency. The temperature in the quasi-anechoic room ranged from 18.7 to 20.9 °C yielding a sound speed of 343.5 ± 0.7 m/s during the measurements.

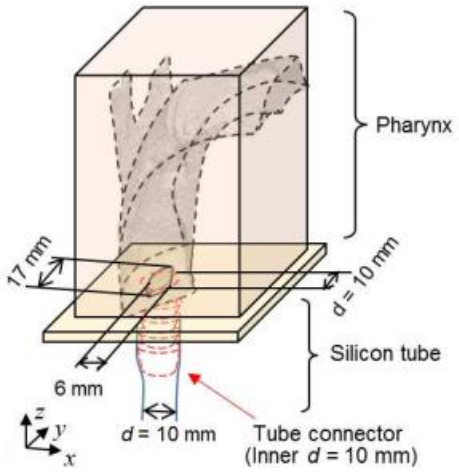


Fig. 2.8 Inlet geometry for the case with flow.

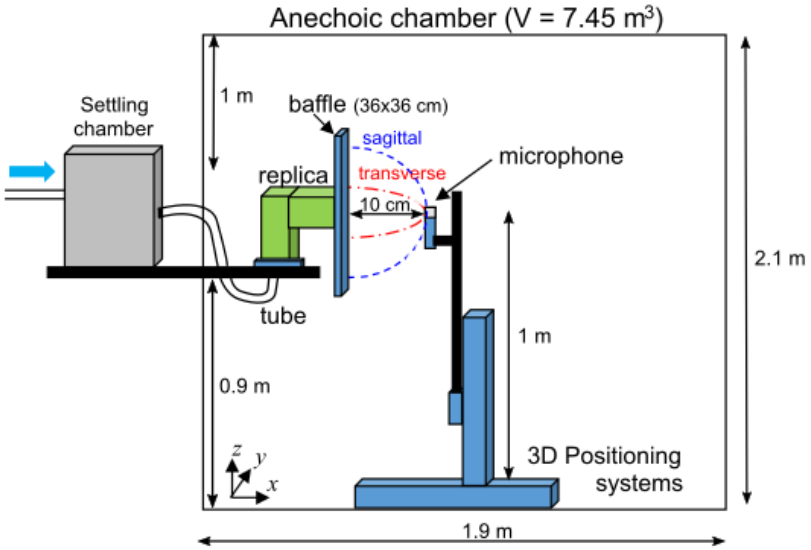


Fig. 2.9 Illustration of the experimental setup used for pressure distribution measurements with acoustic source with flow.

2.1.4 Experimental Setups for Simplified Replica

Steady airflow was delivered to the flow channel of the replica via a compressor (YC-4RS, Yaezaki, Tokyo, Japan) through a mass flow controller (MQV0050, Azbil, Tokyo, Japan). The diameter of the air tube was 16 mm. The tube was connected to the replica with a tube connecter that had an inner diameter of 8 mm. The flow rate was fixed to the subjects average flow rate of sustained /s/ 400 cm^3/s, which was measured with a mass flow meter (Series 4000, TSI, MN). Straightening vanes were set at the inlet of the replica as illustrated in Fig. 2.2. The far field sound generated by the replica was measured by a 1/4-inch omnidirectional microphone (type 4939, Bruel & Kjaer, Nærum, Denmark), which was placed 30 cm along the *x* axis from the outlet of the replica. In addition, the sounds of the subject's 2-s-sustained /s/ and /ʃ/ were measured by the microphone at the same position. The microphone had a flat response from 4 Hz to 100 kHz (61 dB) and dynamic range of 28 to 164 dB. The 2-s-sustained sounds were recorded with a sampling frequency of 44.1 kHz using a data acquisition system (PXIe-4492, National Instruments, TX).

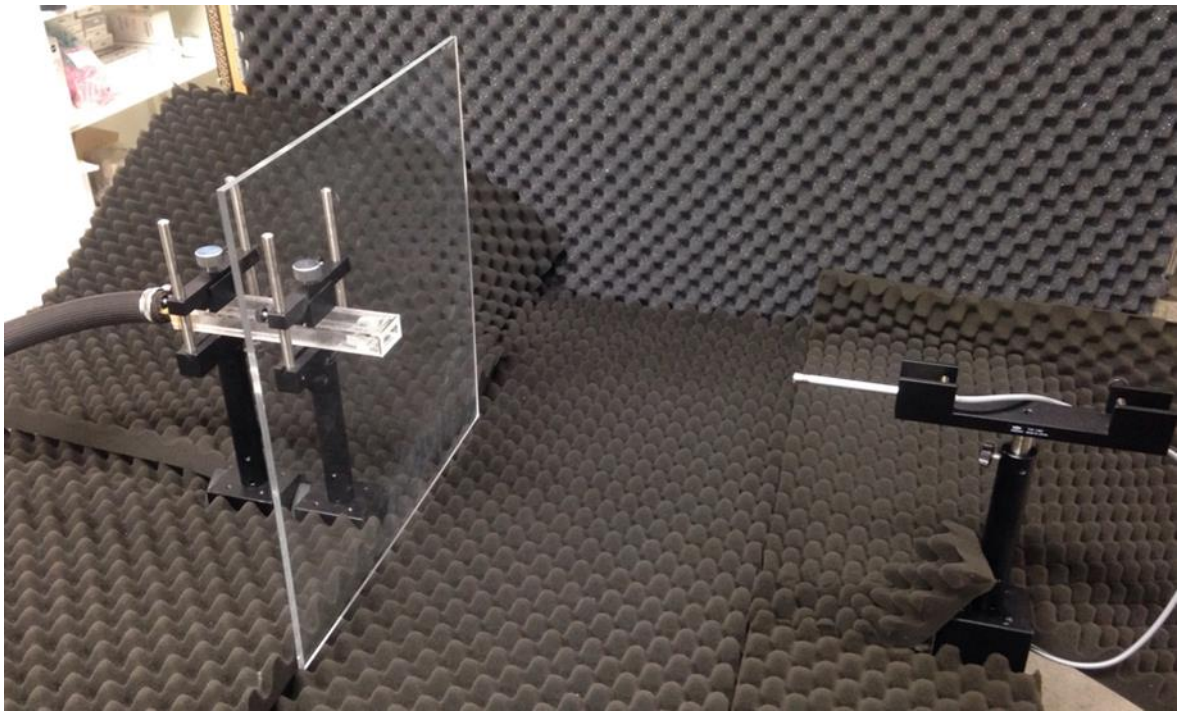


Fig. 2.10 Experimental setups of the sound measurement for simplified replica.

2.1.5 Signal Analysis

For the directivity measurement (section 2.1.3), power spectrum densities (PSDs) were calculated with 863 time segments of Fourier transforms of 1024 sample points (total time 20 s). Measured acoustic pressures were then characterized by considering the PSD amplitude extracted for each frequency. Since the amplitude of the sound emitted by the compression driver depends on the frequency, measured values needed to be normalized for each frequency. For the contour map (Fig. 2.14 and 2.15 shown below), the extracted amplitudes were normalized by their spatial mean of the pressure pattern at each frequency. For the angular plot (Fig. 2.16, 2.17, and 2.18 shown below), the extracted amplitudes were normalized by their maximum value of the pressure pattern at each frequency. This normalization removes the spurious influence of the sound source which can hinder interpretation of the results.

For the measurement of the simplified replica, spectra of the sounds were calculated using a fast Fourier transform with 512-point signals, which were multiplied by a Hanning window, and a time average of 200 sets of 30% overlap. The sound pressure level (SPL) was calculated based on the reference level 20×10^6 Pa. The spectral mean and overall sound pressure level (OASPL) were calculated in the frequency range 0.5 to 15 kHz.

2.2 Experimental Results

2.2.1 Effects of Lip Horn of the Realistic Replica

The spectra of the sound generated by the realistic replica with acoustic source are shown in Fig. 2.11 for the microphone position at an angle of 90° in transverse plane. Because the sound was generated from the compression driver, the spectral features are different from the actual sibilant /s/. By removing the lips, the amplitude decreased in the frequency ranges of 3.5–5.5 kHz and 9.5–11.5 kHz. In contrast, the amplitude at a frequency of approximately 8 kHz increased by 5 dB.

The spectra of sounds generated by flow supply for the realistic replica with and without lips are presented in Fig. 2.12 for the microphone position at an angle of 90° in transverse plane. For both with and without lips, the spectral energy increases in the range of 3–7 kHz; the amplitude reaches a maximum at 8 kHz with a large trough at 9 kHz; and the spectral energy gradually decreases between 9 and 15 kHz. The amplitude for the case without lips is approximately 5 dB lower than for the case with lips in the frequency ranges of 3–7 kHz and 9–14 kHz, and the minimum amplitude observed at 8.6 kHz without lips is shifted to a higher frequency (by approximately 0.3 kHz) when the lips are present.

To clarify the effect of the lips in the realistic replica, the spectral differences (PSDs measured with lips minus those measured without lips) are shown in Fig. 2.13 for the cases of flow and with the acoustic source. Although the overall tendency is similar for the two sound generation mechanisms, some differences were observed in the frequency ranges of 5.5–7 kHz and 12–14 kHz.

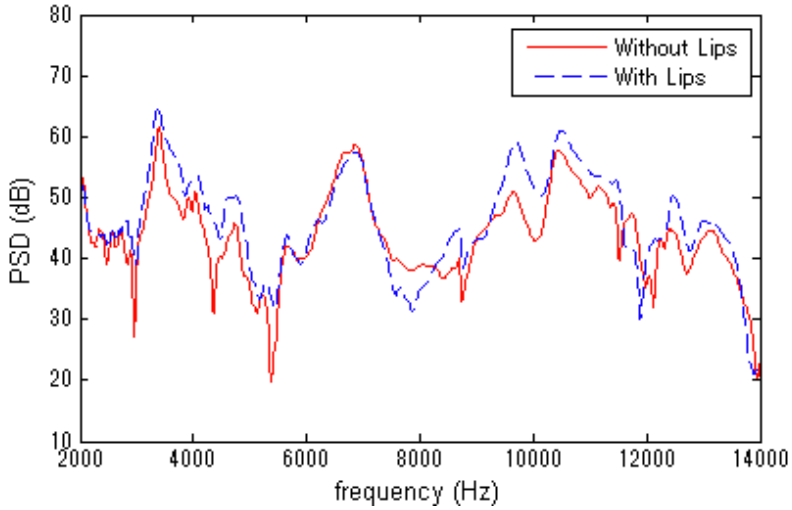


Fig. 2.11 Frequency spectrum of the sound generated with an acoustic source for the replica with and without lips. The microphone was positioned 4 cm from the exit of the replica at 90° in the transverse plane.

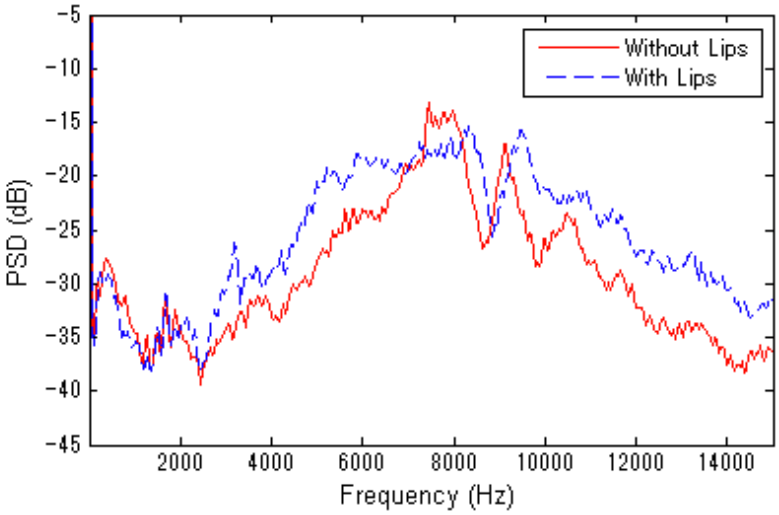


Fig. 2.12 Frequency spectrum of the sound generated with flow for the replica with and without lips. The microphone was positioned 10 cm from the exit of the replica at 90° in the transverse plane.

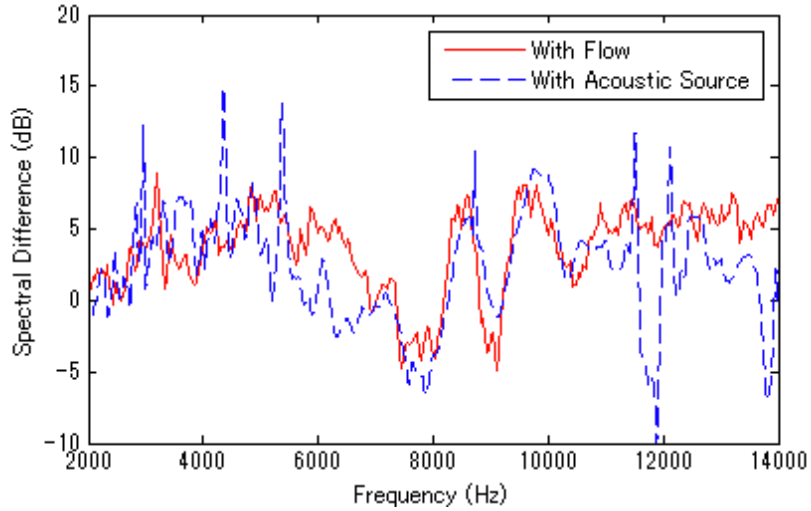


Fig. 2.13 Spectral differences between sounds generated with and without lips. The amplitudes (PSDs) measured with lips were subtracted from those measured without lips

Normalized acoustic pressure amplitudes of the sounds generated by the replica with and without lips along the transverse plane are mapped in Fig. 2.14 for the case with acoustic source in near-field (a), (b), in far-field (c), (d), and for the case with flow (e), (f). The pressure distributions observed with and without lips were mainly changed above 5 kHz for all measurement conditions. The amplitudes at the angle around 90° increased and the amplitudes near 0° and 180° decreased when the lips were present. The position of the maximum pressure amplitude was shifted from angle 0° to 140° above 12 kHz for the case with acoustic source.

Normalized acoustic pressure amplitudes of the sounds generated by the replica with and without lips along the sagittal plane are mapped in Fig. 2.15 for the case with acoustic source in near-field (a), (b), in far-field (c), (d), and for the case with flow (e), (f). For the

case without lips, the large trough at angles between 100° and 140° in the frequency range from 9 up to 10 kHz was observed for all measurement conditions. The amplitudes at the angle around 90° increased and the amplitudes at the angle around 0° and 180° decreased above 5 kHz when the lips were present.

The acoustic pressure amplitudes measured with and without the lips at 6.1 kHz and 9.5 kHz are plotted in Fig. 2.16 as a function of angle for the near-field measurement with acoustic source. The amplitudes in each measurement were normalized by the maximum value at each frequency. In the transverse plane, the amplitudes at the angle around 30° and 160° at 6.1 kHz decreased by 15 dB and 10 dB, respectively, when the lips were present. The amplitudes at the angle around 0° and 170° at 9.5 kHz decreased by 8 dB and 5 dB, respectively, when the lips were present. In the sagittal plane, the amplitudes at the angle around 0° and 150° at 6.1 kHz decreased by 7 dB and 10 dB, respectively, when the lips were present. The amplitudes at the angle around 20° and 110° at 9.5 kHz decreased by 10 dB and 15 dB, respectively, when the lips were present.

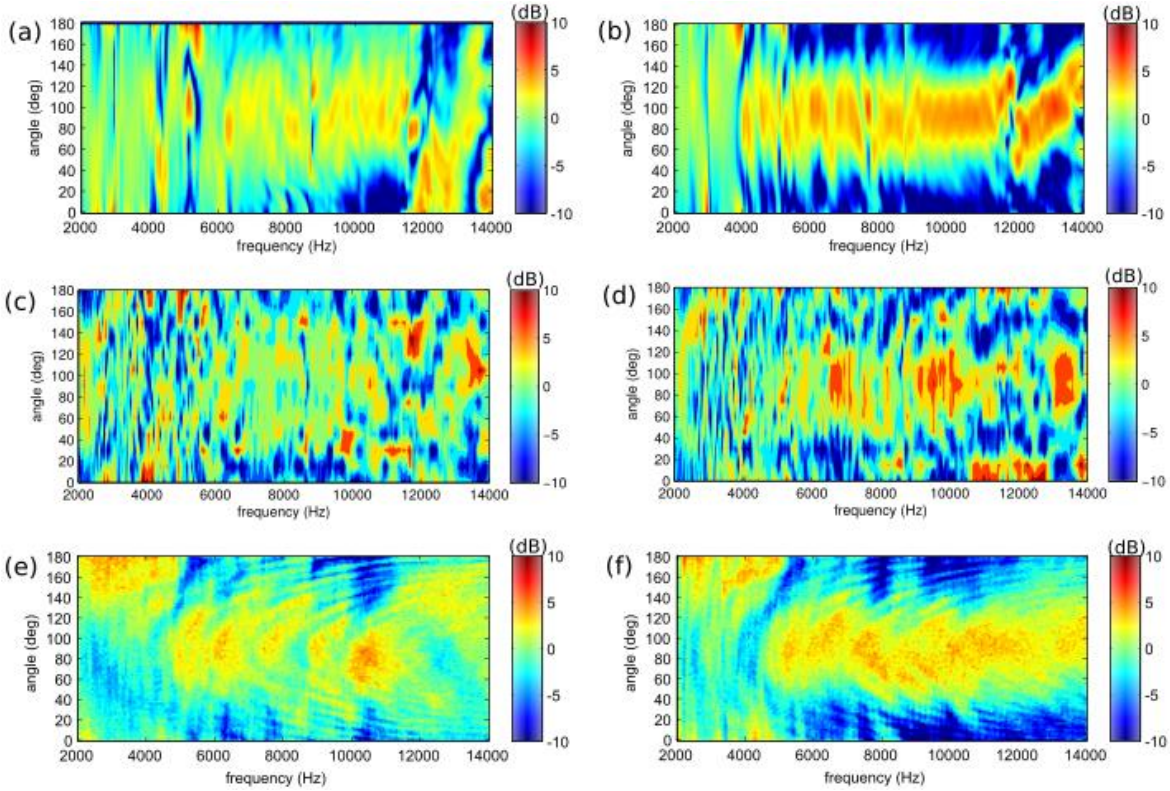


Fig. 2.14 Normalized acoustic pressure amplitudes measured along the transverse plane for the case with acoustic source in near-field (radius 4 cm, every 2°) without lips (a), with lips (b), in far-field (radius 48 cm, every 15°) without lips (c), with lips (d), and for the case with flow (radius 10 cm, every 2°) without lips (e), with lips (f).

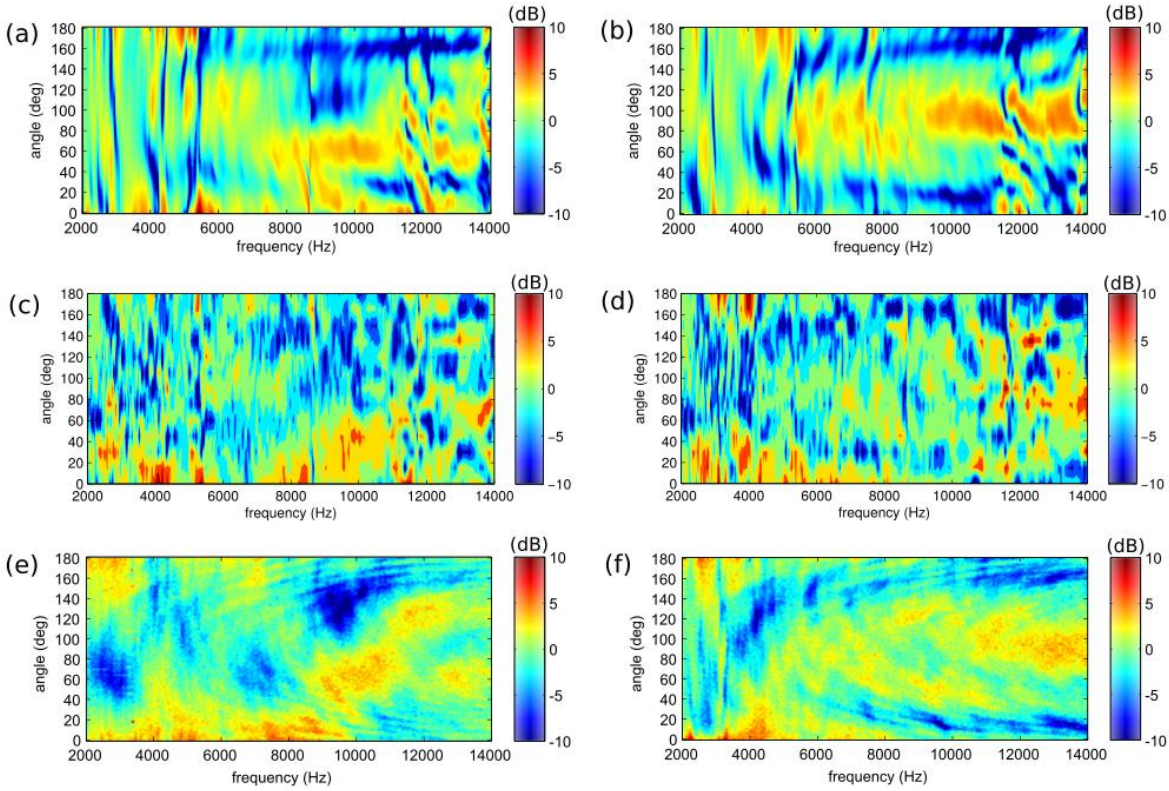


Fig. 2.15 Normalized acoustic pressure amplitudes measured along the sagittal plane for the case with acoustic source in near-field (radius 4 cm, every 2°) without lips (a), with lips (b), in far-field (radius 48 cm, every 15°) without lips (c), with lips (d), and for the case with flow (radius 10 cm, every 2°) without lips (e), with lips (f).

The acoustic pressure amplitudes measured with lips at 6.1 kHz and 9.5 kHz are plotted in Fig. 2.17 as a function of angle for the near-field and far-field measurements with acoustic source. As before, the amplitudes in each measurement were normalized by the maximum value at each frequency. In the transverse plane, the amplitudes in far-field at the angle around 0° and 180° at 6.1 kHz were 10 dB larger than those in the near-field. The amplitudes at the angle around 0° and 180° at 9.5 kHz were 15 dB smaller than the amplitudes at the angle around 90° for both far-field and near-field. In the sagittal plane, the trough in amplitudes at the angle 150° was observed at 6.1 kHz for both far-field and near-field. The troughs in amplitudes at angles around 20° and 175° were observed at 9.5 kHz for both far-field and near-field.

The acoustic pressure amplitudes measured with lips at 6.1 kHz and 9.5 kHz are plotted in Fig. 2.18 as a function of angle for the case with acoustic source and flow. As before, the amplitudes in each measurement were normalized by the maximum value at each frequency. In the transverse plane, the amplitudes measured with flow at the angle 30° and 160° at 6.1 kHz were 5 dB and 10 dB larger, respectively, than those measured with acoustic source. The amplitudes at angles around 0° and 180° at 9.5 kHz were 15 dB smaller than that at angles around 0° for both the flow case and the acoustic source case. In the sagittal plane, the amplitudes measured with flow at the angle 15° and 150° at 6.1 kHz were 7 dB larger than those measured with acoustic source. The troughs in amplitude at the

angle around 20° and 175° were observed at 9.5 kHz for both cases, *i.e.* with acoustic source and with air flow.

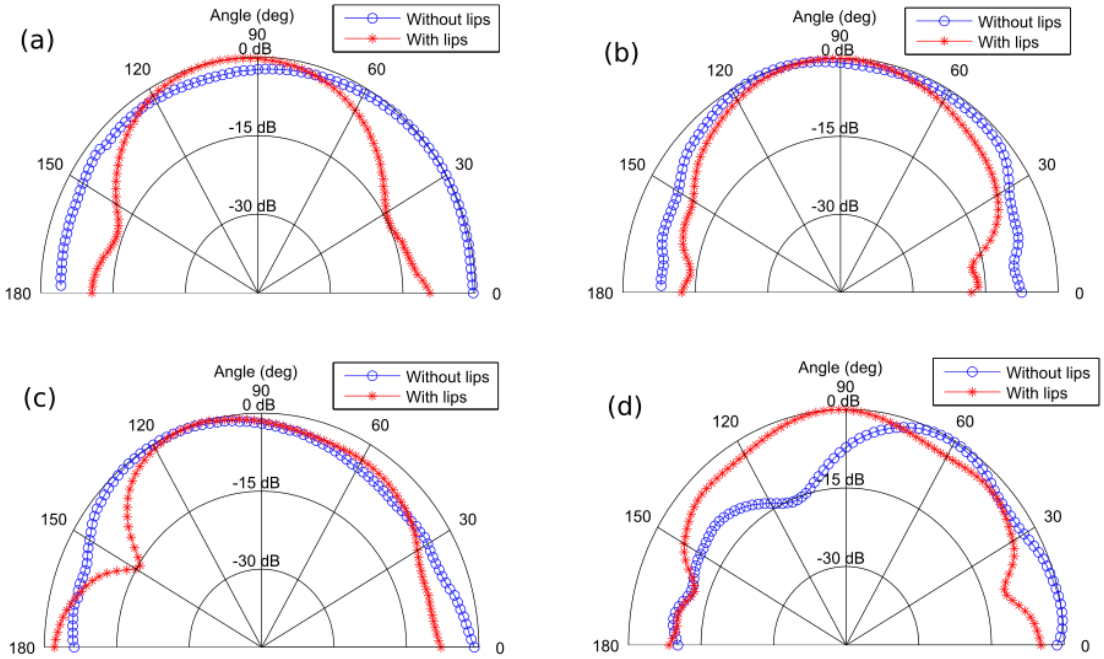


Fig. 2.16 Normalized acoustic pressure amplitudes as a function of angle for the near-field (radius 4 cm) measurement with acoustic source. Measurements along the transverse plane are plotted in (a) at 6.1 kHz, and (b) at 9.5 kHz. Measurements along the sagittal plane are plotted in (c) at 6.1 kHz, and (d) at 9.5 kHz.

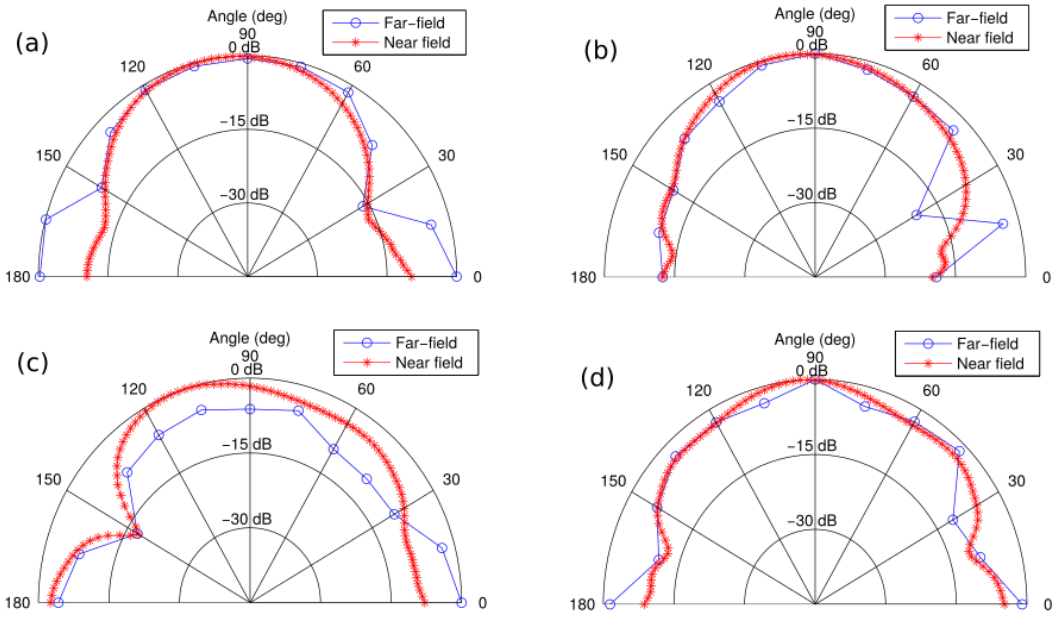


Fig. 2.17 Normalized acoustic pressure amplitudes as a function of angle for the far-field (radius 48 cm) and near-field (radius 4 cm) measurements with acoustic source and lips. Measurements along the transverse plane are plotted at (a) 6.1 kHz, and (b) 9.5 kHz. Measurements along the sagittal plane are plotted at (c) 6.1 kHz, and (d) 9.5 kHz.

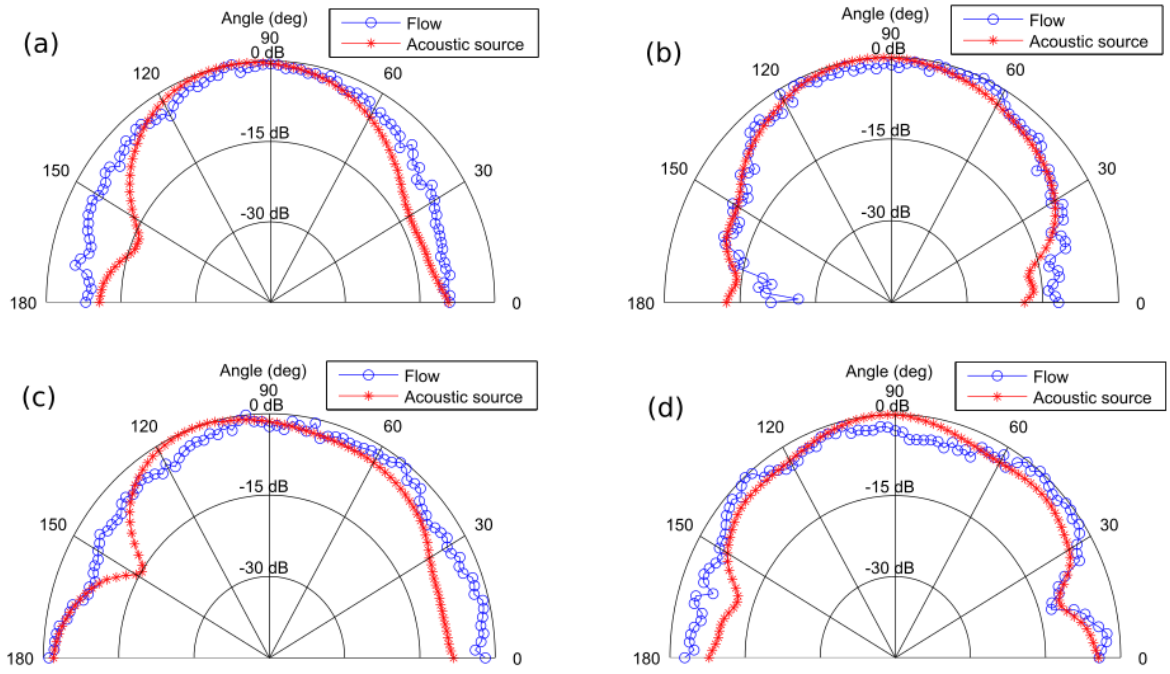


Fig. 2.18 Normalized acoustic pressure amplitudes as a function of angle for the measurements with acoustic source (radius 4 cm) and flow (radius 10 cm). Both plots are for the case with lips. The measurements along the transverse plane are plotted at (a) 6.1 kHz, and (b) 9.5 kHz. The measurements along the sagittal plane are plotted at (c) 6.1 kHz, and (d) 9.5 kHz.

2.2.2 Effects of Tongue Position in the Simplified Replica

The frequency spectrum of sound generated by the simplified replica with tongue model 1 (/s/ constriction width 8 mm and /s/ tip 4 × 4 mm, see Fig. 2.3) is shown in Fig. 2.19. The spectrum of /s/ pronounced in word context /usui/ by the subject of CT images is also plotted in Fig. 2.19. When the tongue was positioned at L_{CA} = 4 mm, broadband noise in the frequency above 4 kHz was generated by the simplified replica. The differences in SPLs between the replica and the subject were below 6 dB in the frequency range 2 to 10 kHz. The spectra of sound measured when L_{CA} = 4, 7, 11 mm are plotted in Fig. 2.20. The spectrum of back ground noise (BGN) is also plotted in Fig. 2.20. By shifting the tongue position in the posterior direction from L_{CA} = 4 to 11 mm, the frequency range of the noise was shifted to a lower frequency range with a minimum of 1.5 kHz. The characteristic peaks of the sibilant fricatives, which is one of the major factors that we use to distinguish the /s/ and /ʃ/ sounds (Jesus and Shadle, 2002), were observed at 4.8 kHz when L_{CA} = 4 mm. By shifting the tongue from L_{CA} = 4 to 11 mm, the peak frequency decreased from 4.8 kHz to 3.2 kHz.

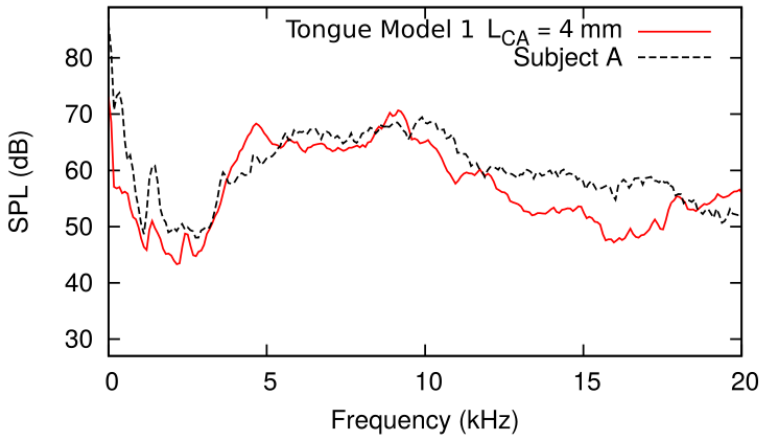


Fig. 2.19 Spectra of sound generated by the simplified replica with tongue model 1. The sound of /s/ pronounced by the subject of CT images is also plotted. The sound was measured at 30 cm from the lip outlet.

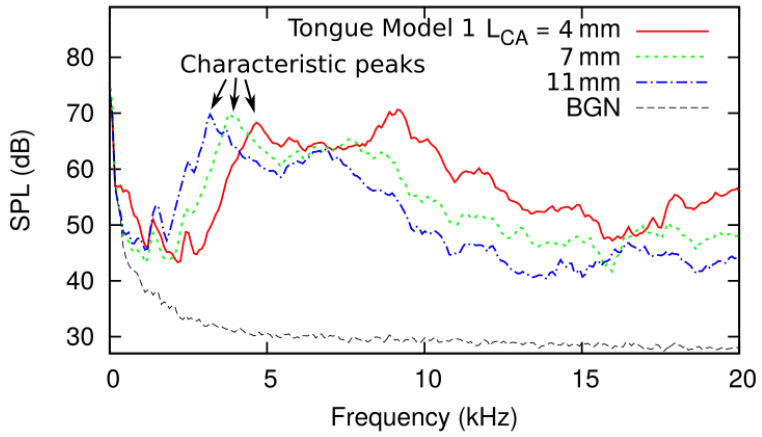


Fig. 2.20 Spectra of sound generated when the tongue model 1 was at L_{CA} = 4, 7, 11 mm. The amplitude of back ground noise (BGN) is also plotted.

The characteristic peak frequency, spectral mean, and OASPL of the sound spectra measured for the four tongue models at seven positions ($0 \leq L_{CA} \leq 10$ mm) are shown in Fig. 2.21. The values calculated from the subject's sustained /s/ and /ʃ/ are plotted in Fig. 2.21 with L_{CA} observed in CT images of Fig. 2.3 (a) [mean ± standard deviation (S.D.) for five trials]. The peak frequencies for tongue models 1 and 2 decreased from 4.7 to 3.8 kHz while the L_{CA} increased from 4 to 10 mm. The peak frequencies for tongue models 3 and 4 decreased from 5.6 to 3.7 kHz while the L_{CA} increased from 0 to 6 mm. The mean value of the subjects peak frequency was 4.8 kHz for /s/ at L_{CA} = 4 mm and 3.7 kHz for /ʃ/ at L_{CA} = 6 mm. Using the L_f and assuming the constriction as an acoustically-closed duct end, the main resonance frequency was calculated with 1/4 wavelength resonances and end correction

$$f = \frac{c}{4(L_f + \delta)}, \tag{2.1}$$

where c is the speed of sound (344 m/s) and δ is the end correction of the rectangular channel $\delta = 3.4$ mm (Ingard, 1953). The frequency of Eq. (2.1) plotted in Fig. 2.21 (a) decreased from 5.2 to 3.2 kHz while the L_{CA} increased from 0 to 10 mm (L_f from 13 to 23 mm).

The spectral means decreased as follows: 8.5 to 5.4 kHz for tongue model 1, 7.8 to 5.2 kHz for tongue model 2, 11.0 to 6.0 kHz for tongue model 3, and 9.5 to 5.5 kHz for tongue model 4. The mean value of the subjects spectral mean was 8.5 kHz for /s/ at L_{CA} = 4 mm and 5.4 kHz for /ʃ/ at L_{CA} = 6 mm. The OASPLs for tongue models 1 and 2 decreased from 93 to 89 dB and 84 to 82 dB, respectively, while the L_{CA} increased from 4 to 10 mm. The peak frequencies for tongue models 3 and 4 first increased from 89 to 95 dB and 83 to 89 dB for L_{CA} = 0 to 2 mm. They were later decreased from 95 to 92 dB and 89 to 86 dB for L_{CA} = 2 to 6 mm, respectively. The mean value of the subjects OASPL was 92 dB for /s/ at L_{CA} = 4 mm, and 87 dB for /ʃ/ at L_{CA} = 6 mm.

The spectra of sound generated by each tongue model at the position where the spectrum has a closest spectral mean value to the subject's /s/ and /ʃ/ are shown in Fig. 2.22 and Fig. 2.23, respectively. The subject's spectra are chosen from the one that has the closest spectral shape to the model's spectra (tongue model 1 for /s/ and tongue model 4 for /ʃ/). The SPL of tongue model 2 at L_{CA} = 4 mm was approximately 3 to 10 dB smaller than that of subject's /s/ whereas the SPL of tongue model 3 at L_{CA} = 2 mm was approximately 5 to 10 dB larger than that of subject's /s/ above 4 kHz. The SPLs of tongue model 1 at L_{CA} = 4 mm and tongue model 4 at L_{CA} = 2 mm were roughly the same as that of the subjects /s/, though the sharp peaks were observed for tongue model 4 at 9 and 11.5 kHz. The SPLs of tongue model 1 at L_{CA} = 10 mm and tongue model 3 at L_{CA} = 6 mm were approximately 2 to 10 dB and 5 dB to 15 dB larger, respectively, than that of the subjects /ʃ/ above 3 kHz. The SPLs of tongue model 2 at L_{CA} = 9 mm and tongue model 4 at L_{CA} = 6 mm were roughly the same as that of subjects /ʃ/.

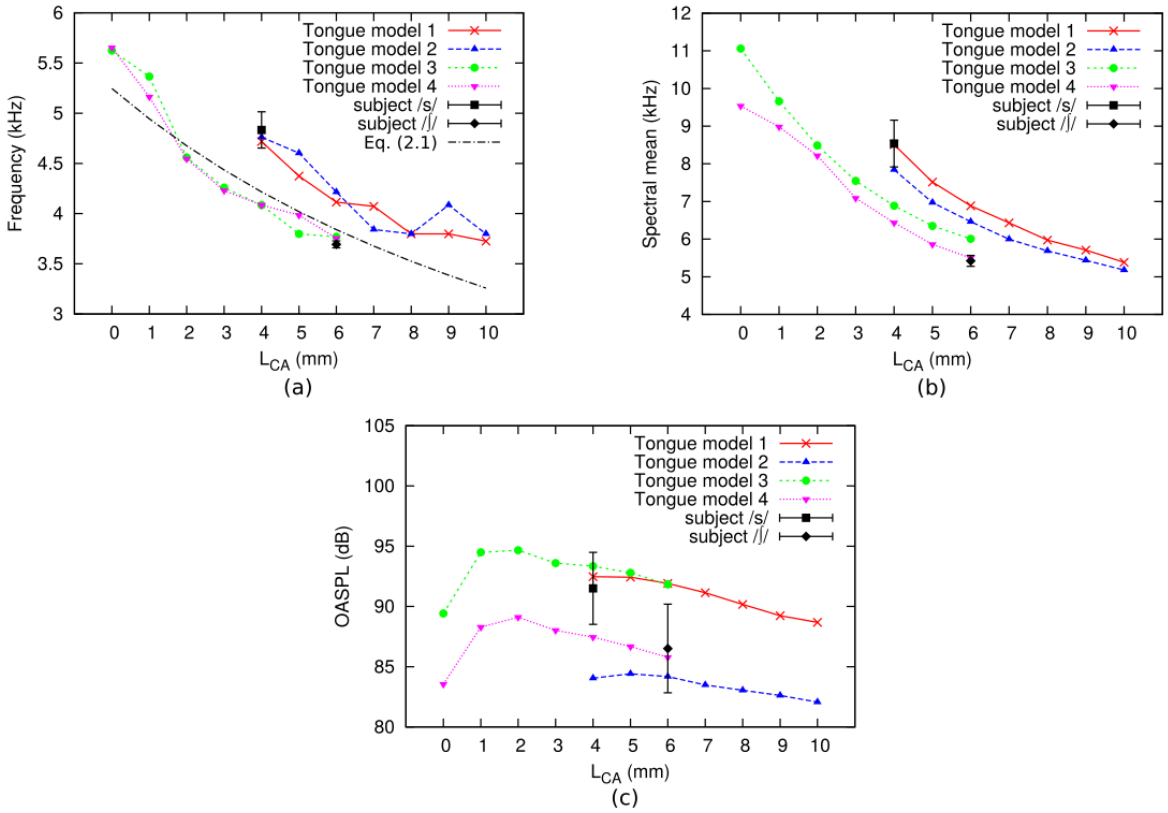


Fig. 2.21 The main peak frequency (a), spectral mean (b), and OASPL (c) of the sound spectra measured for the four tongue models at seven positions ($0 \leq L_{CA} \leq 10$ mm). The subject's mean ± S.D. values (n = 5) of sustained /s/ and /ʃ/ are also plotted with L_{CA} observed in CT images of Fig 2.3 (a). The frequency calculated with Eq. (2.1) is plotted in (a).

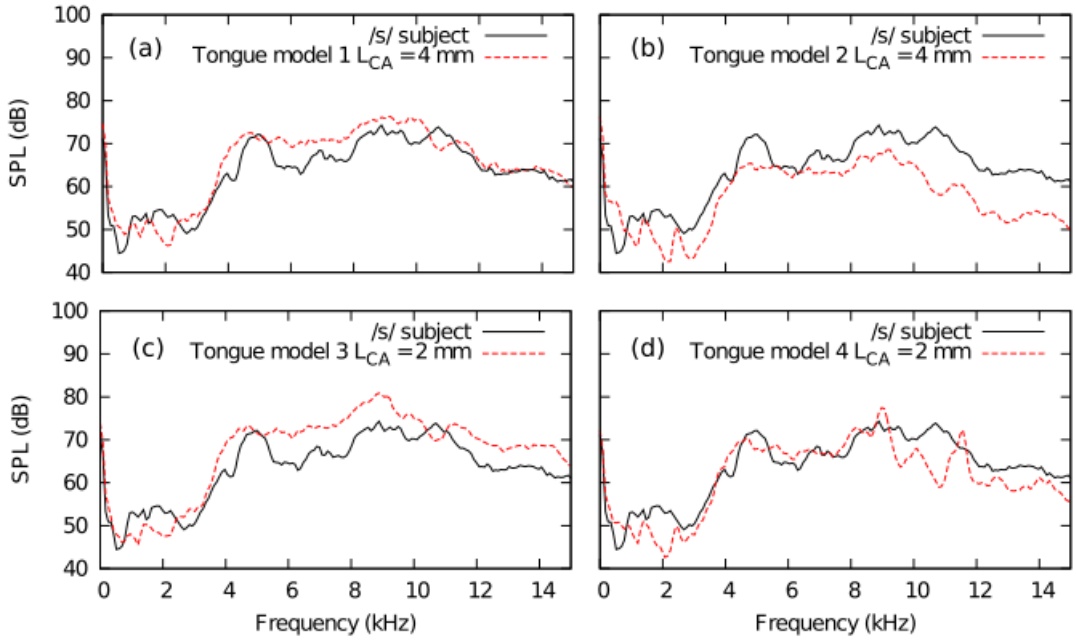


Fig. 2.22 The spectra of sound generated by tongue model 1 (a), tongue model 2 (b), tongue model 3 (c), and tongue model 4 (d) at the position where the spectrum has a closest spectral mean value to the subject's /s/. Each of the subject's sustained /s/ is plotted.

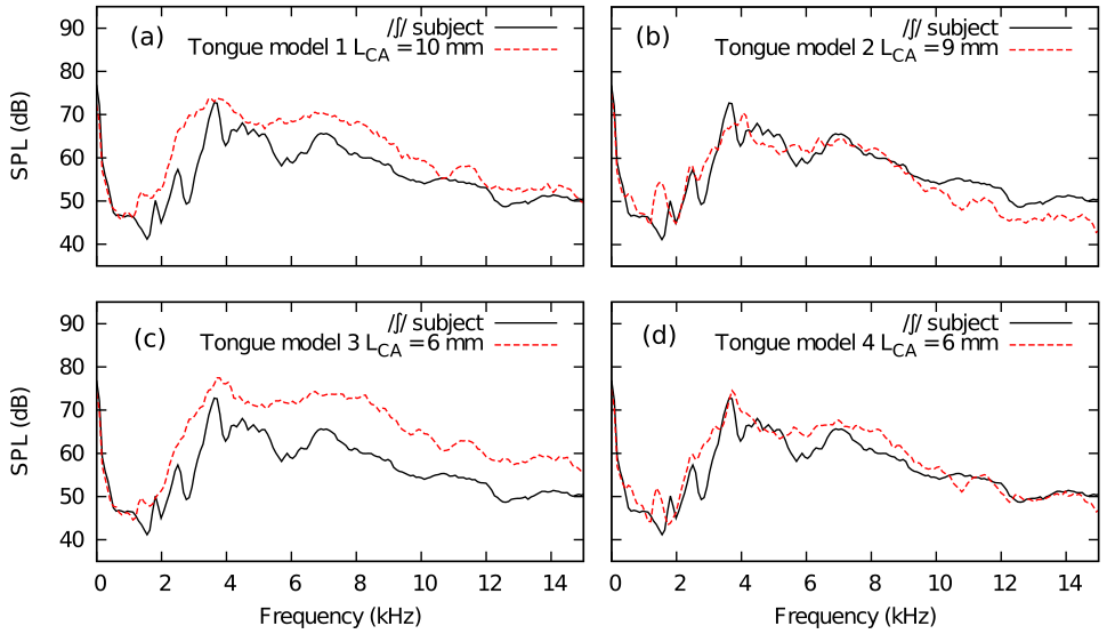


Fig. 2.23 The spectra of sound generated by tongue model 1 (a), tongue model 2 (b), tongue model 3 (c), and tongue model 4 (d) at the position where the spectrum has a closest spectral mean value to the subject's /ʃ/. Each of the subject's sustained /ʃ/ is plotted.

2.3 Discussion

2.3.1 Realistic Replica of Sibilant /s/

The general spectral tendencies of sound generated by the realistic replica with lips and flow supply corresponded to sibilant /s/ uttered by the subject of CT images (Nozaki *et al*., 2014) as well as the English speakers (Shadle and Scully, 1995). This result indicates that the realistic replica is valid to study the acoustic properties of sibilant /s/. By comparing the realistic replica with and without lips, some important differences in generated sounds were observed. First, amplitudes in the frequency range 3 to 7 kHz and 9 to 15 kHz were decreased. In addition, the large trough at 9 kHz was shifted to lower frequency range when the lips were absent. These results indicate that the changes in geometry of the replica with and without lips produced the difference in acoustic resonance and source generation inside the vocal tract. In particular, shift of the peak and trough in the high frequency range observed in Fig. 2.11 and Fig. 2.12 were caused by the higher order acoustic mode which is affected by the both geometry and source position (Motoki, 2013). We note that the observed frequency shift is not due to a change in temperature during the experiments (estimated shift is 34 Hz). Further investigation on the cause of spectral difference is possible by applying the multimodal modeling (see Chapter 3) on the realistic vocal tract geometry.

When the sound was generated by the compression driver, overall spectral differences between the replica with and without lips were similar to those measured with the flow source. This indicates that the differences of the frequency characteristics between the replica with lips and without lips were mainly caused by changes in the vocal tract geometry. Meanwhile, the spectral differences in the frequency ranges 5.5–7 kHz and 12–14 kHz (Fig. 2.13) were slightly increased by changing the source from compression driver to flow source. These were caused by the difference of source characteristics, *i.e.* the amplitude of source was increased by the additional source generation where the jet flow impinged on the lip surface. This experimental evidence supports the results of a previous study that showed the potential role of the lip cavity as a sound source of /s/ in a human speaker (Shadle and Scully, 1995).

The pressure distribution measurements in the transverse and sagittal planes showed that the pressure amplitudes at angles associated with the center of the lips (*i.e.*, the angles around 90° in both planes) increased up to 15 dB when the lips were present in the frequency range above 5 kHz. This increase was observed for all measurement conditions (*i.e.*, near-field and far-field with the acoustic source or with flow supply). Far-field pressure fields of a concentric rigid two-tube simplification of vowel /ɑ/ vocal tract with an infinite baffle (no lips) (Blandin *et al.*, 2016), showed that the amplitude at 90° was larger than those at 0° and 180° above 6.5 kHz, and the maximum difference of amplitudes between the center (90°) and side regions (0° and 180°) was approximately 10 dB in the frequency range from 2 to 10 kHz. The maximum difference in the amplitude for the replica of /s/ was 15 dB for this frequency range (2 to 10 kHz), and these results indicate that the lip horn plays a role to enhance the directivity at the center compared to the repica of /ɑ/. This enhanced directivity in front of the mouth was also observed in the previous

measurements of sibilant /s/ on human speakers (Monson *et al*., 2012) and might therefore be partly due to the lip horn. It is of interest to further investigate the effect of the lips on other sounds in order to investigate the role of the lip horn for the sound radiation. Another potential parameter for the future study is the geometry of the lip horn (*e.g.* parameters for the curvature and protrusion of the radiating surface).

Directivity measurements conducted with an eccentric two tube simplification of /ɑ/ without lips showed that complex directivity patterns, with changes in directivity within short frequency and angle intervals, occur in the frequency range above 6.5 kHz (Blandin, *et al.*, 2016). The influence of the lips on the transfer function of the /ɑ/ vocal tract was mainly observed above 5.5 kHz (Arnela *et al.*, 2016). In this study, complex pressure patterns were observed with and without lips for frequencies above 4 kHz (normalized pressure amplitude differences up to 10 dB), which is significantly lower than that found in the vocal tract of /ɑ/. This suggests that the higher order modes of sibilant /s/ affect the pressure pattern at lower frequencies than that observed for vowels. This is in agreement with the argument that the higher order modes affect the pressure pattern below the first cut-on frequency, and plane-wave theory holds for frequencies below 4 kHz for the vocal tract geometries of sibilant fricatives (Motoki, 2013). Further acoustic modeling of wave propagation through the replica's geometry is needed in order to further examine the effect of higher order modes in sibilant fricatives.

A similar tendency was observed between the near-field and far-field measurements, although the resolution in the far-field (15°) was lower than that in the near-field (2°). This indicates that more precise measurement of the directivity pattern compared to the far-field measurement was achieved by using the 3D positioning system in the near field of the vocal tract. This result encourages a future development of a setup enabling a precise measurement of the directivity pattern in the far-field in order to confirm the current findings using other replicas and eventually on human speakers to investigate the perceptual relevance of the current findings.

The comparison between the case with acoustic source and flow revealed a similar tendency in both the transverse and sagittal plane. In particular, similar pressure distributions were observed in the frequency range from 7 to 10 kHz, which coincides with the frequency range of the highest acoustic energy observed on replica's sibilant /s/ (Nozaki *et al.*, 2014). When the flow was supplied to the replica, the noise source was probably generated downstream from the sibilant groove, *i.e.* around the teeth and lips. Nevertheless, similar pressure distribution patterns were observed for the different sound source positions, *i.e.* at the inlet of the pharynx or downstream from the sibilant groove. This indicates that the pressure pattern outside the vocal tract of sibilant /s/ is less affected by the position of the sound source than by the vocal tract geometry. It is of interest to confirm this finding for other fricatives since usage of an acoustic source allows a detailed spatial measurement of the acoustic pattern inside and outside of the vocal tract geometry which is difficult to achieve in the presence of flow.

2.3.2 Simplified Replica of /s/ and /ʃ/

The spectral shape of sound generated by the simplified replica with tongue model 1 was similar to the spectrum of /s/ in word context up to 12 kHz (Fig. 2.19). Moreover, the spectral shape of the sound generated by the replica with tongue model 1 matched with the subject's sustained /s/ up to 15 kHz including first and second characteristic peaks (4.8 and 9 kHz in Fig. 2.22). This indicates that the simplified replica reproduced the essential phenomena occurring in the vocal tract of sibilant /s/ considering the three-dimensional flow configuration observed in the realistic replica. In this simplified replica, the geometry of constriction, tongue tip, upper and lower teeth was mainly considered for the simplification. In the previous simplified models, the geometry was varied from the simple tube (Stevens, 1971) to 32-cross-section model (Motoki, 2013). Although in the model only with constriction (Stevens, 1971) or constriction and obstacle (Shadle, 1971) produce the first characteristic peak of sibilant /s/ around 4 to 5 kHz, detailed spectral shapes of the sound produced by the model were different from the sound pronounced by the speakers. This indicates that the proposed simplified geometry has better ability to produce the flow and acoustic fields in the actual vocal tract of /s/. By considering the 4 sections from throat to lips (Howe and McGowan, 2005), the spectral shape became much closer to those of the sound produced by the speakers. Meanwhile, the second characteristic peak around 8 to 9 kHz was not observed in Howe's model. With the model proposed by Toda and Maeda (2006) which consists of the back cavity, constriction, and front cavity, the second characteristic peak was predicted at 7 kHz. These results indicate that the dimensions of the front cavity (*i.e.* the cavity downstream from the constriction) and source position are key factors to produce the frequency characteristics of the sibilant /s/. The effects of the source position in the simplified geometry were investigated by the multimodal modeling in the next chapter (Chapter 3).

The spectral analysis from Fig. 2.21 to Fig. 2.23 showed that the vocal tract replica with tongue model 1 at the position L_{CA} = 4 mm and tongue model 4 at the position L_{CA} = 6 mm reproduced the frequency characteristics (*i.e.*, main peak frequency, spectral mean, and spectral shape) of the subjects /s/ and /ʃ/, respectively. These geometries were consistent with the dimensions estimated in the subjects CT images. Meanwhile, the peak frequency or spectral mean of the subject's /s/ and /ʃ/ were also reproduced by changing the position of tongue model 2 or 3. However, the spectral shapes of tongue models 2 and 3 were slightly different from those of the subject's /s/ and /ʃ/. Therefore, these results indicate that the dimensions of the replica with tongue model 1 at L_{CA} = 4 mm and tongue model 4 at L_{CA} = 6 mm are physiologically reasonable and represent the geometric features of the vocal tract for this subject pronouncing /s/ and /ʃ/, respectively.

By examining the calculated spectral parameters in Fig. 2.21, the effect of geometrical factors on the generated sound can be summarized as follows. The position of the tongue model (L_{CA}) that corresponds to the constriction position, length, and the sublingual space changed the main peak frequency. Changes in the amount of sublingual space, which are represented in presence or absence of the tongue tip, changed the peak frequency further. The constriction width slightly affected the peak frequency, though the changes were

smaller than those of the constriction position and sublingual space. In contrast, the constriction width mainly changed the OASPL. Although the tongue position L_{CA} = 4 to 6 and presence of the tongue tip decreased the OASPL as well, the decreased levels were smaller than those of the constriction width and the subject's /s/ to /ʃ/. The spectral mean was attenuated by the combination of the tongue position, sublingual space, and constriction width.

When the position of the tongue model was changed, the decreased ratio of the peak frequency was consistent with the resonance frequency estimated with the constriction position (L_f) in Eq. (2.1). This indicates that the channel between the constriction and outlet can be considered as a main resonator, and the position of the constriction is correlated with the resonance frequency of the main peak. In addition, the presence or absence of the tongue tip changed the volume of the main resonator and the resonant frequency as well. These results agree with that of the model proposed by Shadle (1985). The reason for the changes in OASPL can be speculated by considering the jet flow generated from the constriction. The decrease of mean velocity at the constriction decreases the amplitude of the sound source occurring around the teeth (Nozaki *et al.*, 2012). The widened constriction width decreased the mean velocity and hence decreased the OASPL. When the position of the tongue shifted backward, the magnitude of the velocity fluctuation, as well as the sound source around the teeth, probably decreased because of the larger L_{CA} (Van Hirtum *et al.*, 2009).

Consequently, the simplified replica reproduced the change in main peak frequency by changing the tongue position L_{CA} = 4 to 6 and removing the tongue tip, whereas the replica reproduced the change in OASPL by widening the constriction width. In future work, based on the effect of geometrical factors observed in this experiment, we can estimate the tongue shapes and positions for the different subjects' sustained /s/ and /ʃ/ through employing a mechanical experiment that includes the measurement of dimensions of the subject's lips and teeth. In addition, underlying physical mechanisms in the difference between /s/ and /ʃ/ are investigated by applying the aeroacoustic simulation on this geometry (Chapter 4).

2.4 Summary

With the realistic replica, acoustic pressure distribution patterns were measured on a vocal tract replica of sibilant /s/ with and without lips. It was found that complex pressure patterns with differences in amplitude of approximately 10 dB occur with and without lips for frequencies above 4 kHz. The lip horn enhances the pressure amplitude up to 15 dB at the center of the lips in both transverse and sagittal plane in the frequency range above 5 kHz. These tendencies were observed in the near-field and far-field measurements with the acoustic source, and in the measurements with flow supply. The comparison between the near-field and far-field measurements showed that more precise directivity pattern can be achieved by the near-field measurement compared to the far-field measurement. The comparison between the acoustic source and flow source showed that the pressure distribution pattern is affected by the vocal tract geometry rather than by the source characteristics. The presented experimental results motivate further studies involving spatially detailed directivity pattern measurements for different phonemes using the vocal tract replicas in combination with an acoustic source in order to further study the effect of the lip horn as well as to study the acoustic pressure patterns for different phoneme geometries. Furthermore, the perceptual relevance of these findings needs to be further investigated.

With the simplified replica, the effects of tongue position as well as the tongue shapes on the acoustic properties were assessed. The simplified replica reproduced the change in main peak frequency of /s/ and /ʃ/ by changing the tongue position L_{CA} = 4 to 6 and removing the tongue tip, whereas the model reproduced the change in OASPL of /s/ and /ʃ/ by widening the constriction width. These geometries were consistent with the dimensions estimated in the subjects CT images, indicating that the dimensions of the simplified replica with tongue model 1 at L_{CA} = 4 mm and tongue model 4 at L_{CA} = 6 mm are physiologically reasonable and represent the geometric features of the vocal tract pronouncing /s/ and /ʃ/, respectively. In future work, based on the effect of geometrical factors observed in this experiment, the tongue shapes and positions for the different subjects' sustained /s/ and /ʃ/ can be predicted through employing the mechanical experiments.

.

Chapter 3. Multimodal Acoustic Modeling and Analysis

In this chapter, we investigated the acoustic properties of proposed simplified vocal tract geometry by using the multimodal acoustic theory. The multimodal theory has been developed to compute the acoustic characteristics in higher frequency range where the assumption of plane wave (1D) propagation does not hold. Since the theoretical modeling can be applicable only with the simple cross-sections, *e.g*. rectangular, circular and elliptic sections, the vocal tract geometry has to be simplified to a concatenation of waveguides. Therefore, results in representing cross-section of the vocal tract are limited in accuracy. However, this modeling can represent the effects of the transverse dimensions of the vocal tract at higher frequencies where the most of acoustic energy of sibilant fricatives is generated. In addition, this method has an advantage of the fast computation compared to the numerical simulation, *e.g.* finite element method (FEM) for acoustics. With the multimodal modeling, the effects of source position on the far-field sound spectra as well as on the pressure distribution inside the vocal tract were examined. The established methodology is expected to be used as a fast prediction for the acoustic characteristics of the sibilant fricative production.

3.1 Method

3.1.1 Multimodal Theory

Multimodal theory has been developed and implemented by several researchers, *e.g.* Kergomard *et al*., (1989), Pagneux, *et al*., (1996), and Amir, *et al*., (1997). Vocal tract geometry is simplified as the concatenation of waveguides with constant cross-sections. In this chapter, *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 $\boldsymbol{v}(x,y,z)$ are defined as

$$p(x,y,z) = j\omega\rho\phi(x,y,z), \tag{3.1}$$

$$\boldsymbol{v}(x,y,z) = -\nabla\phi(x,y,z), \tag{3.2}$$

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

$$\nabla^2\phi(x,y,z) + k^2\phi(x,y,z) = 0, \tag{3.3}$$

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 $\psi_{mn}(x,y)$ weighted by forward and backward propagation amplitudes a_{mn}, b_{mn} as:

$$\phi(x,y,z) = \sum_{m,n=0}^{\infty} \psi_{mn}(x,y)\left\{a_{mn}\exp\left(-\gamma_{z,mn}z\right) + b_{mn}\exp\left(\gamma_{z,mn}z\right)\right\}, \tag{3.4}$$

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 $\psi_{mn}(x,y)$ 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 (Pierce, 1989);

$$\psi_{mn}(x,y) = \frac{\cos\left(\frac{m\pi x}{L_x}\right)}{\sqrt{L_x\sigma_m}} \frac{\cos\left(\frac{n\pi y}{L_y}\right)}{\sqrt{L_y\sigma_n}} \tag{3.5}$$
$$x \in [0, L_x], \quad y \in [0, L_y]$$

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

$$\gamma_{z,mn} = \sqrt{\left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2 - k^2}. \tag{3.6}$$

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

$$f_{c,mn} = \frac{c}{2\pi}\sqrt{\left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2}. \tag{3.7}$$

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. (3.4) is truncated to a certain value depending on the frequency range of interest. From Eq. (3.1), (3.2), and (3.4), the sound pressure and particle velocity along the *z*-axis v_z are calculated as

$$p(x,y,z) \approx \boldsymbol{\psi}^T(x,y)\{\boldsymbol{D}(-z)\boldsymbol{a} + \boldsymbol{D}(z)\boldsymbol{b}\}, \tag{3.8}$$

$$\begin{aligned} v_z(x,y,z) &= -\frac{\partial\phi(x,y,z)}{\partial z} \\ &\approx \boldsymbol{\psi}^T(x,y)\boldsymbol{Z}_C^{-1}\{\boldsymbol{D}(-z)\boldsymbol{a} - \boldsymbol{D}(z)\boldsymbol{b}\}, \end{aligned} \tag{3.9}$$

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

$$\boldsymbol{P} = \boldsymbol{D}(-z)\boldsymbol{a} + \boldsymbol{D}(z)\boldsymbol{b}, \tag{3.10}$$

$$\boldsymbol{V} = \boldsymbol{Z}_C^{-1}\{\boldsymbol{D}(-z)\boldsymbol{a} - \boldsymbol{D}(z)\boldsymbol{b}\}. \tag{3.11}$$

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

$$\boldsymbol{\Psi}_{i,i+1} = \frac{1}{S_i}\int_{S_i} \boldsymbol{\psi}_i(x,y)\boldsymbol{\psi}_{i+1}^T(x,y)dS \tag{3.12}$$

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

$$\boldsymbol{P}_i = \boldsymbol{\Psi}_{i,i+1}\boldsymbol{P}_{i+1}, \tag{3.13}$$

$$\boldsymbol{V}_{i+1} = \boldsymbol{\Psi}_{i,i+1}^{T}\boldsymbol{V}_i , \tag{3.14}$$

$$\boldsymbol{Z}_i = \boldsymbol{\Psi}_{i,i+1}\boldsymbol{Z}_{i+1}\boldsymbol{\Psi}_{i,i+1}^{\mathrm{T}}. \tag{3.15}$$

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 $\boldsymbol{Z}_{rad}$ at the outlet is obtained as

$$\begin{aligned}\boldsymbol{Z}_{rad} &= [Z_{mn,pq}] \\ &= \left[\frac{jk\rho c}{2\pi S}\int_S \int_S \psi_{mn}(x,y)\psi_{pq}(x',y')\frac{e^{-jkr}}{r}dS'dS\right] \\ r &= \sqrt{(x-x')^2+(y-y')^2}\end{aligned} \tag{3.16}$$

where S is the area of the outlet, and (x, y) and (x', y') are coordinates of points on S (Muehleisen, 1996). The relationship between the modal pressure, modal velocity, and radiation impedance at the outlet is written as

$$\boldsymbol{P}_{\boldsymbol{out}} = \boldsymbol{Z}_{rad}\boldsymbol{V}_{\boldsymbol{out}}. \tag{3.17}$$

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 (Pierce, 1989)

$$\begin{aligned}p(x,y,z) &= \frac{jk\rho c}{2\pi S}\int_S \boldsymbol{V}_{\boldsymbol{out}}\cdot\boldsymbol{\psi}_{\boldsymbol{out}}(x',y')\frac{e^{jkh}}{h}dS', \\ h &= \sqrt{(x-x')^2+(y-y')^2+(z-z')^2}.\end{aligned} \tag{3.18}$$

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).

3.1.2 Vocal Tract Geometry

A simplified vocal tract geometry consists of a concatenation of six sections, each with uniform cross-sectional area, and a rectangular cross-sectional shape (related to simplified geometries in section 2.1.2). The geometry of the vocal tract pronouncing /s/ obtained by CT scan (Nozaki *et al.*, 2014) is depicted in Fig. 3.1 (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. 3.2 (b) and dimensions are given in Table 3.1. 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^3s^{-1} reproduces the main spectral features of sound /s/, and the maximum discrepancy between the spectra was less than 9 dB in the frequency range 0.5 to 15 kHz.

In addition to the simplified geometry shown in Fig. 3.1, another geometry was used to assess the effects of the tongue position on the frequency characteristics of generated sound. It was outlined in Chapter 2 that the distance between the tongue and teeth is an important parameter suggested to change the sibilant fricative sound from /s/ to /ʃ/. Therefore, the distance between the tongue and teeth L_{CA} was varied from 0 to 9 mm in the simplified vocal tract geometry. Concretely, the geometry consists of a concatenation of nine rectangular uniform sections. The geometry is illustrated in Fig. 3.2 and dimensions are given in Table 3.2. The total length from inlet to outlet yields 167 mm. When L_{CA} = 0, the longitudinal length *Lz* of each section from throat to upper teeth (section 1-8) is the same as for the vocal tract shown in Fig. 3.1, whereas the longitudinal length *Lz* of lip cavity (section 9) is 5 mm shorter than the geometry of Fig. 3.1 to mimic a lip opening of both /s/ and /ʃ/. In addition, transverse length *Lx* of each section except section 3 is longer than that of Fig. 3.1 so that the cutoff frequencies of Fig. 3.2 are smaller than those of Fig. 3.1.

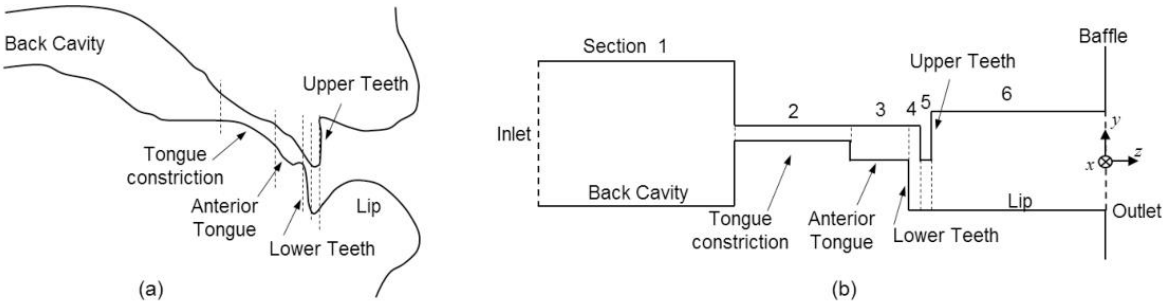


Fig. 3.1 (a) Mid-sagittal section of a vocal tract of male Japanese subject pronouncing /s/, (b) Simplified rectangular vocal tract geometry of sibilant /s/.

Table 3.1 Dimensions (mm) of each section of the sibilant /s/ vocal tract geometry.

Section	1	2	3	4	5	6
Lx	16.8	8	23	21	21	23.5
Ly	12.5	1.25	3	7.25	4.25	8.5
Lz	140	10	5	1	1	14

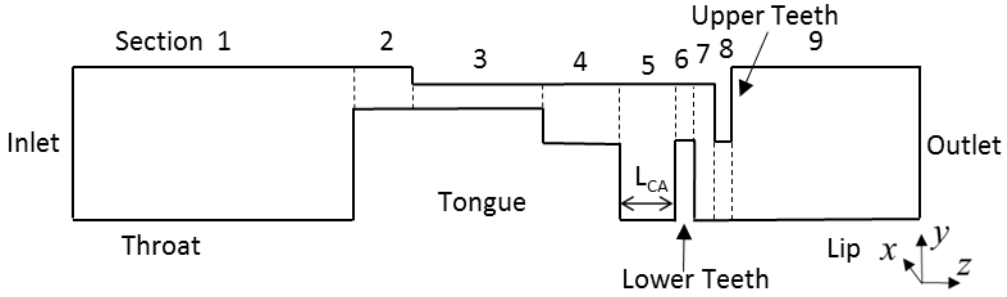


Fig. 3.2 Simplified rectangular vocal tract geometry of sibilant /s/ and /ʃ/.

Table 3.2 Dimensions (mm) of each section in the sibilant /s/ - /ʃ/ vocal tract geometry.

Section	1	2	3	4	5	6	7	8	9
Lx	25	25	8	25	25	25	25	25	25
Ly	8	2.3	1.3	3	7	3	7	4	8
Lz	140 - L_{CA}	L_{CA}	10 - L_{CA}	4	L_{CA}	1	1	1	10

3.1.3 Source Position

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

$$\boldsymbol{V}_{in} = \int_{\Omega_0} v_0 \boldsymbol{\psi}_{in}(x,y) dS, \tag{3.19}$$

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 mm s^{-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 3.3, 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.

When the turbulent flow is generated in the vocal tract, monopole, dipole, and quadrupole sound sources of flow fluctuation are assumed to be produced (Stevens, 1971; Shadle, 1985; Howe and McGowan, 2005). 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. 3.1 (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

$$\boldsymbol{P}^+ - \boldsymbol{P}^- = 0 \tag{3.20}$$

$$\boldsymbol{V}^+ - \boldsymbol{V}^- = \boldsymbol{Q} \tag{3.21}$$

$$\boldsymbol{Q} = Q\boldsymbol{\psi}(x, y) \tag{3.22}$$

where + and - represent the variables downstream and upstream of the source section, and Q is the volume flow rate supplied at the inlet (Amir *et al.*, 1997). We imposed a fluctuating volume flow rate Q = 190 mm^3s^{-1} as a monopole sound source to reproduce the SPL of the flow source (Howe and McGowan, 2005).

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 (Howe and McGowan, 2005; McGowan and Howe, 2007) 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. 3.3. 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 $\boldsymbol{Z}_C$ was calculated with the same area of inlet used for the acoustic driver.

Table 3.3 Mode (*m*,*n*) and corresponding 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)
fc (kHz)	0	7.3	14.6	20.2	21.5	21.9

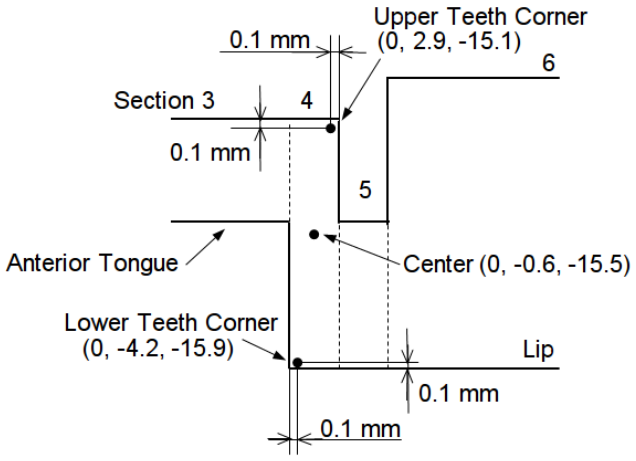


Fig. 3.3 Position of the sound source in section 4.

3.2 Experimental Validation

3.2.1 Experimental Method

For physical model 1, a replica of the simplified 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. 3.4.

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 $-10 \text{ mm} \leq x \leq 10 \text{ mm}$, $-13.8 \text{ mm} \leq z \leq 0.2 \text{ mm}$ for the horizontal plane and in $-3 \text{ mm} \leq y \leq 3 \text{ mm}$, $-13.8 \text{ mm} \leq z \leq 0.2 \text{ mm}$ 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 duration of 20 s. The acoustic pressure p at the microphone probe was recorded during each sweep signal using a data 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;

$$G_{lip}(f) = 20\log_{10}(p(0,0,-1.8)/p(0,0,-13.8)). \quad (3.23)$$

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).

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. 3.1 (b) constructed using rapid plaster prototyping was used in this flow experiment. The geometry shown in Fig. 3.2 was constructed by acrylic boards. 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 $\text{cm}^3\,\text{s}^{-1}$ (Fujiso *et al.*, 2015). A rectangular baffle (350 × 350 mm) was attached at the outlet of the replica to be consistent with the infinite baffle of the

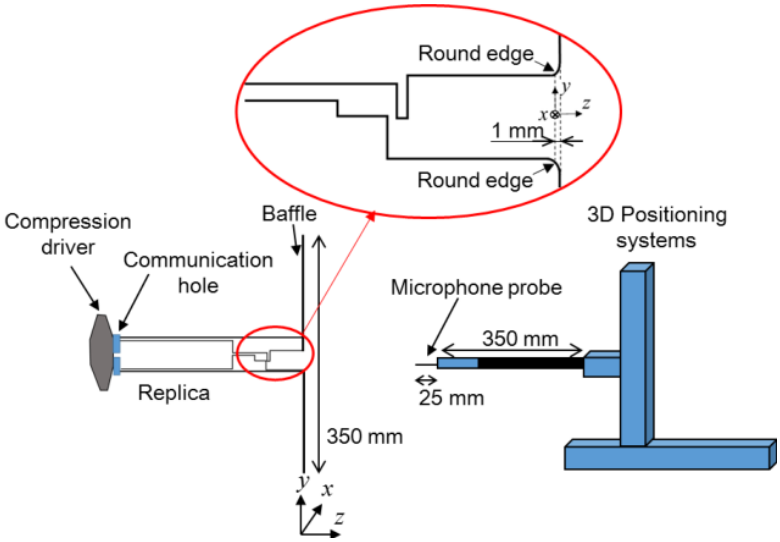


Fig. 3.4 Schematics of experimental setup.

theoretical 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.

3.2.2 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. 3.5. Note that as in Fig. 3.1 (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 second-order mode (7.3 kHz in Table 3.3). This indicates that the cutoff frequency is slightly changed from the estimated value due to the complex geometry, 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.

Figure 3.6 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. 3.5 (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. 3.5) and vertical (Fig. 3.6) 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.

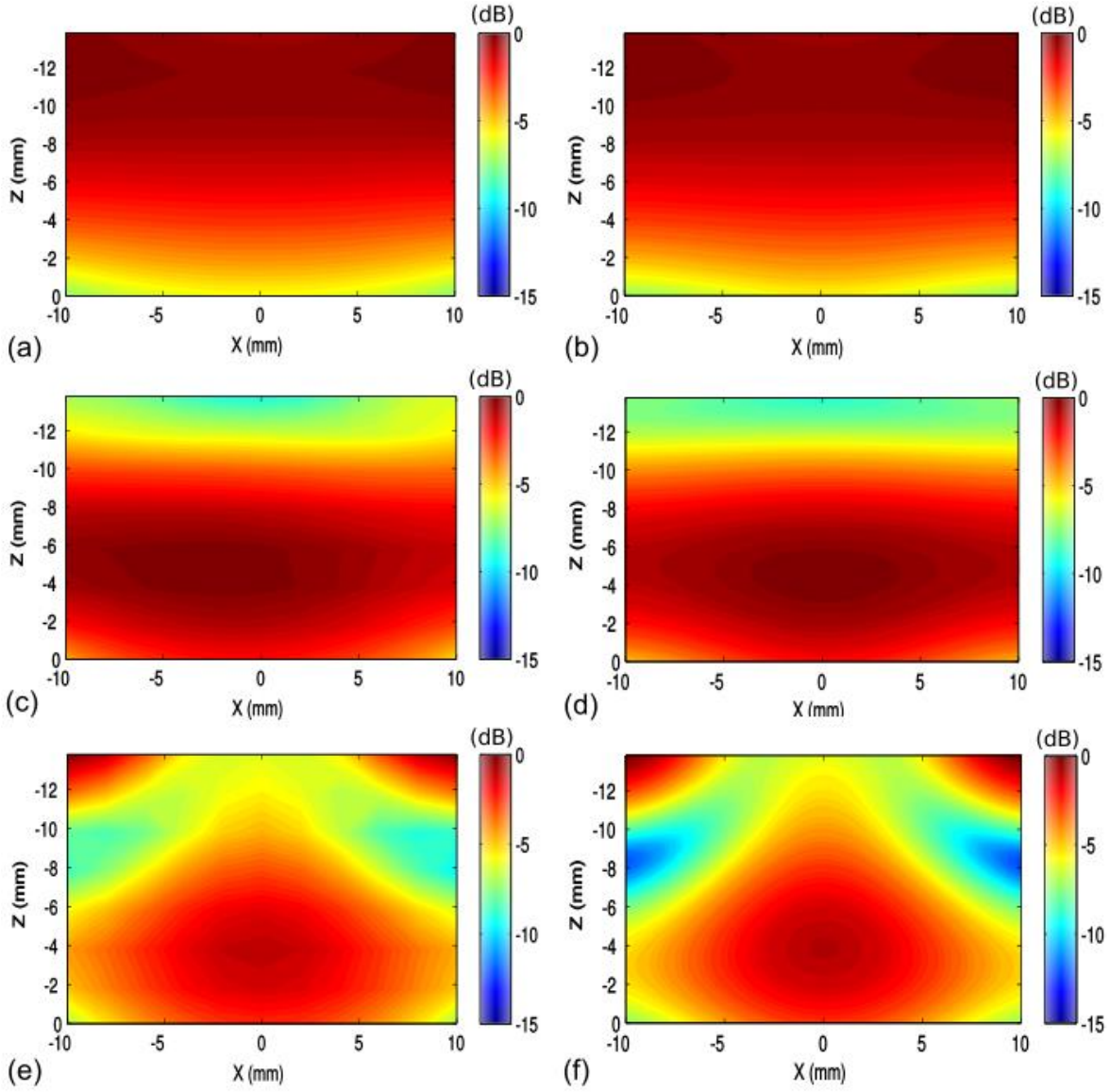


Fig. 3.5 Measured (Physical Model 1, left column) and Modeled (Computational Model 1, right column) pressure distribution along the horizontal plane at y = 0 in lip section (section 6 in Fig. 3.1) for frequency 5050 Hz (a–b), 10050 Hz (c–d), and 13050 Hz (e–f). Amplitudes were normalized by the maximal value on the plane. The z-axis corresponds to the main propagation direction with the outlet at z = 0.

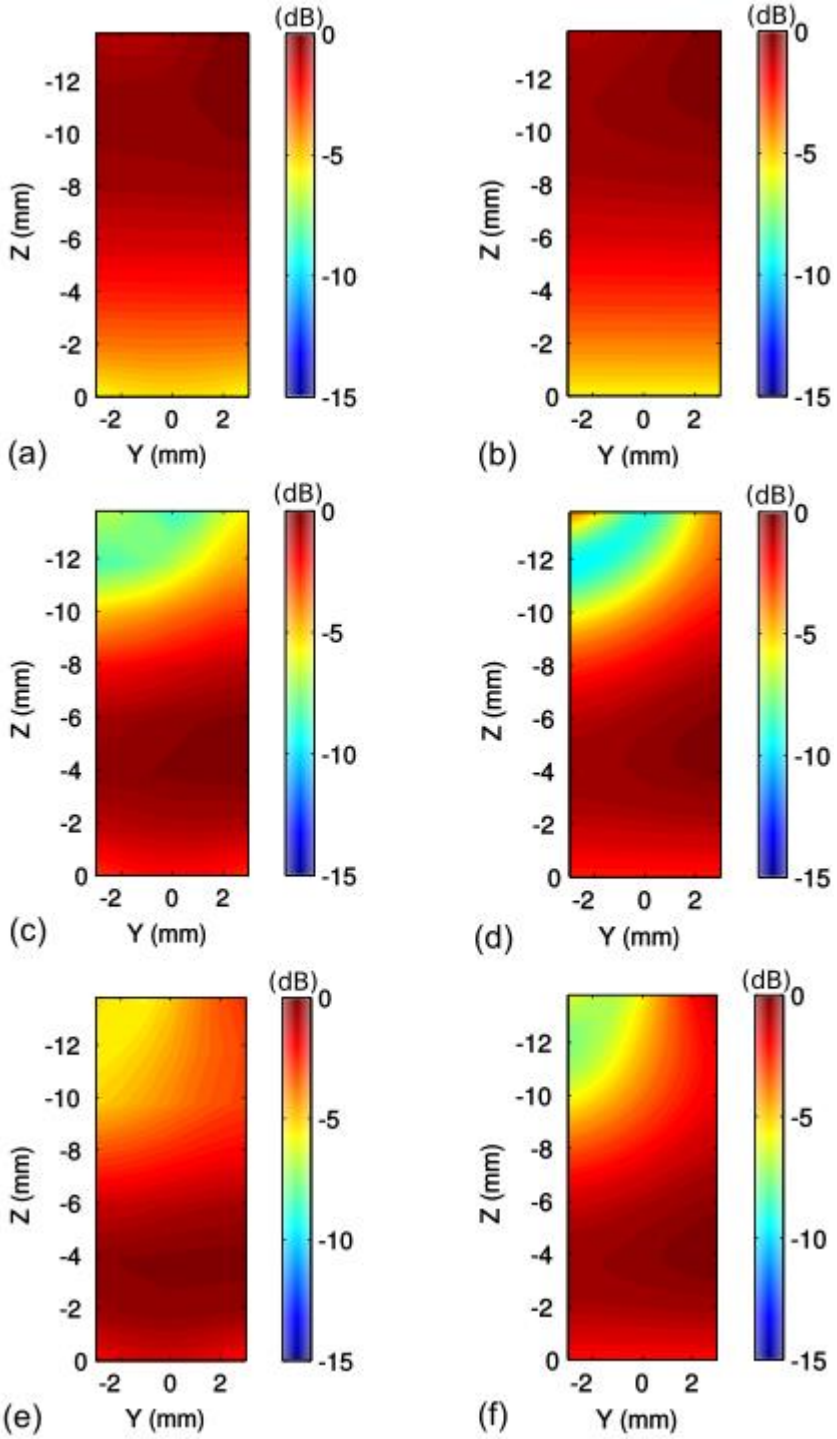


Fig. 3.6 Measured (Physical Model 1, left column) and modeled (Computational Model 1, right column) pressure distribution along the vertical plane at x = 0 in lip section (section 6 in Fig. 3.1) for frequency 5050 Hz (a–b), at 10050 Hz (c–d), and 13050 Hz (e–f). Amplitudes were normalized by the maximal value on the plane. The z-axis corresponds to the main propagation direction with the outlet at z = 0.

3.2.3 Flow Source

When the 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 to 6 in Fig. 3.1 (b)) (Stevens, 1971; Shadle, 1985; Howe and McGowan, 2005). The sound spectrum measured and modeled at 30 cm from the outlet along the centerline of the replica is plotted in Fig. 3.7. The replica produced a sound similar to the subject's sibilant /s/, which is characterized as broadband noise above 4 kHz with the first characteristic peak at 4 kHz and overall spectral peak at 8 kHz. General tendencies of the measured sound spectra were captured by applying the multimodal model when the sound source is located near the upper teeth in section 4 (y = 2.9, z = -15.1). However, the amplitude in the frequency range 4.5 to 12 kHz was lower than that measured in the flow experiment (approximately 5-10 dB). This might be due to the difference of the source characteristics between the flow source and modeled monopole source. Therefore, further agreement might be achieved by accounting for dipole or quadrupole source distributions in the 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.

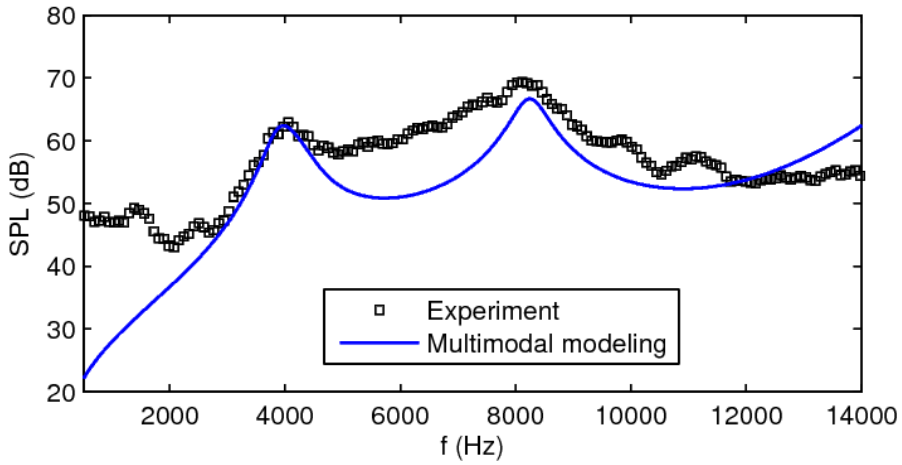


Fig. 3.7 Experimentally measured (Physical Model 2, rectangle) and theoretically modeled spectra (Computational Model 2: blue line) for the source near the upper teeth corner (Y = 2.9, Z = − 15.1), observed at 30 cm from the outlet along the centerline (z-axis). The experiment was conducted by supplying air to the mechanical replica. SPL is based on the reference level 20×10^{-6} Pa.

3.3 Detailed Results

3.3.1 Effects of Mode Number

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. 3.8. 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 transfer function obtained using the multimodal model approach matches with the measured transfer function 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 (Motoki, 2013).

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. 3.9. 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). The second characteristic peak of plane wave model matched with the peak measured at 11 kHz in the experiment. In order to assess the cause of this match, the spectra predicted by two, three, and four modes in the theoretical model were compared with experimentally measured spectrum in Fig. 3.10. The modeling with four modes (blue solid line) is corresponding to the multimodal model. By increasing the number of modes in the vocal tract, the second characteristic peak frequency was shifted to lower frequencies from 11 kHz to 8 kHz. This result indicates that the overall peak at 8 kHz was captured by shifting the peak at 11 kHz predicted by the plane wave model. From these results, we speculate that the amplitude of plane wave model at 11 kHz was accidentally matched with the peak measured in the flow experiment.

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 (Howe and McGowan, 2005). In addition, the second characteristic peak was predicted by the multimodal model for this source location. 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.

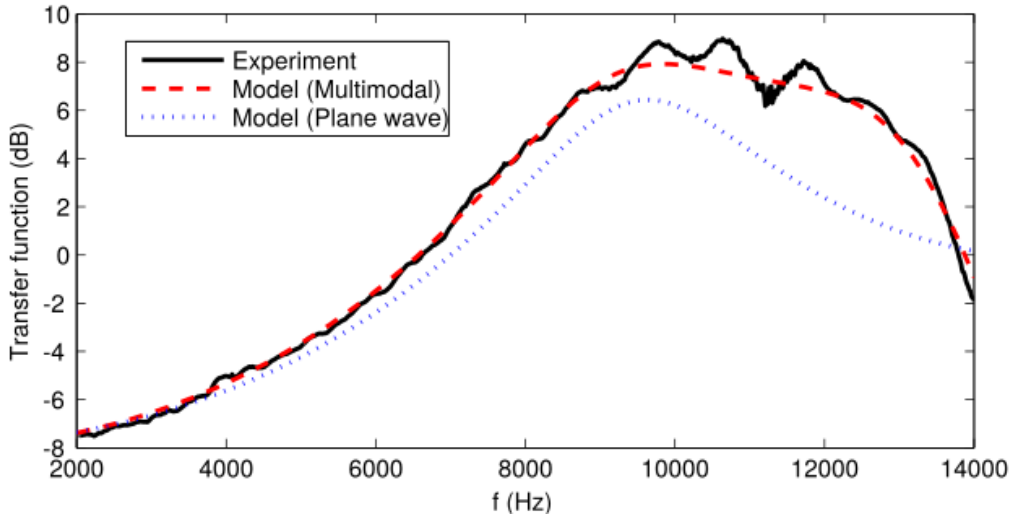


Fig. 3.8 Experimentally measured (Physical Model 1, black line) and theoretically modeled (Computational Model 1, plane wave: blue dot, multimodal: red dash) pressure-pressure transfer function G lip as a function of frequency between positions z = -13.8 mm (near teeth) and z = − 1.8 mm (near outlet). Experiment was conducted with the compression driver at the inlet of the mechanical replica.

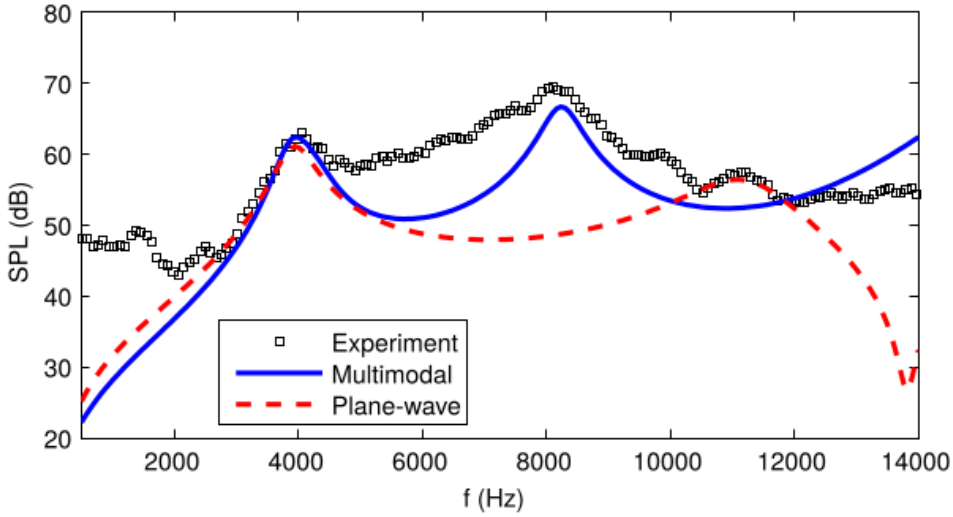


Fig. 3.9 Experimentally measured (Physical Model 2, rectangle) and theoretically modeled spectra (Computational Model 2, plane wave mode: red dash, multimodal: blue line) for the source near the upper teeth corner (Y = 2.9, Z = − 15.1), observed at 30 cm from the outlet along the centerline (z-axis).

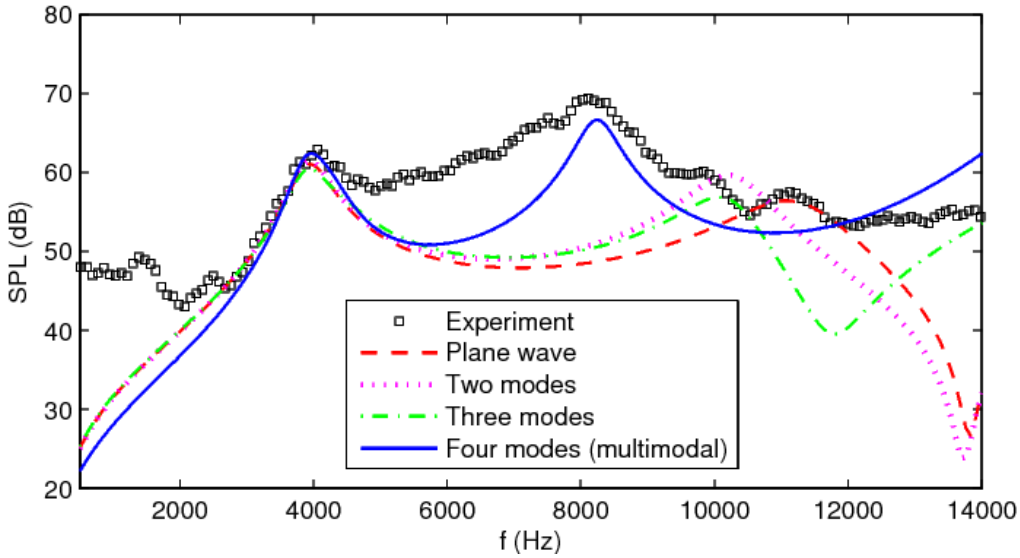


Fig. 3.10 Experimentally measured and theoretically modeled spectra (plane wave mode: red dash, two modes; pink dot, three modes: green dot-dash, multimodal: blue line) for the source near the upper teeth corner (Y = 2.9, Z = − 15.1), observed at 30 cm from the outlet.

3.3.2 Effects of Source Position

To assess the effect of the source position in the multimodal model (Eq. (3.21–3.23)), far-field pressure was calculated with Eq. (3.18) 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. 3.11. 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). The 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 center were assessed in the model. In particular, positions near the wall of teeth (corner positions) are considered as illustrated in Fig. 3.3. 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. 3.12. 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.

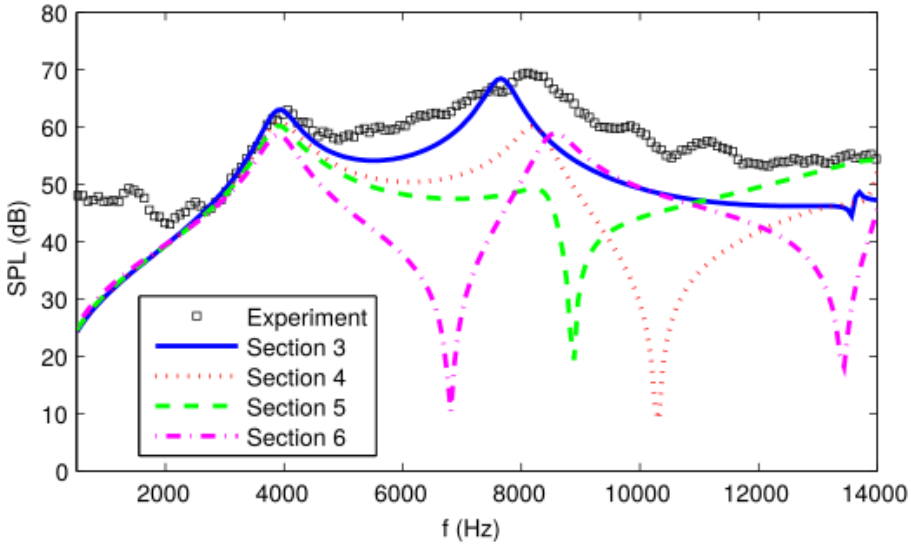


Fig. 3.11 Experimentally measured spectrum (Physical Model 2, rectangle) and theoretically predicted spectra (Computational Model 2) with multimodal model and monopole at the center of sections 3-6 (the center of section 3: blue line, section 4: red dot, section 5: green dash, and section 6: pink dot-dash) at 30 cm from the outlet along the source 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.

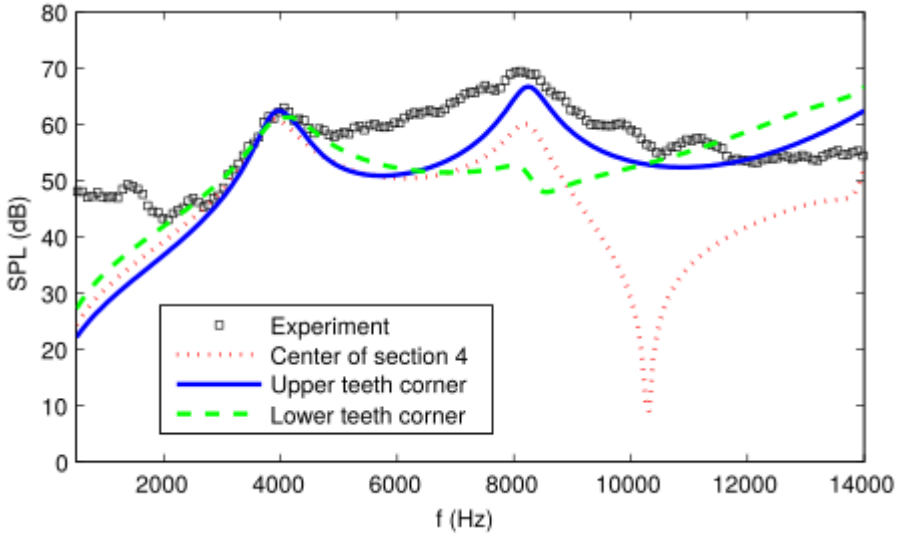


Fig. 3.12 Experimentally measured spectrum (Physical Model 2, rectangle) and theoretically predicted spectra (Computational Model 2) with multimodal model at 30 cm from the outlet along the centerline (z-axis). A monopole source was located within section 4: near the upper teeth corner (Y = 2.9, Z = − 15.1, blue line); at the center (Y = 0, Z = − 15.5, red dot); and near the lower teeth corner (Y = − 4.15, Z = − 15.9).

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. 3.13 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 Fig. 3.9 to 3.12) appeared within the cavity between the constriction and the upper teeth. 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). 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. 3.14 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 to 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 node and antinode in the pressure distribution inside the vocal tract for the sibilant /s/.

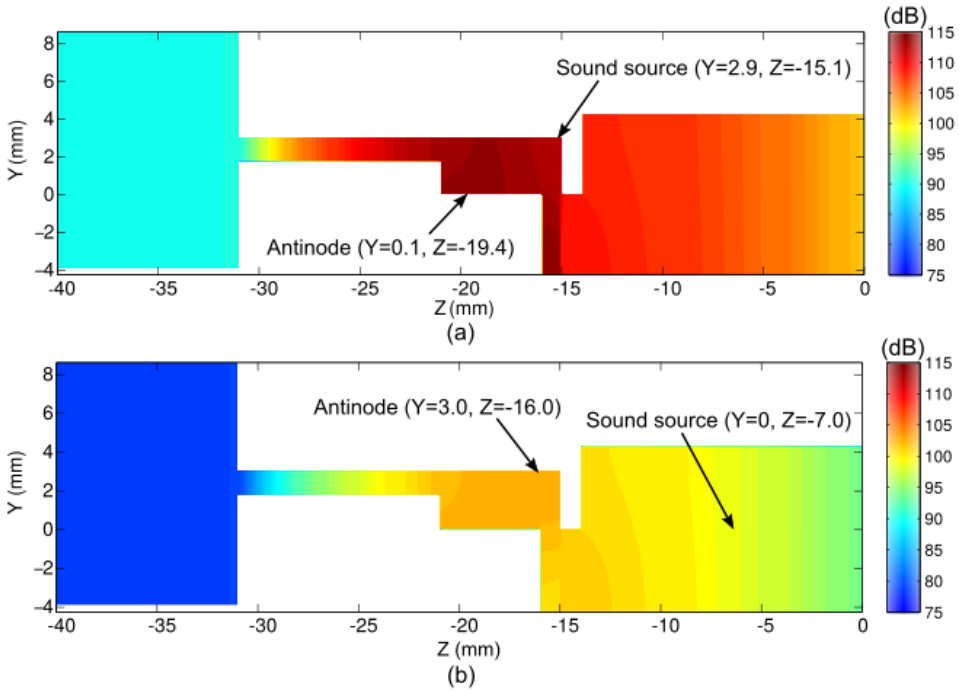


Fig. 3.13 Pressure distribution predicted by the multimodal model (Computational Model 2) along the vertical center plane (x = 0) of a portion of the vocal tract geometry for 4 kHz. The monopole source was positioned near the upper teeth corner in section 4 (a), and at the center of section 6 (b).

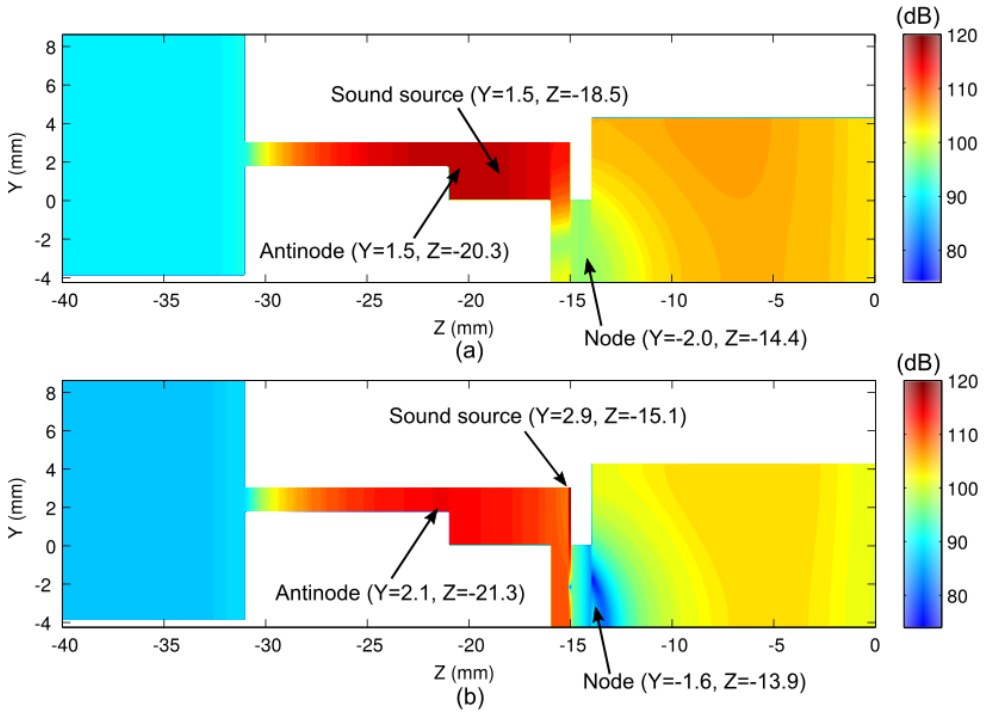


Fig. 3.14 Pressure distribution predicted by the multimodal model (Computational Model 2) along the vertical center plane (X = 0) for the frequency of the second peak 7.6 kHz (a) and 8.2 kHz (b). The monopole source was positioned at the center of section 3 for (a) and near the upper teeth corner for (b).

3.3.3 Effects of Tongue Position

The modeled spectra for the sound source positioned at section 4 in simplified geometry of /s/ and /ʃ/ (Fig. 3.2) and measured spectra with flow are plotted in Fig. 3.15. The length between tongue and teeth L_{CA} was changed from L_{CA} = 0 to 2 mm and 7 mm. As the length L_{CA} increased from 0 to 7 mm, the frequency of the first characteristic peak frequency of the modeled spectrum decreased from 4.7 to 3.4 kHz, and its amplitude increased from 68 up to 82 dB. In addition, the frequency and amplitude of the second characteristic peak decreased from 8.5 to 6.8 kHz and from 70 to 60 dB, respectively. The general tendencies of the measured spectra, which shifted toward the lower frequency ranges, were captured by the multimodal model. The observed frequency shift for the first characteristic peak as a function of L_{CA} corresponds to the peak shift observed on human speakers uttering sibilant fricatives /s/ (5-6 kHz) or /ʃ/ (2-3 kHz) as shown in Chapter 2. Meanwhile, the first peak of the measured spectrum when L_{CA} = 7 mm was broader and its frequency was lower than that of the modeled spectrum.

The modeled spectra for the sound source positioned at upper teeth corner in section 7 and measured spectra with flow are plotted in Fig. 3.16. By positioning the source at the upper teeth corner, the second characteristic peak at 9 kHz when L_{CA} = 0 was captured in the same way as the simplified geometry shown in Fig. 3.1 (b). Moreover, the first characteristic peak at 3.1 kHz when L_{CA} = 7 mm was more precisely captured by positioning the source at the upper teeth corner. The amplitude of the first peak was decreased from 82 to 78 dB by changing the source position from section 4 to 7.

Meanwhile, the second characteristic peak captured by the source at the center of section 4 at 6.9 kHz was shifted to lower frequency 5.8 kHz.

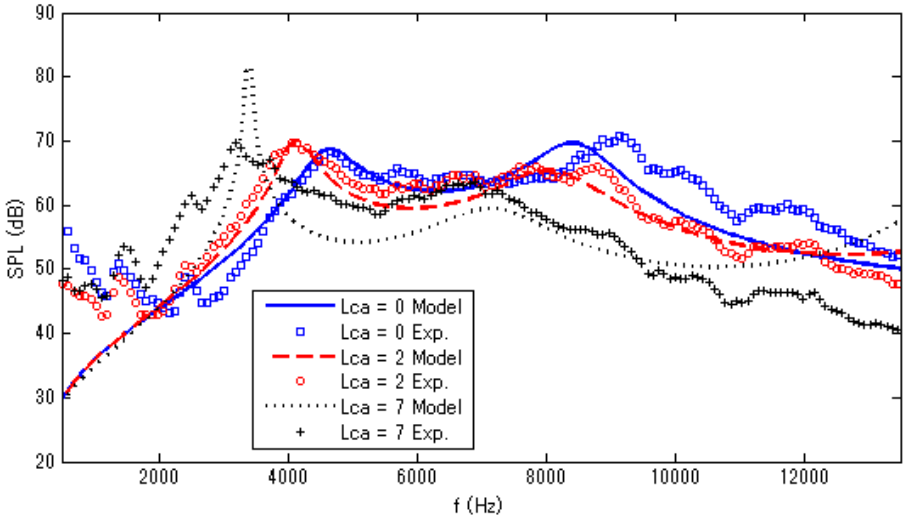


Fig. 3.15 Experimentally measured spectrum (Physical Model 2) and theoretically predicted spectra (Computational Model 2) with multimodal model for lower mouth cavity lengths L_{CA} = 0, 2, and 7 mm. The sound source in the model was positioned at the center of section 4.

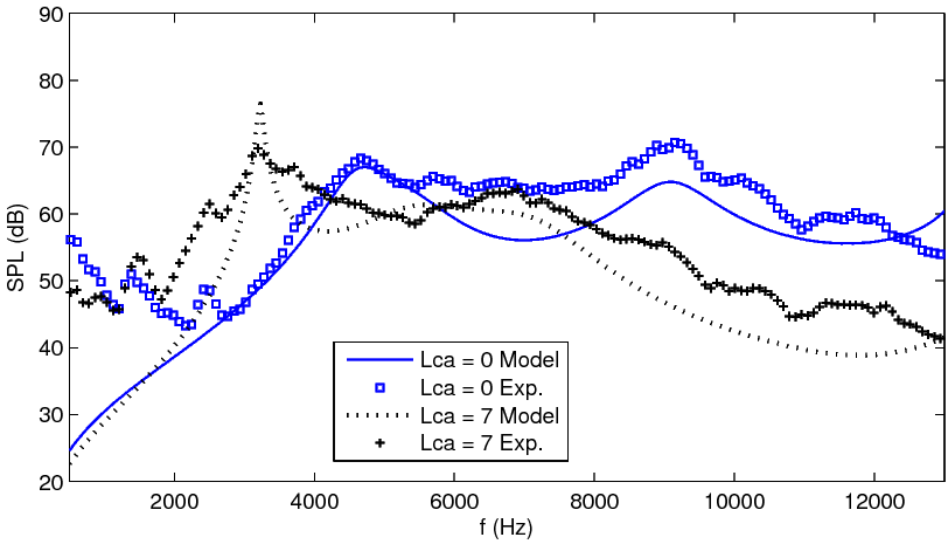


Fig. 3.16 Experimentally measured spectrum (Physical Model 2) and theoretically predicted spectra (Computational Model 2) with multimodal model when the lower mouth cavity lengths L_{CA} = 0, and 7 mm. The sound source in the model was positioned at the upper teeth corner of section 7.

Modeled first characteristic peak frequencies for the source within section 4 and 9 as a function of cavity length L_{CA} are compared with the measured peak frequency values in Fig. 3.17 (a). Modeled and measured frequencies decreased from 4.7 kHz to 3.4 kHz when L_{CA} increased from 0 to 7 mm. At this time, the peak frequency decreased with 65 Hz by changing the source position from section 4 to 9. For $L_{CA} \geq$ 7mm, the modeled peak frequency for the source within section 4 remains roughly constant at 3.4 kHz whereas the peak frequency for the source within section 9 decreased to 3.2 kHz. The error bar of measured values shows the standard deviation of 5 measurement trials, and modeled value for both source positions falls within the range of measured values.

The longitudinal position (*z*-direction) of the maximum value in the pressure distributions along the vertical center plane ($x = 0$) for the first peak frequency is plotted as a function of L_{CA} in Fig. 3.17 (b). The maximum position corresponds to the antinode of the first characteristic peak frequency as illustrated in Fig. 3.13. The position was shifted from z = -13.5 mm up to -20 mm when L_{CA} increased from 1 to 7 mm, and remains constant (z = -20 mm) for $7 \leq L_{CA} \leq 9$ mm for both source positions within section 4 and 9. In contrast, when L_{CA} = 0 mm, the maximum position for the source within section 4 was 3.5 mm backward from the position for the source within section 9. The position of the tongue tip in the vocal tract geometry (boundary between section 4 and 5) is also plotted in Fig. 3.17 (b). The position of tongue tip roughly matched with the position of maximum pressure positions for $1 \leq L_{CA} \leq 7$ mm.

The modeled pressure distributions on the vertical center plane with the source at the center of section 4 for L_{CA} = 1, 7 and 9 mm at 4.4, 3.4, and 3.4 kHz (first peak frequency), respectively, are shown in Fig. 3.18. The maximum pressure was observed at the bottom of the cavity between tongue and lower teeth for all tongue positions. For $L_{CA} \geq 7$ mm, the maximum position is maintained at the same position (z = 20 mm), though the tongue tip was shifted backward. As a result, the frequency of first characteristic peak, which corresponds to the maximum pressure position, remained the same for $7 \leq L_{CA} \leq 9$ mm.

In order to investigate the cause of the second characteristic peak when the L_{CA} = 7 mm, the modeled pressure distributions on the vertical center plane when the source was positioned at the center of section 4 and at the upper teeth corner are shown in Fig. 3.19. When the source was positioned at the center of section 4, the maximum amplitude appeared at the upper corner of lip cavity (z = -10 mm). Meanwhile, the acoustic node and antinode appeared in the space between the constriction and teeth.

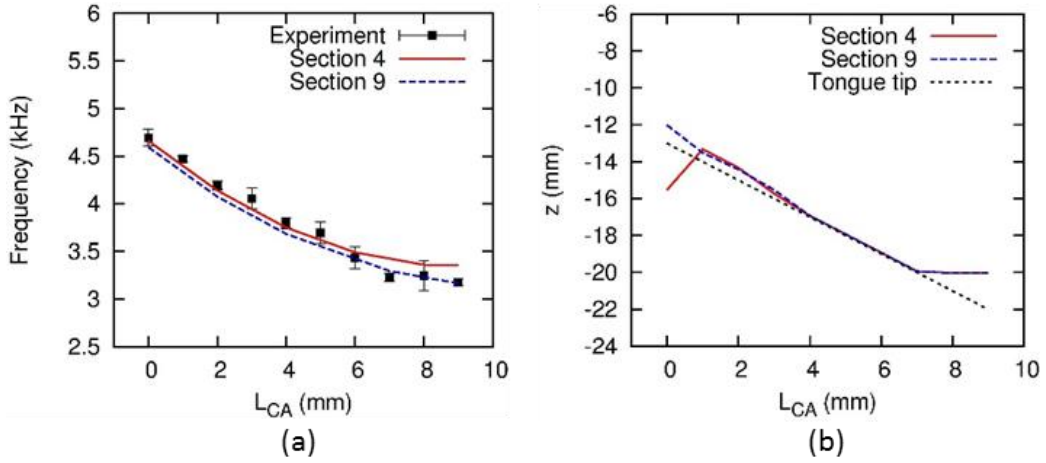


Fig. 3.17 (a) First characteristic peak frequency as a function of cavity length L_{CA} for modeled (source at section 4 and 9) and measured pressure spectra. (b) The longitudinal position (z-direction) of the maximum value in the pressure distributions along the vertical center plane (x = 0) for the first peak frequency. The position of the tongue tip (boundary of section between 4 and 5) is also plotted in (b).

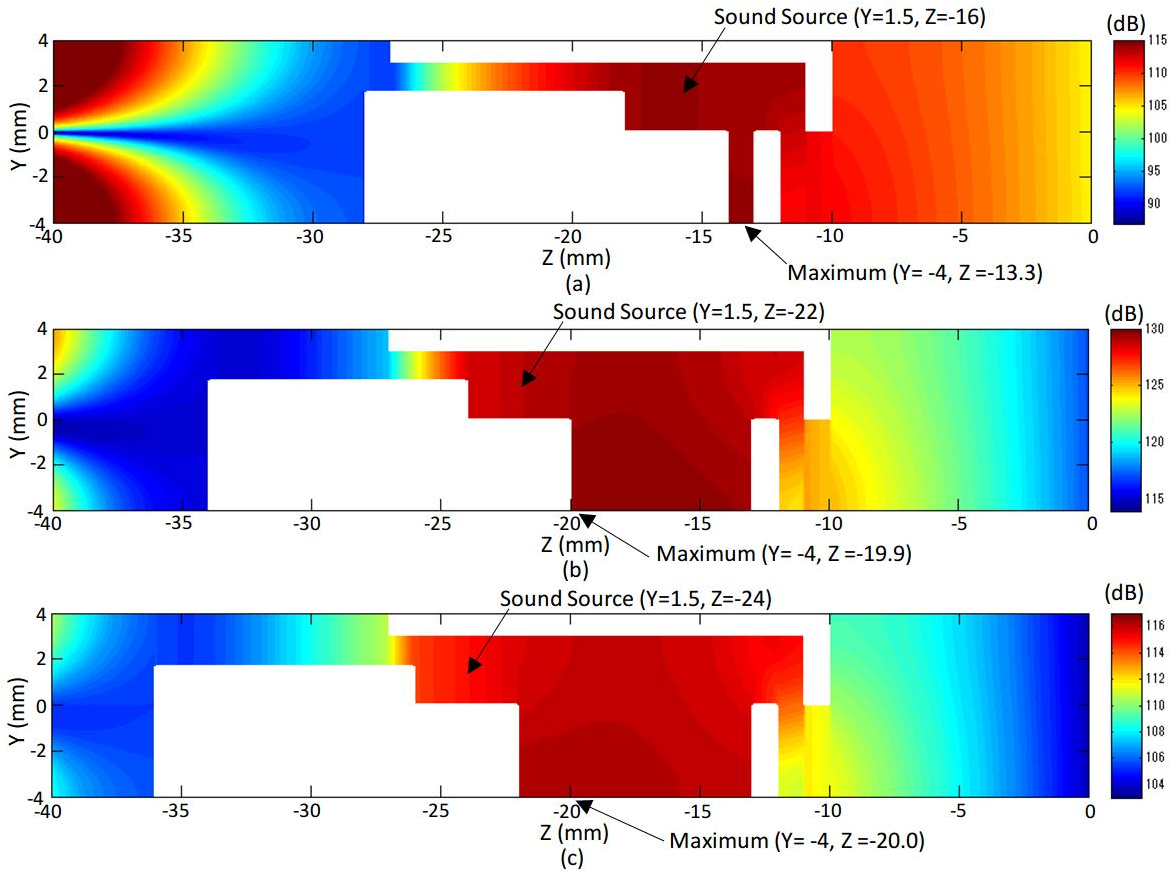


Fig. 3.18 Pressure distribution along vertical center plane (x = 0 mm) of the simplified vocal tract geometry for L_{CA} = 1 mm at 4.4 kHz (a), 7 mm at 3.4 kHz (b), and 9 mm at 3.4 kHz (c) for the source at the center of section 4.

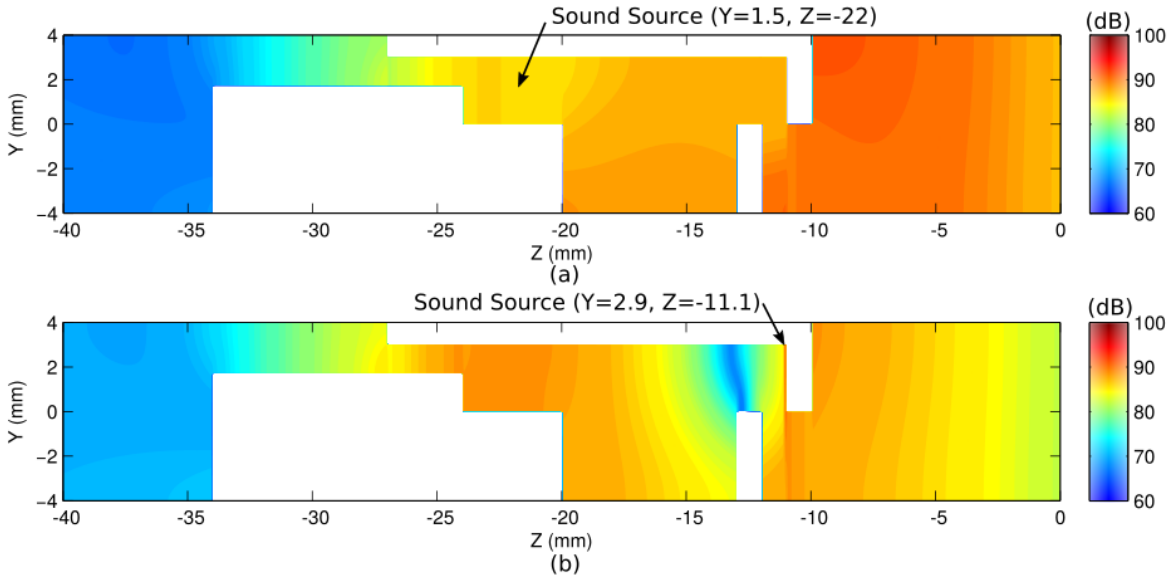


Fig. 3.19 Pressure distribution along vertical center plane (x = 0 mm) of the simplified vocal tract geometry for L_{CA} = 7 mm at 6.9 kHz when the source was positioned at the center of section 4 (a) and at the upper teeth corner of section 7 (b).

3.4 Discussion

Experimental validation in section 3.2 showed that the predicted and measured pressure distributions for the source at the inlet agreed well when the acoustic higher-order modes were taken into account. The pressure-pressure transfer function showed that the difference between the plane wave and the measured transfer function appeared above 4 kHz. Meanwhile, the transfer function obtained using the multimodal model approach matched with the measured transfer function up to 14 kHz. The difference between the plane wave and measured values appeared for frequencies lower than it did in the case of vowels (4.5 kHz) (Blandin, *et al.,* 2016). This indicates that the smaller cross-sectional distances of sibilants have a greater effect on the pressure distribution than those of vowels.

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). This peak was not observed in Motoki's model (Motoki, 2013). Moreover, general tendencies of the measured spectra including the second characteristic peak were 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. However, the amplitude in the frequency range 4.5 to 12 kHz was lower than the amplitude measured in the flow experiment (approximately 5-10 dB). This discrepancy over 5 dB is significant for listeners with normal hearing (Monson, *et al*., 2014). 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 source distributions, owing to the airflow velocity fluctuations (Lighthill, 1952). 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 (McGowan and Howe, 2007).

Note that the first characteristic peak was captured by the plane wave model in the same way as with the one-dimensional model (Howe and McGowan, 2005). However, the comparison with the flow experiment suggests that higher-order modes have to be taken into account to be able to capture the higher mode peak. In addition, the overall spectral peak observed at 8 kHz matches the spectra measured with European Portuguese speakers' /s/ (Jesus and Shadle, 2002) and Shadle's /s/ model (Shadle, 1985). 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, and indeed confirms previous findings in literature (Shadle, 1985; Howe and McGowan, 2005; McGowan and Howe, 2007).

For the frequency of the first characteristic 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. This finding is consistent with Shadle's simplified model (Shadle, 1985) that showed the dependence of the main resonance frequency on the position of the constriction.

For the frequency of the second characteristic 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 node and antinode in the pressure distribution inside the vocal tract as well as the amplitude and frequency of the peaks observed in the subject's /s/. The relationship between these nodes or antinodes and the peak frequencies will be the subject of further study.

The results of the second simplified geometry (Fig. 3.2) showed that the decreases of frequencies and amplitudes of the characteristic peak due to the tongue position were captured by the multimodal modeling. This suggests that the multimodal modeling is capable of assessing the effects of geometrical changes on the far-field sound spectrum of the sibilant fricatives. Dependence of the peak frequency and antinode position on the tongue position showed that the change of the antinode positions are correlated with the change of the first characteristic peak frequency (Fig. 3.17 (a)). This result indicates that the first characteristic frequency peak of sibilant fricatives is determined by the position of the antinode occurring between the tongue tip and teeth. This also suggests that the difference between /s/ and /ʃ/ occurs due to a shift of antinode position determined by the position of the tongue tip in the vocal tract.

For $L_{CA} \geq 7$ mm, the maximum position was maintained at the same position, although the tongue tip was shifted backward. As a result, the frequency of the first characteristic peak remained in the constant value for $7 \leq L_{CA} \leq 9$ mm. This indicates that the acoustic resonance occurs not only in the region between the tongue tip and outlet, but also in the cavity between the constriction and teeth for $L_{CA} \geq 7$ mm. In order to reduce the peak frequency, sections downstream from the constriction, *i.e.* longitudinal length L_z or spanwise length L_x at section 4-5, need to be increased.

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 node and antinode 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.

3.5 Summary

In this chapter, the 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. For the experimental validation, the acoustic driver and flow supply were applied to the inlet of the simplified geometrical approximation. 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 were 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. Moreover, the comparison with the flow experiment suggests that higher-order modes have to be taken into account to be able to capture the higher frequency peaks. 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 appeared 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 positions of the node and antinode were 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/.

Chapter 4. Computational Analysis of Aeroacoustic Fields

In this chapter, we investigated the relationship among the vocal tract geometry, flow configuration, and acoustic properties in the vocal tract. The aeroacoustic simulations were conducted in two ways: coupling method and direct method. Firstly, the frequency characteristics of the sound generated by the flow inside the simplified vocal tract were predicted by the coupling method since the computational cost is lower than the direct method. The predicted velocity distributions and sound spectra were validated by the experimental measurement. Secondly, the direct method was conducted to include the acoustic feedback on the flow into the simulation. The spectra and velocity distributions of the direct method were also validated by the experiment. After confirming that the computational accuracy is enough to distinguish the frequency characteristics of /s/ and /ʃ/, the flow and acoustic fields in the simplified and realistic geometries were analyzed. By comparing the flow and acoustic fields in the simplified vocal tract geometry, the cause of the frequency characteristics of /s/ and /ʃ/ was investigated. Moreover, differences of the flow and acoustic fields between /s/ and /ʃ/ were examined in the realistic geometry. Finally, by comparing the simplified and realistic replicas, the aeroacoustic generation mechanisms of sibilant fricatives were discussed.

4.1 Method

4.1.1 Governing Equations

In the coupling method, the flow and sound in the simplified replica were simulated separately with the assumption that feedback effects of the generated sound on flow are negligible in the vocal tract. Firstly, flow fields were simulated by LES of incompressible fluid. The governing equations are spatially filtered incompressible Navier-Stokes equations:

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0, \tag{4.1}$$

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial \bar{u}_i \bar{u}_j}{\partial x_j} = -\frac{1}{\rho_0}\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial \bar{\sigma}_{ij}}{\partial x_j}, \tag{4.2}$$

where, $\bar{u}_i$ $(i = 1, 2, 3)$ are grid-scale (GS) velocity components, $\bar{p}$ is GS pressure, ρ_0 is constant pressure. The GS stress tensor $\bar{\sigma}_{ij}$ is calculated as

$$\bar{\sigma}_{ij} = 2(\nu_{SGS} + \nu)\bar{s}_{ij}, \tag{4.3}$$

with GS kinematic viscosity ν, and subgrid-scale (SGS) viscosity ν_{SGS}. The kinematic viscosity is the values of air at atmospheric pressure and 20°C. The strain rate tensor $\bar{s}_{ij}$ is calculated as $\bar{s}_{ij} = 1/2\,(\partial \bar{u}_i/\partial x_j + \partial \bar{u}_j/\partial x_i)$. The SGS viscosity is estimated by Smagorinsky model as

$$\nu_{SGS} = (C_{\mathrm{s}}\Delta)^2\sqrt{2\bar{s}_{ij}\bar{s}_{ij}}, \tag{4.4}$$

where, C_{s} is Smagorinsky coefficient and Δ is the scale filter length. The coefficient C_{s}

is calculated locally in time and space by dynamic Smagorinsky model (DSM), (Germano *et al*.,1991; Lilly, 1992).

After the flow computation, acoustic fields were calculated by using Lighthill's acoustic analogy in frequency domain (Oberai *et al*, 2000):

$$-\nabla^2\rho' - k^2\rho' = M^2\frac{\partial^2\bar{u}_i\bar{u}_j}{\partial x_i\partial x_j}, \tag{4.5}$$

where ρ' is the amplitude of density fluctuation in frequency domain, *k* is wave number, and M is Mach number. Sound source in frequency domain (right-hand side in Eq. (4.5)) was calculated by discretized Fourier transform (DFT) using flow velocity at each grid. Both flow and acoustic simulations were conducted by second-order scheme FEM softoware FrontFlow/blue ver. 8.1 (Guo *et al*., 2006).

To consider the feedback of the generated sound on the flow, LES in a 3D compressible fluid was conducted in the direct method. The governing equations are spatially filtered compressible Navier-Stokes equations and the equation of state:

$$\frac{\partial\bar{\rho}}{\partial t} = -\frac{\partial\bar{\varphi}_j}{\partial x_j}, \tag{4.6}$$

$$\frac{\partial\bar{\varphi}_i}{\partial t} + \frac{\partial\bar{\varphi}_i\tilde{u}_j}{\partial x_j} = -\frac{\partial\bar{p}}{\partial x_i} + \frac{\partial\bar{\tau}_{ij}}{\partial x_j}, \tag{4.7}$$

$$\frac{\partial\bar{\rho}\bar{E}}{\partial t} + \frac{\partial\bar{\varphi}_j\bar{E}}{\partial x_j} = -\frac{\partial\bar{p}\tilde{u}_j}{\partial x_j} + \frac{\partial\bar{\tau}_{ij}\tilde{u}_j}{\partial x_j} - \frac{\partial\bar{q}_j}{\partial x_j}, \tag{4.8}$$

$$\bar{p} = (\gamma-1)\bar{\rho}\bar{e}, \qquad \bar{e} = C_{\mathrm{v}}\bar{T}. \tag{4.9}$$

where $\bar{\rho}$ is density, $\bar{\varphi}_i = \overline{\rho u_i}$ is momentum, $\bar{E} = 1/2\,|\tilde{u}_i|^2 + \bar{e}$ is total energy per unit mass, $\bar{e}$ is internal energy, and $\bar{T}$ is temperature. The symbol $\tilde{\cdot}$ represent Favre mean filtered value (Favre, 1969):

$$\begin{aligned}\bar{\rho}(\boldsymbol{x}) &= \iiint G(\boldsymbol{x}-\boldsymbol{x}')\rho(\boldsymbol{x}')d^3\boldsymbol{x}',\\ \tilde{u}_i(\boldsymbol{x}) &= \frac{\iiint G(\boldsymbol{x}-\boldsymbol{x}')\rho u_i(\boldsymbol{x}')d^3\boldsymbol{x}'}{\iiint G(\boldsymbol{x}-\boldsymbol{x}')\rho(\boldsymbol{x}')d^3\boldsymbol{x}'}.\end{aligned} \tag{4.10}$$

The GS viscous stress tensor $\bar{\tau}_{ij}$ and heat flux $\bar{q}_j$ are calculated as

$$\bar{\tau}_{ij} = 2\bar{\rho}(\nu+\nu_{\mathrm{SGS}})\left(\tilde{s}_{ij} - \frac{1}{3}\delta_{ij}\tilde{s}_{ll}\right), \tag{4.11}$$

$$\bar{q}_j = -\bar{\rho}\gamma C_{\mathrm{v}}(\alpha+\alpha_{\mathrm{SGS}})\frac{\partial\bar{T}}{\partial x_j} \tag{4.12}$$

with GS thermal diffusivity α, and SGS thermal diffusivity α_{SGS}. The strain rate tensor $\tilde{s}_{ij}$ is calculated as $\tilde{s}_{ij} = 1/2\,(\partial\tilde{u}_i/\partial x_j + \partial\tilde{u}_j/\partial x_i)$. The gas constants $C_{\mathrm{v}}\,and\,\gamma$ are specific heat and specific heat ratio, respectively. The gas constants are the values of air at atmospheric pressure and 20°C. In this study, ν_{SGS} and α_{SGS} were estimated with one equation type subgrid-scale model (Fureby, 1996):

$$\nu_{\mathrm{SGS}} = C_k\Delta\sqrt{k_{\mathrm{SGS}}}, \tag{4.13}$$

$$\alpha_{\mathrm{SGS}} = \nu_{\mathrm{SGS}}/P_{rt}, \tag{4.14}$$

$$\frac{\partial \bar{\rho} k_{\mathrm{SGS}}}{\partial t} + \frac{\partial k_{\mathrm{SGS}} \bar{\varphi}_j}{\partial x_j} + \frac{\partial}{\partial x_j}\left\{\bar{\rho}(\nu + \nu_{\mathrm{SGS}}) \frac{\partial k_{\mathrm{SGS}}}{\partial x_j}\right\} = -\tau_{ij} \frac{\partial \tilde{u}_i}{\partial x_j} - C_\epsilon \frac{\bar{\rho} k_{\mathrm{SGS}}^{3/2}}{\Delta}, \quad (4.15)$$

where

$$\tau_{ij} = \frac{2}{3} \bar{\rho} k_{\mathrm{SGS}} \delta_{ij} - 2\bar{\rho} \nu_{\mathrm{SGS}} \left(\tilde{s}_{ij} - \frac{1}{3} \delta_{ij} \tilde{s}_{ll}\right), \quad (4.16)$$

and dimensional constants were C_k = 0.094, C_ϵ = 1.04, and the turbulence Prandtl number was P_{rt} = 1.0.

The spatial derivatives were discretized by the second-order accurate central differencing scheme, and the time integration was performed by the second-order accurate Crank-Nicolson method. The equations were implemented and solved in the finite volume method software OpenFOAM 2.3.1 (OpenCFD Ltd).

4.1.2 Computational Grids

For the coupling method, the simplified vocal tract geometry consists of teeth blades and a rectangular flow channel representing a throat, constriction formed by tongue and upper jaw, space behind the teeth, and lip cavity. Overall geometry is illustrated in Fig. 4.1. Dimensions of the channel were determined based on the vocal tract geometry of 32 years old Japanese male subject measured by CT images in the same way as Chapter 3. The CT images were obtained while the subject sustained the sibilant /s/ for 9.6 s in a seated position. The cross-sectional areas and vertical heights at five positions (throat, constriction, cervical area, gap between teeth, lip cavity) were used to construct the model. The dimensions of each section are summarized in Table 3.1. It has been confirmed that the model reproduces the subject's sibilant /s/ in the frequency range 0.5 to 14 kHz.

To validate the computational accuracy, three sets of computational grids were prepared for the coupling method. The grids from the constriction exit to the middle of lip cavity are shown in Fig. 4.2. Comparing to 10 million grids (10M), minimum element size of 30 million grids (30M) is larger since grid stretching nearby the surface of the model is moderate. To capture the flow separation from the constriction exit and teeth gap, the grid size near the walls was decreased in 45 million grids (45M). Computational parameters for three grid sets are summarized in Table 4.1. In 45M, the time step Δt was smaller than those of other grid sets because of the minimum element size. For each set of the computational grids, a far-field region was constructed to simulate the sound propagation from the outlet of the model. The far-field region of 45M was enlarged to evaluate the influence of the region size on the sound propagation. At the inlet of the constriction in 30M and 45M, round edges with radius 0.1 mm were formed to smooth the contraction flow between the throat and constriction.

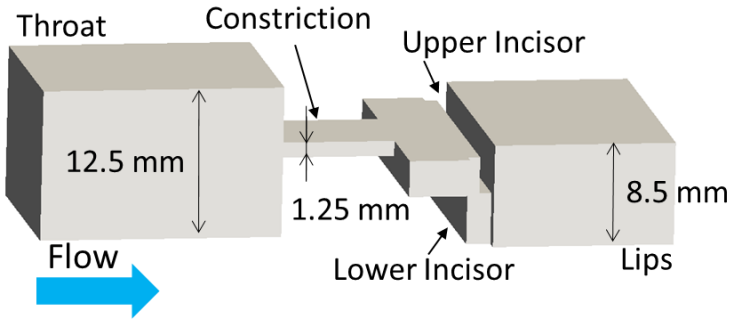


Fig.4.1 Schematic illustration for simplified vocal tract geometry of sibilant /s/.

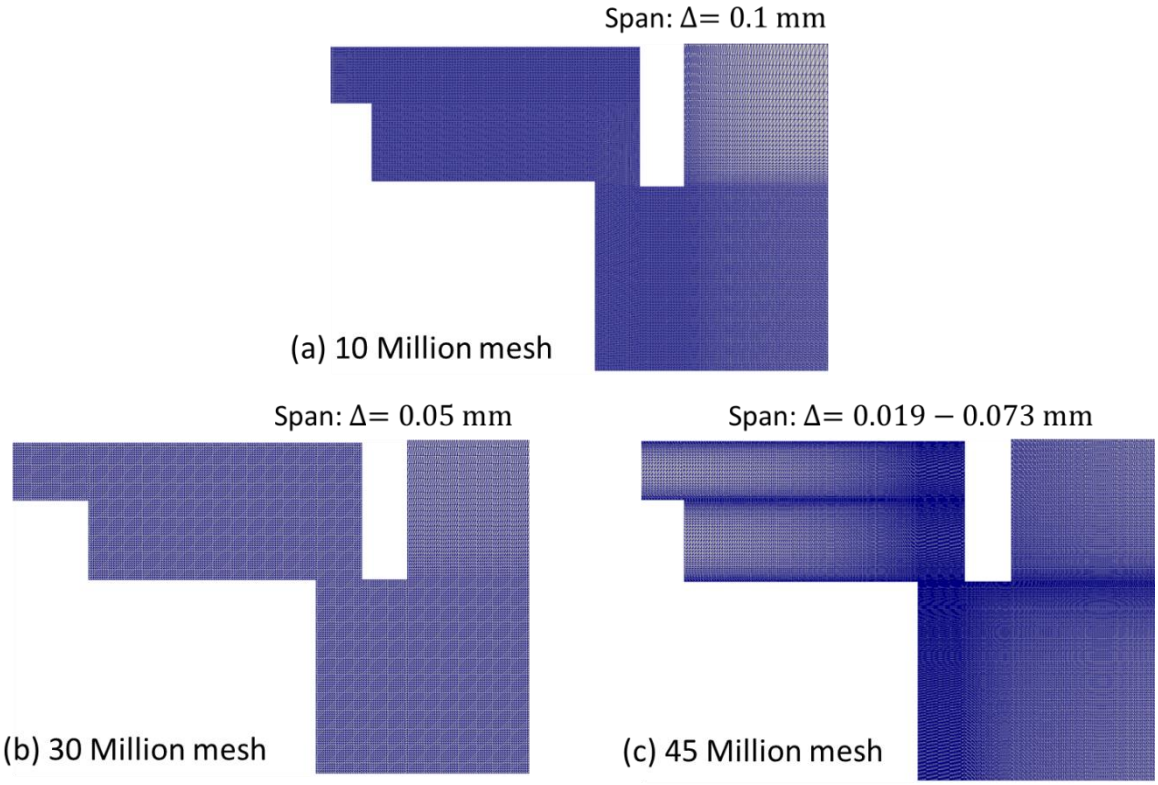


Fig. 4.2 Computational grids for coupling method. (a) 10M, (b) 30 M, and (c) 40 M.

Table 4.1 Parameters for the flow simulation of the coupling method.

Total number of elements	Minimum element size (mm)	Δt (s)	Simulated time (s)	Size of far-field region (mm)
10 million	2.9×10^{-2}	5×10^{-7}	0.005 - 0.018	100 × 68 × 63
30 million	4.9×10^{-2}	5×10^{-7}	0.028 - 0.045	100 × 68 × 63
45 million	1.7×10^{-2}	1×10^{-7}	0.016 - 0.020	200 × 200 × 200

For the direct method (the compressible flow simulation), effects of the tongue position on the flow and sound generated by the simplified geometry were examined. The simplified vocal tract geometry for the direct method is illustrated in Fig. 4.3. The tongue

model, which had a 1.3 × 8 mm groove on the top and tip L_T = 4 mm at the front, was positioned between the back cavity and the lower teeth. The length between the tongue tip and lower teeth was defined as L_C. When $L_C > 0$, a lower mouth cavity appears between the lower teeth and tongue blade. The length L_C was varied from L_C/L_T = 0–2. The x_1-axis indicates the anterior-posterior direction, the x_2-axis indicates the upper-lower direction, and the x_3-axis indicates the transverse direction in this chapter. The origin of the coordinate system was set at the exit of the constriction.

The computational grids for the direct method are shown in Fig. 4.4. To simulate the sound pressure propagating from the model, a far-field region (100 × 50 × 60 mm) was added to the exit of the lip cavity. Since the frequency range of the speech sounds of interest is less than 15 kHz, the smallest wavelength of interest, 22.9 mm, was captured with more than seven grids (the maximum length of the grids was 3.26 mm at the outlet of the far-field). To capture the flow separation from the constriction and the gap between the teeth, grid sizes near the wall of the constriction and teeth were reduced. The minimum element size was 1.6×10^{-2} mm. The total number of grid points was approximately 40.5×10^6 when L_C/L_T = 0 and 47.2×10^6 when L_C/L_T = 1.25. The time step for the time integration was 1×10^{-7} s, and 13.5×10^4 iterations were performed after 2.5×10^4 preliminary iterations required to achieve developed flow in the vocal tract.

The computational grids of the realistic geometry pronouncing /s/ and /ʃ/ are depicted in Fig. 4.5. For the geometry of /s/, CT images of the Japanese male subject sustained /s/ for 9.6 s were used (accuracy: 0.1 × 0.1 × 0.1 mm voxels). For the geometry of /ʃ/, CT images of the same subject pronouncing in word context /mʲisoʃirɯ/ were used (accuracy: 0.488 × 0.488 × 0.5 mm voxels). The surface of the vocal tract geometry was extracted by using the segmentation software itk-SNAP (http://www.itksnap.org). Because of lower spatial accuracy for the CT scan of /ʃ/, the surface of the vocal tract /ʃ/ was rougher than the surface of /s/. The grid sizes from the constriction to lip cavity were decreased to capture the small vortices in the turbulent flow region. The minimum element size was 6.2×10^{-2} mm. The total number of grid points was approximately 23×10^6 for /s/ and 35×10^6 for /ʃ/.

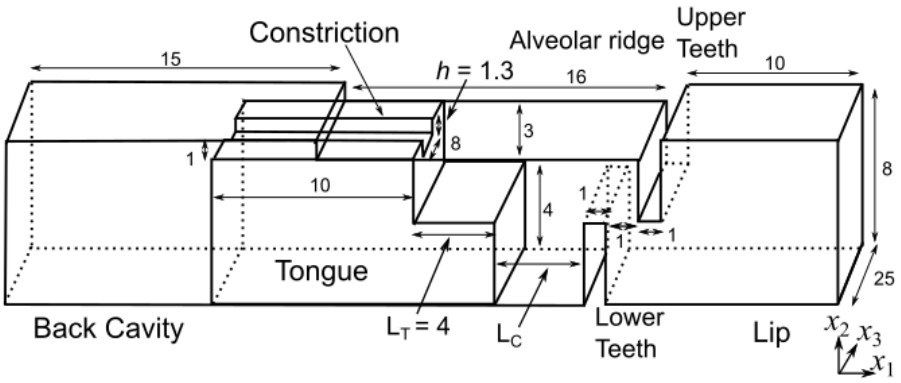


Fig. 4.3 Simplified vocal tract geometry based on the vocal tract of a Japanese male speaker. The units are in mm.

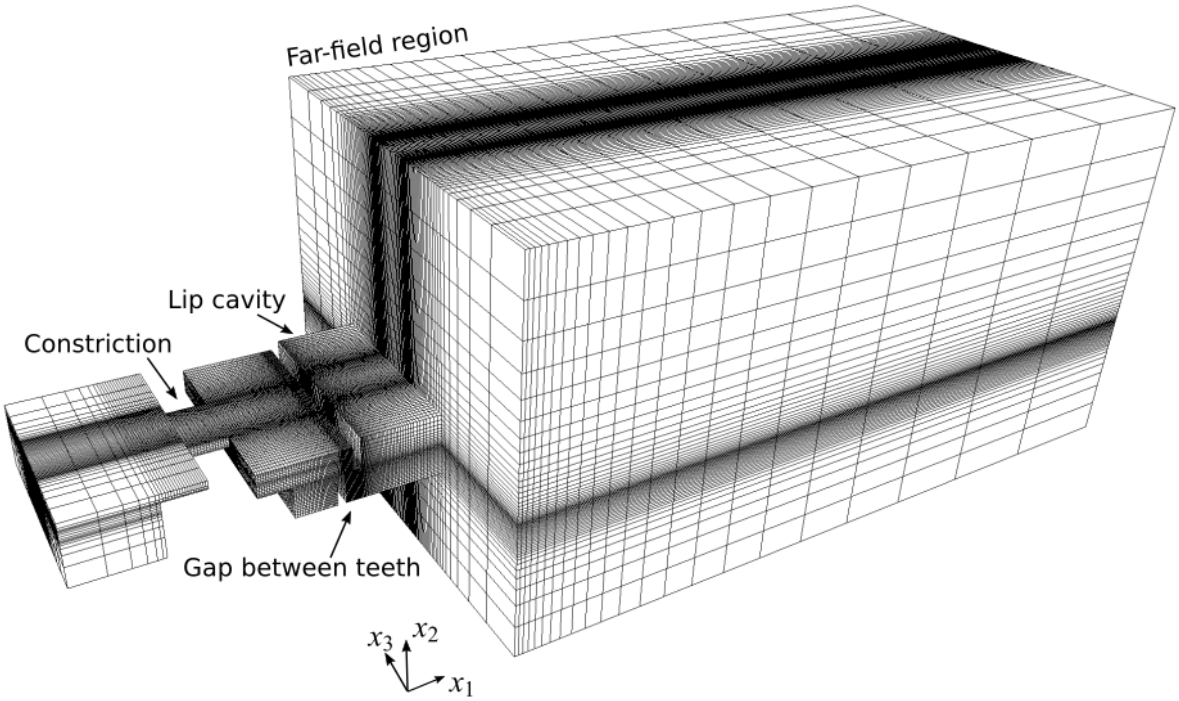


Fig. 4.4 Computational grid for the simple replica L_C/L_T = 1.25 with far-field region (every 5th grid line is shown for clarity).

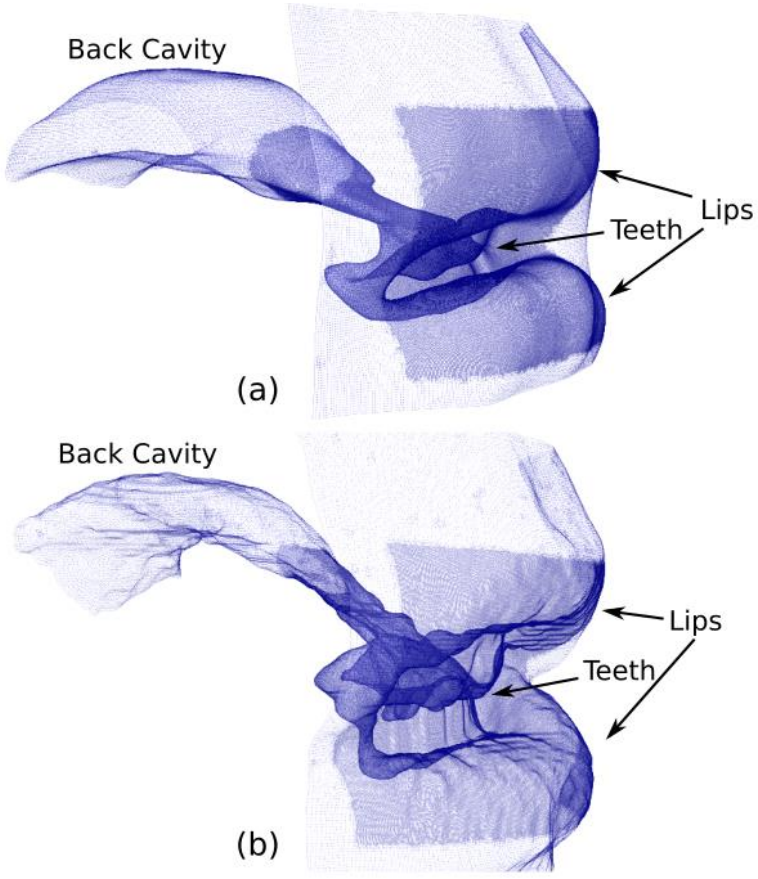


Fig. 4.5 Computational grid for realistic geometry of the subject pronouncing /s/ and /ʃ/.

4.1.3 Boundary Conditions

Boundary conditions for the coupling method are presented in Fig. 4.6. For the flow simulation, uniform velocity and constant pressure were set on inlet and outlet boundaries, respectively, to yield steady flow rate 400 cm^3/s at the inlet. No-slip boundary condition was used on the surface of the vocal tract. For the acoustic simulation, non-reflecting boundary condition (NRBC) was set on the inlet and outlet boundaries, and the surface of the vocal tract was set as rigid wall. The velocity 1.5 mm downstream from the upper teeth wall and sound pressure at 20 mm along the axis x_1 were extracted to validate the computational accuracy.

Boundary conditions used in the direct method are depicted in Fig. 4.7. At the inflow boundary, a uniform velocity of $\tilde{u}_1$ = 1.5 m/s at a constant temperature of $\bar{T}$ = 20 °C was used to produce the subject's flow rate of 300 cm^3/s. No-slip and adiabatic boundaries were used at the model walls. At the outlet of the far-field region, a no-reflecting boundary condition (Poinsot and Lele, 1992) was imposed to allow the acoustic waves to pass smoothly with minimal disturbances. The velocity 0.5 mm downstream from the upper teeth wall and sound pressure at 90 mm along the axis x_1 were extracted to validate the computational accuracy. Mach number and Reynolds number based on the maximum mean velocity and height (h = 1.3 mm) at the constriction were 0.127 and 3697, respectively, when L_C/L_T = 0 ($|\overline{\tilde{u}}|_{max}$ = 43.6 m/s at x_1/h = -7), and 0.121 and 3527, respectively, when L_C/L_T = 1.25 ($|\overline{\tilde{u}}|_{max}$ = 41.6 m/s at x_1/h = -3.5).

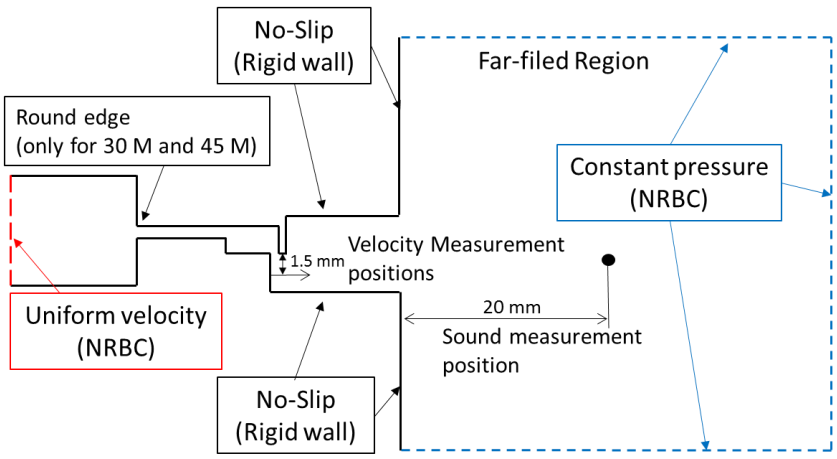


Fig 4.6 Boundary conditions for incompressible flow and acoustic simulation.

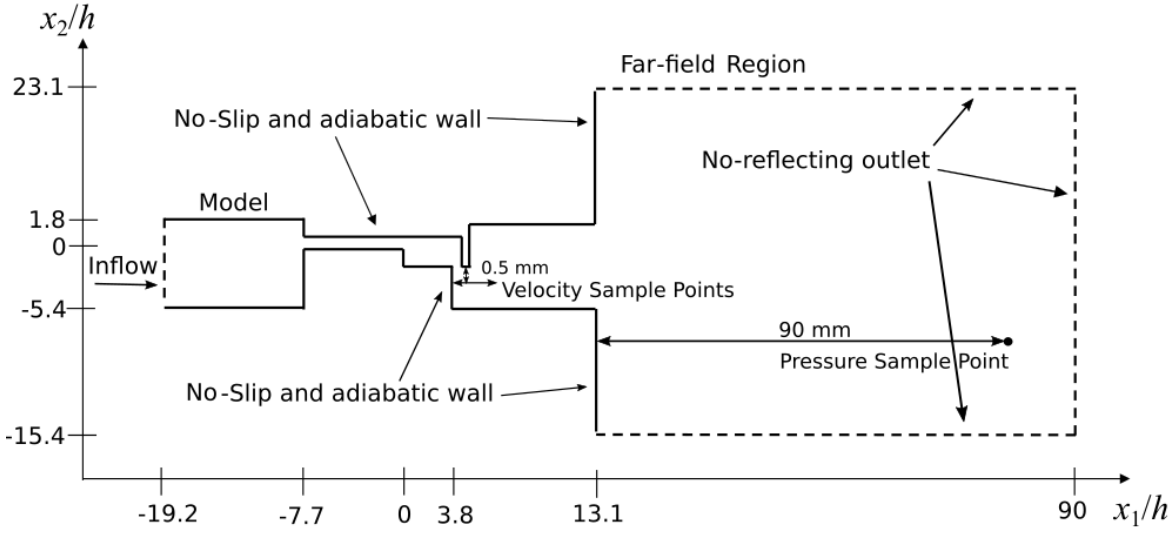


Fig. 4.7. Boundary conditions for compressible flow.

4.1.4 Sound Source and Pressure Amplitude in the Frequency Domain

To evaluate the sound source generated by the flow fluctuation in the direct method, the source term ψ in Lighthill's analogy (Lighthill, 1952)

$$\psi = \frac{\partial^2 T_{ij}}{\partial x_i \partial x_j}, \quad T_{ij} \cong \bar{\rho}\tilde{u}_i\tilde{u}_j \tag{4.17}$$

was calculated from the GS density and velocity components. The second and third term in Lighthill's analogy were neglected because of low Mach number and high Reynolds number of the flow inside the vocal tract.

To examine the cause of the frequency characteristics of the generated sound, the frequency components of the sound source and pressure fluctuation were obtained by DFT. The simulated values were sampled at a rate of 100 kHz and the DFT was performed on 256-point values, which were multiplied by a Hanning window, and averaged with 7 sets of 30% overlapped frames. The frequency resolution of these calculated values was 391 Hz. The sound pressure level (SPL) was calculated as

$$SPL = 20\log_{10}(|\bar{p}'|/p_0), \tag{4.18}$$

where $\bar{p}'$ is the pressure in the frequency domain and p_0 = 20 × 10^{-6} Pa is the reference sound pressure.

The spatial mean values of the sound source and pressure in the transverse (x_2-x_3) plane were calculated by taking the surface integral over the plane surface S_Ω for both the source magnitude and SPL,

$$\psi_\Omega = \frac{1}{S_\Omega}\int_{S_\Omega} 10\log_{10}|\psi'|dS, \quad SPL_\Omega = \frac{1}{S_\Omega}\int_{S_\Omega} SPL\, dS \tag{4.19}$$

where ψ' is the source term in the frequency domain. In discretized form, the spatial mean was estimated as

$$\psi_\Omega \cong \frac{1}{S_\Omega}\sum_{k=1}^{N} 10\log_{10}|\psi_k'|S_k, \quad SPL_\Omega \cong \frac{1}{S_\Omega}\sum_{k=1}^{N} SPL_k S_k, \tag{4.20}$$

where N is the number of elements on S_Ω. ψ_k', SPL_k, and S_k are the sound source, SPL, and surface area at the element k, respectively.

4.2 Experimental Validation of Computational Accuracy

4.2.1 Experimental Method

The sound pressure and flow velocity generated from the simplified model were measured in the experimental setup depicted in Fig. 4.8. The vocal tract geometry shown in Fig. 4.1 was constructed by the plaster using 3D printer (Zprinter, 3D systems), and the geometry shown in Fig. 4.3 was constructed with acrylic boards. Steady airflow was delivered to the inlet of the model from a compressor (YC-4RS, Yaezaki, Tokyo, Japan) via an air tube with an inner diameter of 16 mm and a mass flow controller (MQV0050, Azbil, Tokyo, Japan). The flow rate was fixed in the range of subject's physiological flow rates: 400 cm^3/s for the coupling method and 300 cm^3/s for the direct method. The air tube was connected to the inlet of the back cavity through a tube connecter with an inner diameter of 8 mm and a preliminary rectangular duct with dimensions of 8 × 25 × 125 mm with straightening vanes attached inside. The total length from inlet to outlet of the rectangular duct was 170 mm, and is consistent with the subject's vocal tract length from the vocal fold to the lips along the center line. A rectangular baffle of 350 × 350 mm was attached at the outlet of the model to mimic the speaker's face.

The flow velocity downstream of the teeth was measured with a hot-wire anemometer (d = 5 μm, l = 2 mm wire, 0251R-T5, Kanomax, Osaka, Japan). The tip of the anemometer was placed every 0.2 mm along the velocity sample points depicted in Fig. 4.6 and Fig. 4.7 by using x-y-z axis stages (LS-4042-S1; ALS-115-E1P, Chuo Precision Industrial, Tokyo, Japan) shown in Fig. 4.9 (a). The anemometer was calibrated in a small wind tunnel (Model 1065, Kanomax) every 1 m/s from 2 to 10 m/s, and every 5 m/s from 10 to 55 m/s using the power law (Khan *et al.*, 1987). The sound pressure at 20 mm (for coupling method) and 90 mm (for direct method) from the lip outlet along the x_1-axis was measured by a 1/4 inch omnidirectional microphone (Type 4939, Bruel & Kjaer, Nærum, Denmark) as shown in Fig. 4.9 (b). We used a hot-wire anemometer to confirm that sound was measured in a quiescent medium (flow velocity was less than 1 m/s at 90 mm from the lip). The microphone and the model were placed at least 0.6 m from the walls with soundproof materials in an experimental room.

In addition to the simplified replica, the realistic vocal tract replicas of /s/ and /ʃ/ were constructed to validate the computational accuracy for the simulation with the realistic geometries. The replicas were constructed by using the vocal tract surfaces of the computational grids and 3D printer (Objet30Pro, Stratasys, USA). The constructed realistic replicas are shown in Fig. 4.10. The replicas were made of acrylic resin and connected the air compressor through the air tube. The baffle board (350 × 350 mm) was attached at the edge of the replica's front face. The measurement setups are shown in Fig. 4.11.

The velocity and sound pressure for 1 s were recorded with sampling frequency 100 kHz and 44.1 kHz, respectively, using a data acquisition system (PXIe-4492, National Instruments, Austin, USA). The sound spectrum was calculated by discretized Fourier transform (DFT) with 256-point signals, which were multiplied by a Hanning window, and averaged with 60 sets of 30% overlapped frames. The frequency resolution of the calculated values was 172 Hz.

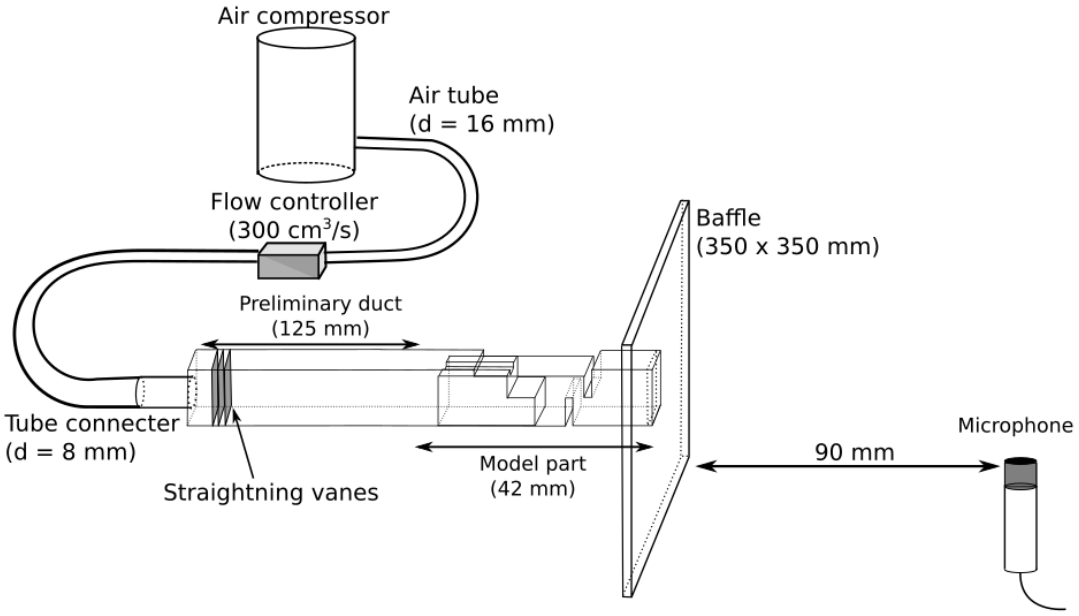
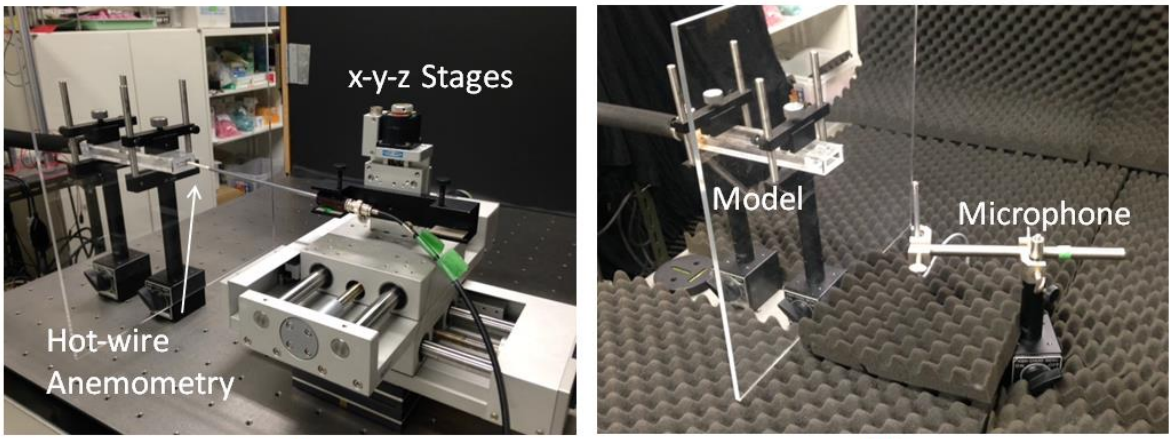
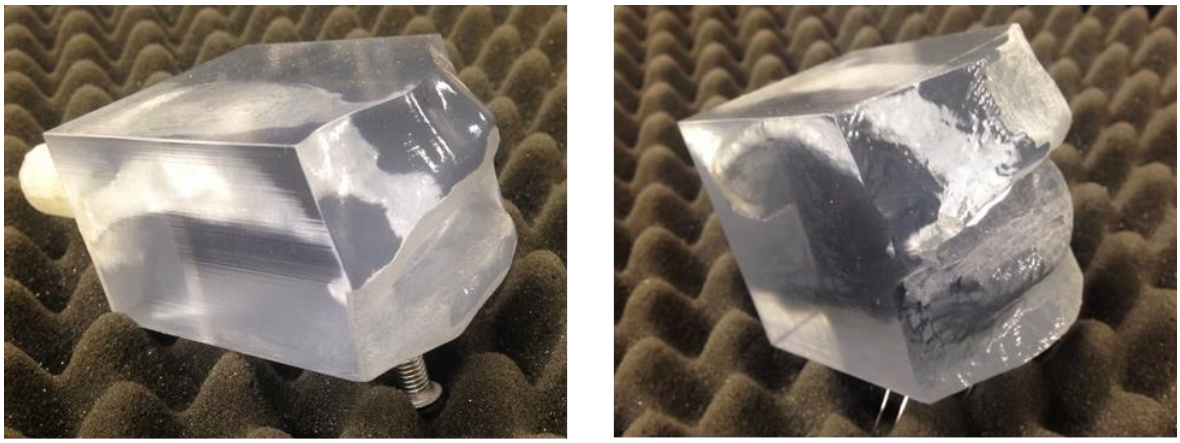


Fig. 4.8 Schematic of the experimental setup for sound measurement.



(a) (b)

Fig. 4.9 Experimental setups for velocity measurement (a) and sound measurement (b).



(a) (b)

Fig. 4.10 Realistic Replica of /s/ (a) and /ʃ/ (b).

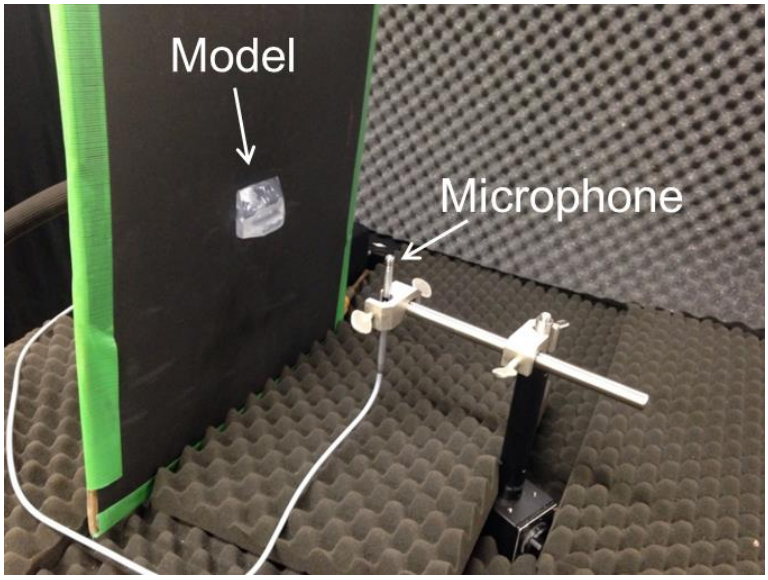


Fig. 4.11 Measurement setups for realistic replica.

4.2.2 Velocity Distribution

For the coupling method, instantaneous velocity fields between the constriction and the lip cavity in mid-sagittal plane are shown in Fig. 4.12. In 10M, flow at the constriction became fully turbulent and passed nearby the surface of upper and lower teeth. In 30M and 45M, flow at the constriction became laminar and formed recirculatory flow nearby the surface of upper teeth. This recirculatory flow widened the angle between the jet flow and the surface of lower teeth. The difference of flow state at the constriction was caused by the difference of edge shape (*i.e.* angular edge in 10M and round edge in 30M and 45M) at the inlet of the constriction. By making the round edge, flow passed smoothly through the edge of constriction and was not disturbed.

Mean and root mean square (RMS) values of the velocity distribution at 1.5 mm below the upper teeth edge for the coupling method are plotted in Fig. 4.13 (a) and (b), respectively. Since the velocity component in the transverse direction (axis x_3) was small ($\bar{u}_3 \ll \bar{u}_1, \bar{u}_2$) at the sample points, the velocity magnitude measured by the hot-wire anemometer was estimated by calculating the velocity vector $u_h = (\bar{u}_1^2 + \bar{u}_2^2)^{1/2}$ in the simulation. In addition, to consider the short simulation time, mean and RMS values of the experiment were calculated for shortened time (*i.e.* 0.004 s) and standard deviations of mean and RMS were calculated for the recorded time (*i.e.* 1 s). The peak of the mean velocity measured by the hot-wire anemometer appeared between 1 and 1.4 mm from the lower teeth, and the position of the peaks in simulation of 30M and 45M agreed with the peak observed in the experiment. Meanwhile, the peak in 10M appeared at 0.8 mm from the lower teeth and this position was different from the experimental observation. This shift of the peak in simulation was caused by the flow state upstream of the jet flow as shown in Fig. 4.12. RMS values at the separation region of the jet flow (distance 1.5-2.5 mm from the lower teeth) in 30M were decreased by increasing the grid resolution to 45M. This tendency agreed with the case of LES for the flow-separation (Kato *et al.*, 2007).

For the direct method, the simulated velocity distribution along the velocity sample points 0.5 mm downstream from the upper teeth was compared with the measured velocities. The velocity magnitude measured by the hot-wire anemometer was estimated by calculating the velocity vector $u_h = (\tilde{u}_1^2 + \tilde{u}_2^2)^{1/2}$. Mean and RMS values of the velocity fluctuation in the replica with L_C/L_T = 0 and 1.75 are plotted in Fig. 4.14. Since the simulation time 0.0132 s was much shorter than the recording time 1.0 s in the experiment, variations of the mean and RMS for shortened recording time 0.0132 s were plotted as bars. Although the mean values at the peak x_1 = -11 mm were slightly overestimated in the simulation for both L_C/L_T = 0 and 1.75, shift of the jet region from L_C/L_T = 0 to 1.75 was captured by the simulation. Moreover, the shift of RMS distributions from L_C/L_T = 0 to 1.75 was also captured in the simulation. The overestimated values in the simulation were probably caused by the coarse grid size at the separation region downstream of the teeth (Kato, *et al.*, 2007). Further agreement might be achieved by refining the grid size nearby the upper and lower teeth.

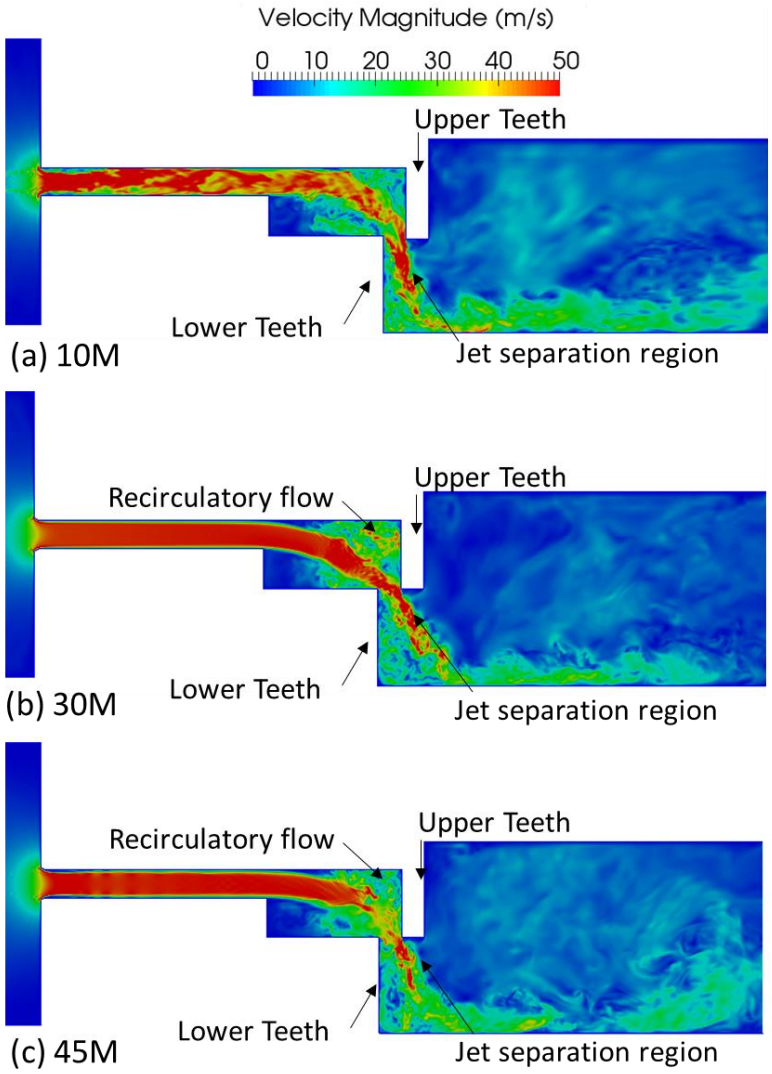


Fig 4.12 Instantaneous velocity magnitude in midsagittal plane of 10 million grids (a), 30 million grids (b), and 45 million grids (c) in coupling method.

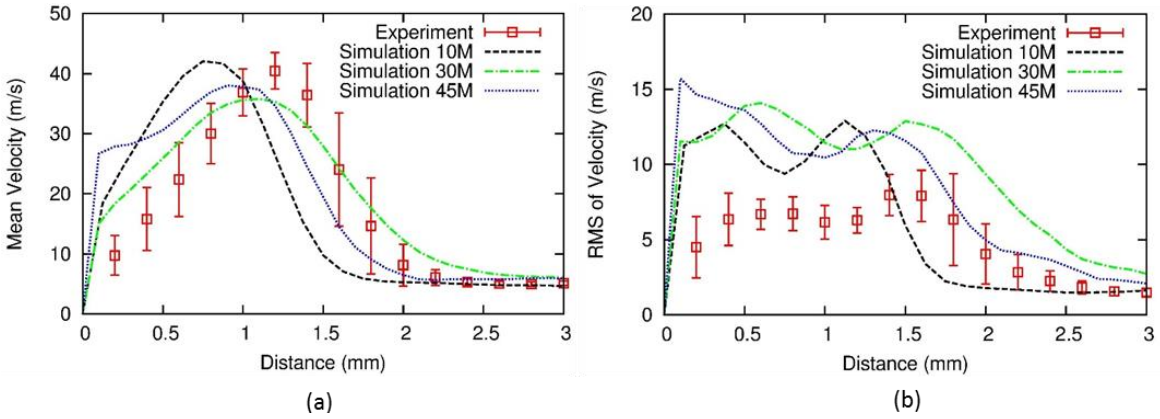


Fig. 4.13 Mean (a) and RMS (b) of velocity distribution predicted by the coupling method at 1.5 mm below the teeth.

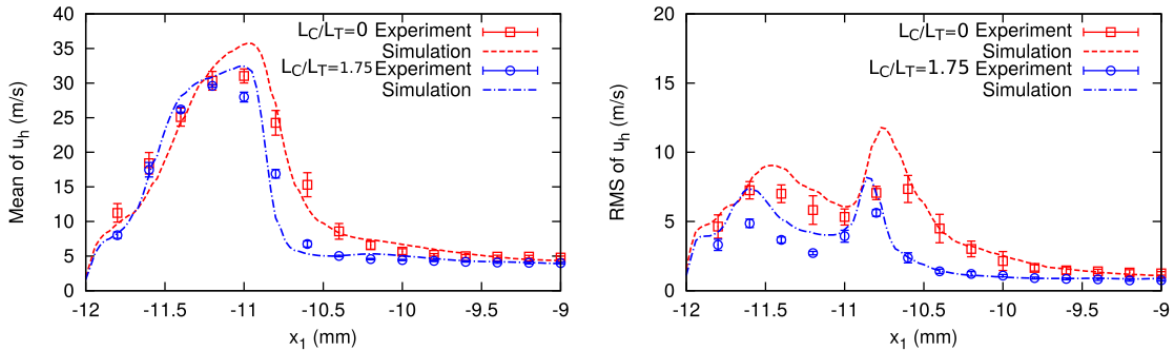


Fig. 4.14 Mean (a) and RMS (b) values of velocity fluctuation predicted by the direct method at 0.5 mm downstream from the upper teeth edge.

4.2.3 Sound Spectrum

The frequency spectra of sound at 20 mm from the outlet in the coupling method are shown in Fig. 4.15. The spectral shape of 10M and 45M roughly agreed with that of the experiment in the frequency range 1 to 5 kHz. The differences between measured and estimated values at the frequency below 3.5 kHz were larger than those over 3.5 kHz. Those larger values might be decreased by increasing the number of averaging values in DFT of flow source since lower frequency sound consists of longer wave length. Meanwhile, agreement of the sound level in 10M indicated that the size of far-field region in 10M is enough to simulate the sound spectrum at 20 mm from the outlet of the model.

The PSD at 3.1 kHz simulated by 30M was 12 dB larger than the measured PSD. The sound source in the vocal tract was mainly generated by the velocity fluctuation in the jet flow downstream from the teeth gap. Therefore, this larger PSD were probably caused by the over prediction of the velocity fluctuation (*i.e.* RMS values) at the flow separation region which was caused by the low grid resolution in 30M. We note that the PSD at 3.1 kHz simulated in 10M was smaller than the PSD of 30M, since the minimum element size near the separation region of 10M was smaller than that of 30M (see Table 4.1). Hence,

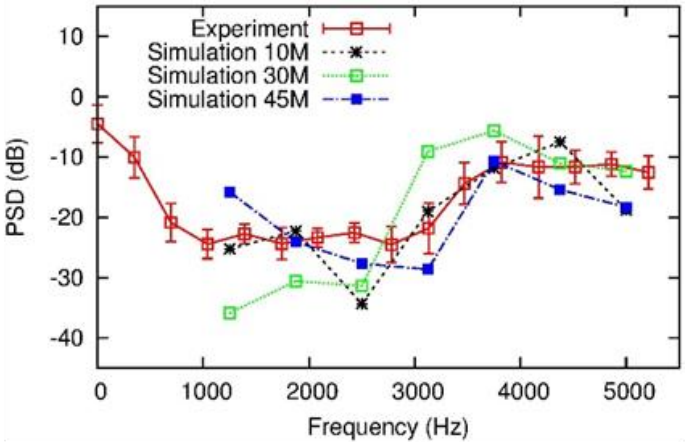


Fig. 4.15 Frequency spectra of sound predicted by the coupling method at 20 mm from outlet of the model.

this result indicates that it is important to increase the grid resolution at the separation region downstream from the teeth gap in order to accurately simulate the sound spectrum of sibilant /s/.

The spectral shapes of the sound estimated by the direct method for the simplified geometry with $L_C/L_T = 0$ and 1.25 are compared with those of the experimental replica and /s/ and /ʃ/ pronounced by the subject in Fig 4.16. The variations of SPL for the subject's pronunciation over the 15 trials are plotted as bars. The maximum SPL discrepancy between the replica and subject was less than 5 dB in the frequency range 1.5–15 kHz. This indicates that the frequency characteristics of /s/ and /ʃ/ can be represented by sounds generated by the replica with $L_C/L_T = 0$ and 1.25, respectively, up to 15 kHz. The largest SPL difference between the experiment and simulation was observed below 1.5 kHz. This difference was probably caused by a small averaging number for DFT because of the high computational cost of the simulation, and the no-reflecting boundary condition in the numerical simulation, which could not pass the low frequency oscillations (see the Appendix B for details).

The spectra predicted by the direct method for the realistic replica are shown in Fig. 4.17. The sounds generated by the acrylic realistic replica as well as by the subject's /s/ and /ʃ/ were compared. The SPLs in the simulation and experiment were collected at 68 mm along axis x_1. For the realistic geometry of /s/, the first characteristic peak appeared at 5 kHz and the overall peak appeared at 9 kHz in the same way as the subject's /s/. The maximum discrepancy between the simulation and measurement was less than 12 dB in the frequency range 2 to 22 kHz. For the realistic geometry of /ʃ/, the characteristic peak at 4 kHz was captured by the direct method. These results indicate that the LES on both the realistic and simplified geometries can express the sound of subject's /s/ up to 22 kHz.

For the case with the realistic geometry of /s/, a large peak, which did not appear in the subject's /s/, was observed at 9 kHz. This peak was caused by the periodic vortices generated at the constriction. In addition, the characteristic peaks observed at 3.5 and 5.5 kHz for the realistic replica of /ʃ/ were 1 kHz shifted to higher frequency range in the simulation. At the constriction of /s/ and /ʃ/, the maximum vertical height of the flow

channel is 1.3 mm, and the spatial resolution of the CT images is 0.1 mm for /s/ and 0.5 mm for /ʃ/. Therefore, we speculate that these phenomena occurred because of the error in the surface extraction from the CT images. Further quantification on quality of the extracted surface is needed to obtain more accurate result.

Nevertheless, the velocity distribution downstream from the teeth and the overall spectral shapes of the subject's /s/ and /ʃ/ were captured by the direct method. Therefore, the direct method can be used to accurately simulate the flow and pressure fields inside the vocal tract during the pronunciation of /s/ and /ʃ/ sounds. From next section, the flow and acoustic fields estimated by the direct method are analyzed and discussed.

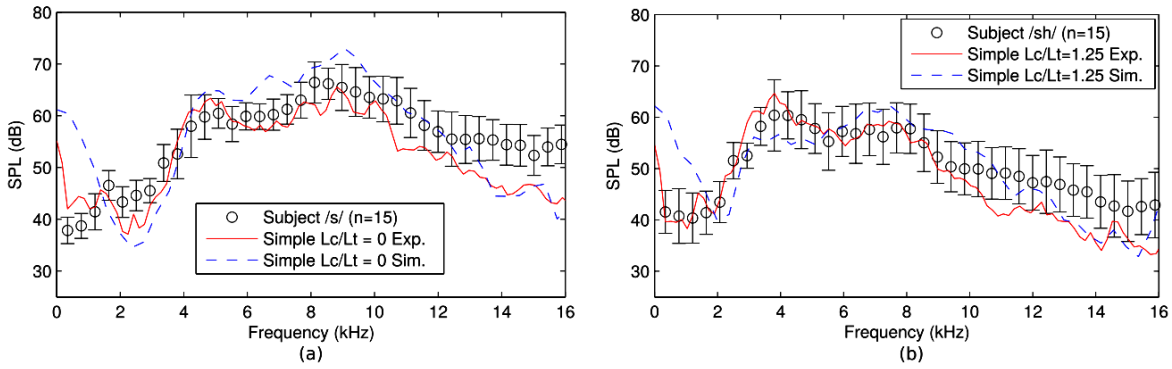


Fig. 4.16 Spectra of sound pressure in the direct method at x_1/h = 82.3 and x_2/h = -12.1. The spectra of experimentally measured (red solid line) and simulated (blue dash) sound pressure with the model at the tongue positions L_C/L_T = 0 and 1.25 are shown in (a) and (b), respectively. Variations of SPL in /s/ and /ʃ/ pronounced by the subject 15 times are plotted with black circle and bars.

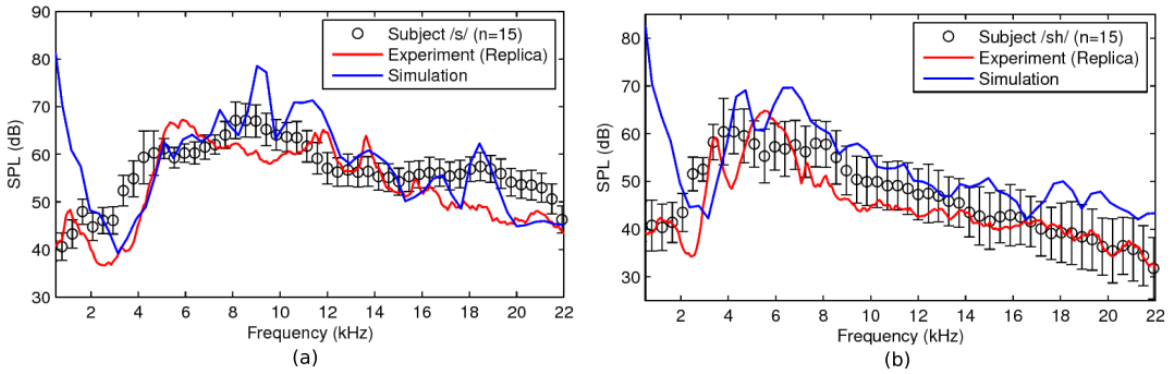


Fig. 4.17 Measured and simulated spectra of sound pressure in the direct method at 68 mm from lips for the realistic geometry of /s/ (a) and /ʃ/ (b). Variations of SPL in /s/ and /ʃ/ pronounced by the subject 15 times are plotted with black circle and bars.

4.3 Detailed Results

4.3.1 Simplified Vocal Tract Geometry

In this section, effects of the tongue position on flow and sound generation in the vocal tract are presented by comparing the fields predicted in the simplified geometry when L_C/L_T = 0 and 1.25. Figure 4.18 and 4.19 show the normalized instantaneous velocity magnitude $|\tilde{u}|/\overline{|\tilde{u}|}_{\max}$, RMS of the velocity fluctuation $|\tilde{u}|_{\mathrm{rms}}/\overline{|\tilde{u}|}_{\max}$, and sound source ψ_{rms} in the vertical (x_1–x_2) plane at the center x_3 = 0. Instantaneous velocity magnitudes (Fig. 4.18 (a) and 4.19 (a)) showed that for both L_C/L_T = 0 and 1.25 the flow that traveled from the back cavity caused the maximum velocity at the entrance of the constriction (x_1/h = -7 for L_C/L_T = 0, x_1/h = -3.5 for L_C/L_T = 1.25), and the flow left from the constriction impinged on the upper teeth (x_1/h = 4.6 for L_C/L_T = 0, x_1/h = 8.5 for L_C/L_T = 1.25). The impinged flow directly passed through the gap between the teeth for L_C/L_T = 0, whereas the flow recirculated in the cavity between the tongue and lower teeth ($3.1 < x_1/h < 6.9$) and disturbed the flow near the tongue tip for L_C/L_T = 1.25. After leaving the gap between the teeth, the flow impinged on the lower lip surface and left the model.

The RMS of the velocity fluctuation (Fig. 4.18 (b) and 4.19 (b)) showed that a large velocity fluctuation appeared at a region above the tongue tip (x_1/h = 3.8) and downstream from the gap between the teeth (x_1/h = 4.6) for L_C/L_T = 0. Meanwhile, for L_C/L_T = 1.25, a large fluctuation appeared above the cavity between the tongue and teeth (x_1/h = 3.1) and downstream from the gap between the teeth (x_1/h = 8.5). The maximum RMS $|\tilde{u}|_{\mathrm{rms}}/\overline{|\tilde{u}|}_{\max}$ = 0.27 observed when L_C/L_T = 0 was larger than the maximum RMS $|\tilde{u}|_{\mathrm{rms}}/\overline{|\tilde{u}|}_{\max}$ = 0.21 when L_C/L_T = 1.25. This difference was caused by the difference in flow configuration in the region between the tongue and teeth. While the flow leaving the constriction instantly impinged on the upper teeth for L_C/L_T = 0, the flow leaving the constriction was decelerated in the cavity between the tongue and teeth for L_C/L_T = 1.25. The decelerated flow in the model with L_C/L_T = 1.25 generated smaller velocity fluctuations compared with the flow in the model with L_C/L_T = 0.

The RMS of the sound source (Fig. 4.18 (c) and 4.19 (c)) showed that a large source fluctuation appeared near the upper and lower teeth ($3.8 < x_1/h < 5.4$) for L_C/L_T = 0, whereas a large fluctuation appeared above the cavity between the tongue and teeth (x_1/h = 3.1) and downstream from the gap between the teeth (x_1/h = 8.5) for L_C/L_T = 1.25. The overall distributions and magnitudes of the sound source were correlated with the RMS of the velocity fluctuation for both L_C/L_T = 0 and 1.25. Since the RMS of the velocity fluctuation for L_C/L_T = 0 was larger than that for L_C/L_T = 1.25, the magnitude of the source for L_C/L_T = 0 was larger than that for L_C/L_T = 1.25, especially near the upper and lower teeth wall.

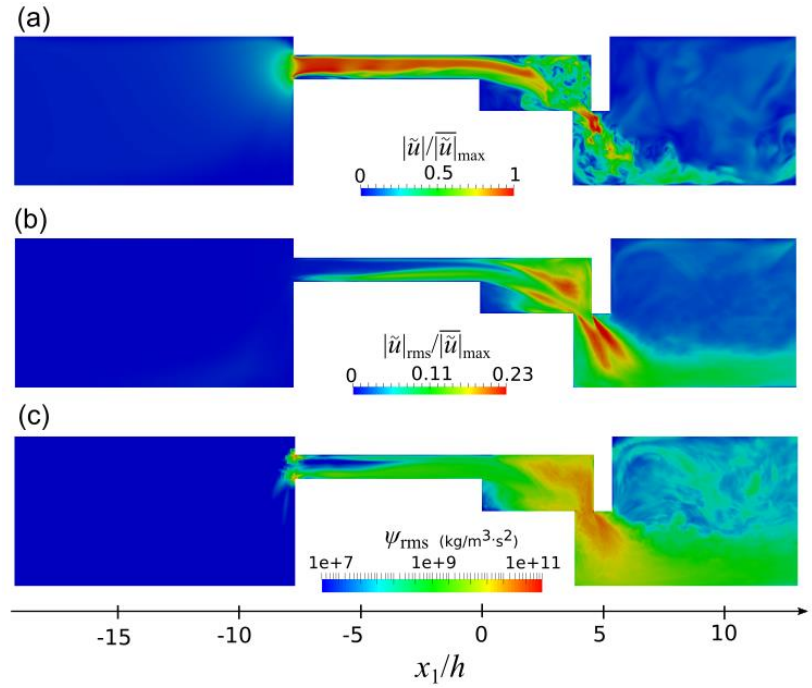


Fig. 4.18 Contour of the flow and source magnitude in the vertical (x_1–x_2) plane at the center x_3 = 0 when L_C/L_T = 0. Instantaneous velocity magnitude (a), RMS of the velocity fluctuation (b), and RMS of the sound source (c) are shown. Magnitude of the source is plotted in log scale.

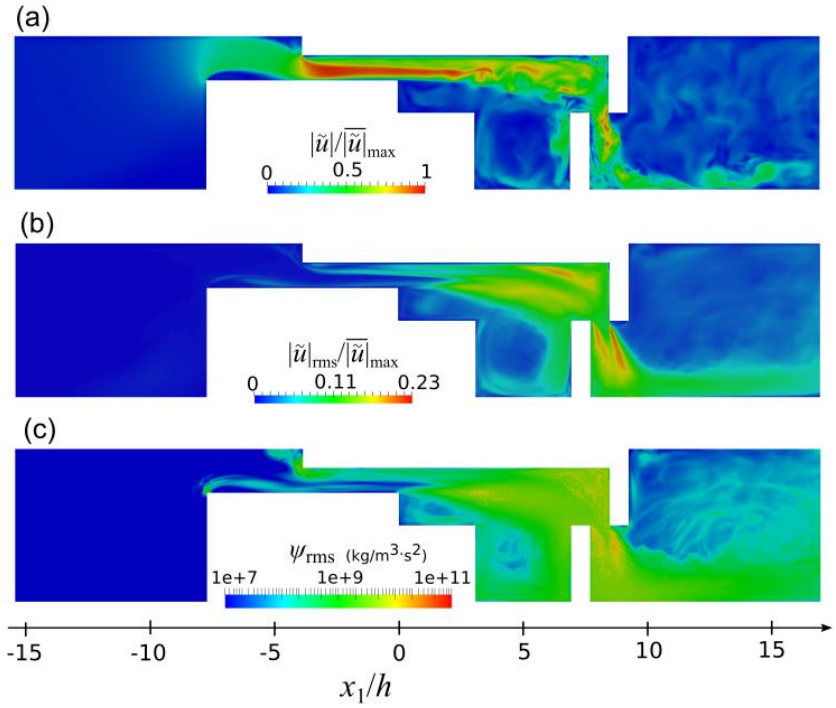


Fig. 4.19 Contour of the flow and source magnitude in the vertical (x_1–x_2) plane at the center x_3 = 0 when L_C/L_T = 1.25. Instantaneous velocity magnitude (a), RMS of the velocity fluctuation (b), and RMS of the sound source (c) are shown. Magnitude of the source is plotted in log scale.

To identify the flow configuration in the region where the sound source was generated, we calculated the second invariant of the velocity gradient tensor, $q = \|\mathbf{\Omega}\|^2 - \|\mathbf{S}\|^2$, where $\mathbf{\Omega}$ and $\mathbf{S}$ are the anti-symmetric and symmetric parts of the velocity gradient tensor, respectively. The regions where $q > 0$ represent vortex tubes. Iso-surfaces for $q/(|\overline{\tilde{u}}|_{\max}/h)^2 = 2.67$ in the simplified model are shown in Fig. 4.20. When $L_C/L_T = 0$, fine-scale vortices were generated in the region where the jet flow impinged on the upper teeth wall ($x_1/h = 4$). In contrast, when $L_C/L_T = 1.25$, coarser vortices were distributed in the region from the constriction inlet to lip cavity ($-4 < x_1/h < 10$). These differences in the flow configuration caused the differences in the source amplitude and distribution that were observed in Fig. 4.18 (c) and Fig. 4.19 (c).

To compare the amplitude in each frequency component, the spatial mean values of the sound source ψ_Ω are plotted along the x_1-axis in Fig. 4.21. For both $L_C/L_T = 0$ and 1.25, the maximum magnitude was observed at the gap between the teeth ($3.8 < x_1/h < 5.4$ for $L_C/L_T = 0$, $7.7 < x_1/h < 9.2$ for $L_C/L_T = 1.25$) in all frequencies. Meanwhile, the magnitudes near the exit of the constriction ($0 < x_1/h < 3.8$ for $L_C/L_T = 0$, $0 < x_1/h < 7.7$ for $L_C/L_T = 1.25$) and lip cavity ($x_1/h > 5.4$ for $L_C/L_T = 0$, $x_1/h < 9.2$ for $L_C/L_T = 1.25$) decreased with increments of the frequency. Large magnitudes appeared at frequencies of 8.2 kHz, 10.8

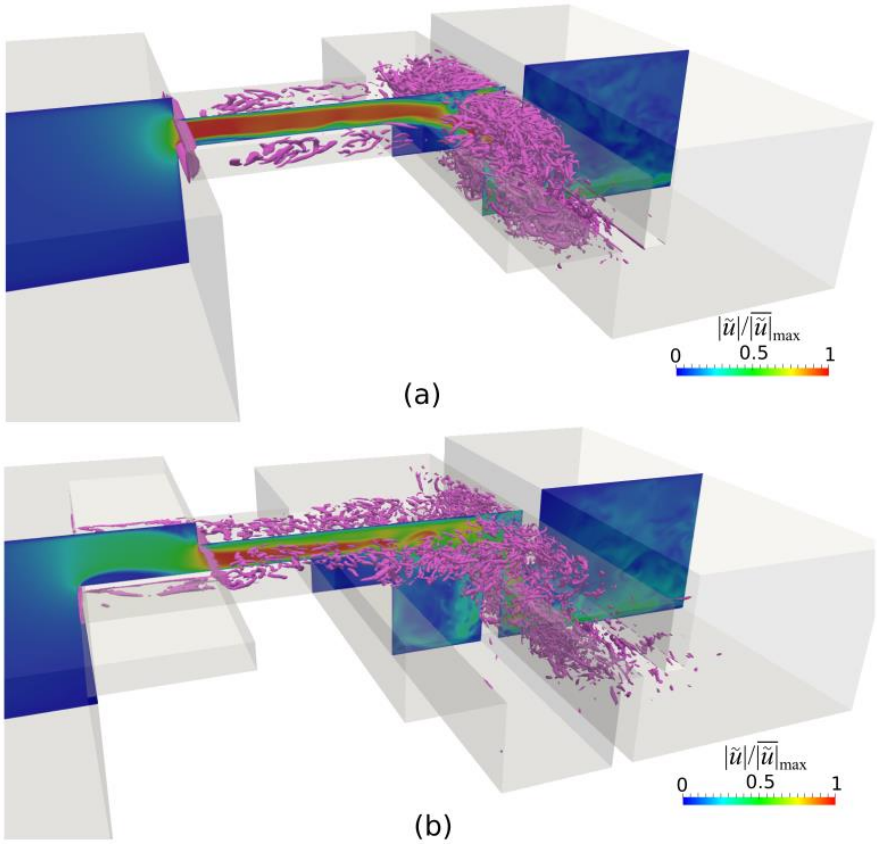


Fig. 4.20 Iso-surfaces of the second invariant $q/(|\overline{\tilde{u}}|_{\max}/h)^2 = 2.67$ in the simplified model when $L_C/L_T = 0$ (a) and $L_C/L_T = 1.25$ (b). The contour of the instantaneous flow magnitude in the vertical (x_1–x_2) plane at the center $x_3 = 0$ is also shown in the model.

kHz, and 12.9 kHz within the constriction (-7.7 < x_1/h < 0) for L_C/L_T = 0, whereas large magnitude appeared at a frequency of 12 kHz upstream from the constriction (-7.7 < x_1/h < -3.9) for L_C/L_T = 1.25. The distinctive periodic fluctuation of the sound source, which causes the characteristic peaks of the sibilant fricatives discussed in the previous section, was not observed for both cases.

The spatial mean values of the pressure field SPL_Ω are plotted along the x_1-axis in Fig. 4.22. Large amplitudes were observed in the frequency range 4.7–9.3 kHz at the channel downstream from the constriction (x_1/h > -5) for L_C/L_T = 0, whereas large amplitudes were observed in the frequency range 3.5–7 kHz downstream from the constriction (x_1/h > -5) for L_C/L_T = 1.25. The large amplitudes at 8.2 kHz and 11 kHz when L_C/L_T = 0 and 12 kHz when L_C/L_T = 1.25 were caused by the periodic source fluctuation in the constriction observed in Fig. 4.21. In contrast, the large amplitudes around 4.7 kHz for L_C/L_T = 0 and around 3.5 kHz for L_C/L_T = 1.25, which correspond to the characteristic peak frequencies of the far-field sound spectrum in Fig. 4.16, were not observed in the source magnitude of Fig. 4.21.

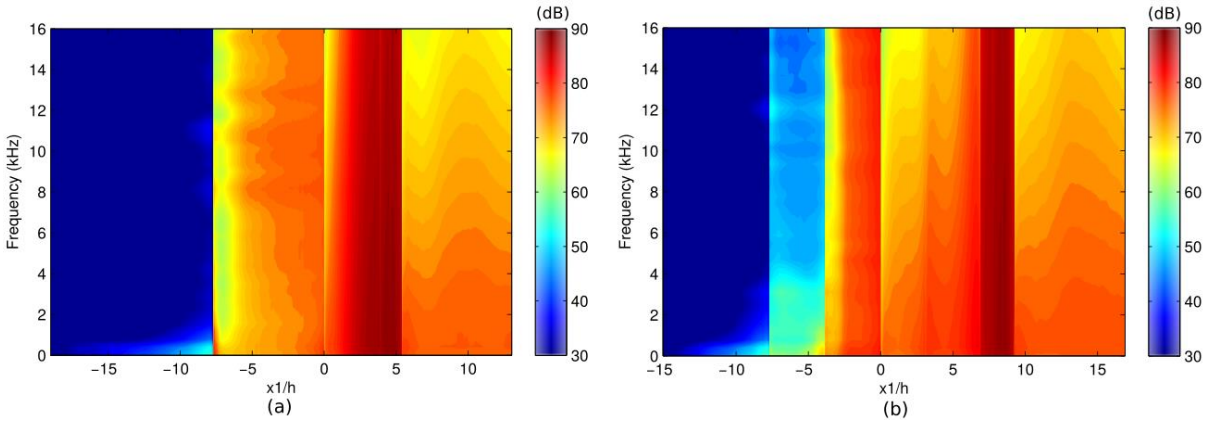


Fig. 4.21 Spatial mean of the sound source ψ_Ω in the transverse (x_2–x_3) plane. (a) L_C/L_T = 0, and (b) L_C/L_T = 1.25. Color bar shows magnitude of sound source along the x_1-axis at each frequency.

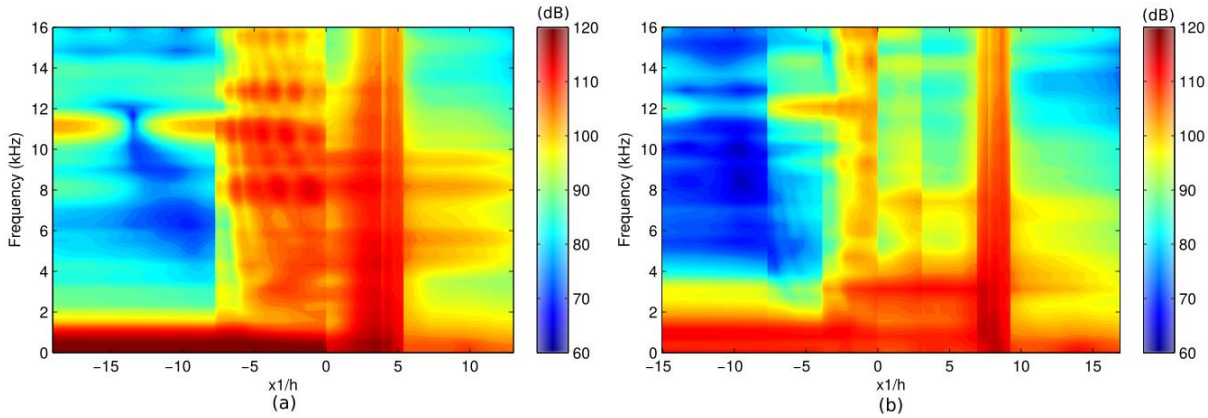


Fig. 4.22 Spatial mean of SPL_Ω in the transverse (x_2–x_3) plane. (a) L_C/L_T = 0, and (b) L_C/L_T = 1.25. Color bar shows the amplitude of SPL along the x_1-axis at each frequency.

4.3.2 Realistic Vocal Tract Geometry

The instantaneous velocity fields simulated in the realistic vocal tract geometry of /s/ and /ʃ/ are shown in Fig. 4.23. In the vocal tract of /s/, the jet flow left from the constriction directly impinged on the upper teeth and was disturbed near the lower lip surface. The disturbed flow left for the upper side of the lip cavity. In the vocal tract of /ʃ/, the jet flow left from the constriction was disturbed in the cavity between the tongue and lower teeth. Then, the jet accelerated at the gap between teeth and left for the lower side of the lip cavity. Since the constriction size of /ʃ/ was wider than that of /s/, the maximum velocity magnitude in the vocal tract of /ʃ/ was lower than the maximum of /s/.

The mean velocity, RMS of velocity fluctuation, and RMS of sound source on the mid-sagittal plane of the realistic geometry of /s/ and /ʃ/ are shown in Fig. 4.24. The maximum mean velocity at the constriction was $|\overline{\tilde{u}}|_{\max}$ = 50.1 m/s in the vocal tract of /s/ and $|\overline{\tilde{u}}|_{\max}$ = 33.4 m/s in the vocal tract of /ʃ/. The cores of the jet flow appeared at the constriction and the gap between teeth in the vocal tract of /ʃ/, whereas the core appeared only at the constriction in the vocal tract of /s/. In the vocal tract of /ʃ/, large RMS values of velocity fluctuation were observed in the mixing layer of the jet flow. Meanwhile, the large RMS values of the velocity fluctuation occurred by the impingement of the jet on the lip surface in the vocal tract of /s/. The large magnitude of the sound source appeared from the gap between teeth to the middle of the lip cavity in /s/ whereas the large magnitude of the source appeared from the cavity between tongue and teeth to the lip cavity.

To examine the directivity pattern of the propagating sound from the realistic geometry in this simulation, the pressure amplitudes in the frequency range 6.3 to 9.8 kHz are depicted in Fig. 4.25. In the same way shown in section 2.2.1, the pressure amplitude in the sagittal plane was larger at the center of the lips than at the above and below of the lip center (the angle around 0° and 180° in section 2.2.1). The directivity was increased with the increment of the frequency in the same way as observed in the experiment. However, the large troughs observed in the experiment with lips at 6.1 kHz (Fig. 2.16) did not appear in the far-field region of the simulation.

The iso-surfaces of the second invariant of velocity gradient tensor in the vocal tract are shown in Fig. 4.26. Since the second invariant consists of an anti-symmetric part which represents the rotational movement and a symmetric part which represents the parallel movement, positive values of the second invariant indicate the vortex configuration. From the comparison of the vortex configuration between /s/ and /ʃ/, we observed that the flow disturbed mainly at the gap between teeth in /s/ whereas the flow disturbed at the cavity between tongue and teeth in /ʃ/.

The velocity field and iso-surfaces of pressure $\bar{p} = 10^5$ are shown in Fig. 4.27. To visualize the 3D velocity configuration from the side view, the volume rendering on the velocity above 2 m/s was calculated by the software ParaView 5.0.1 (https://www.paraview.org). By visualizing the pressure by the iso-surfaces, the sound propagating from the vocal tract can be observed (details: https://youtu.be/qOaH9ssZCcc). The difference in a state of the sound propagation between the vocal tract of /s/ and /ʃ/ was not observed from the videos.

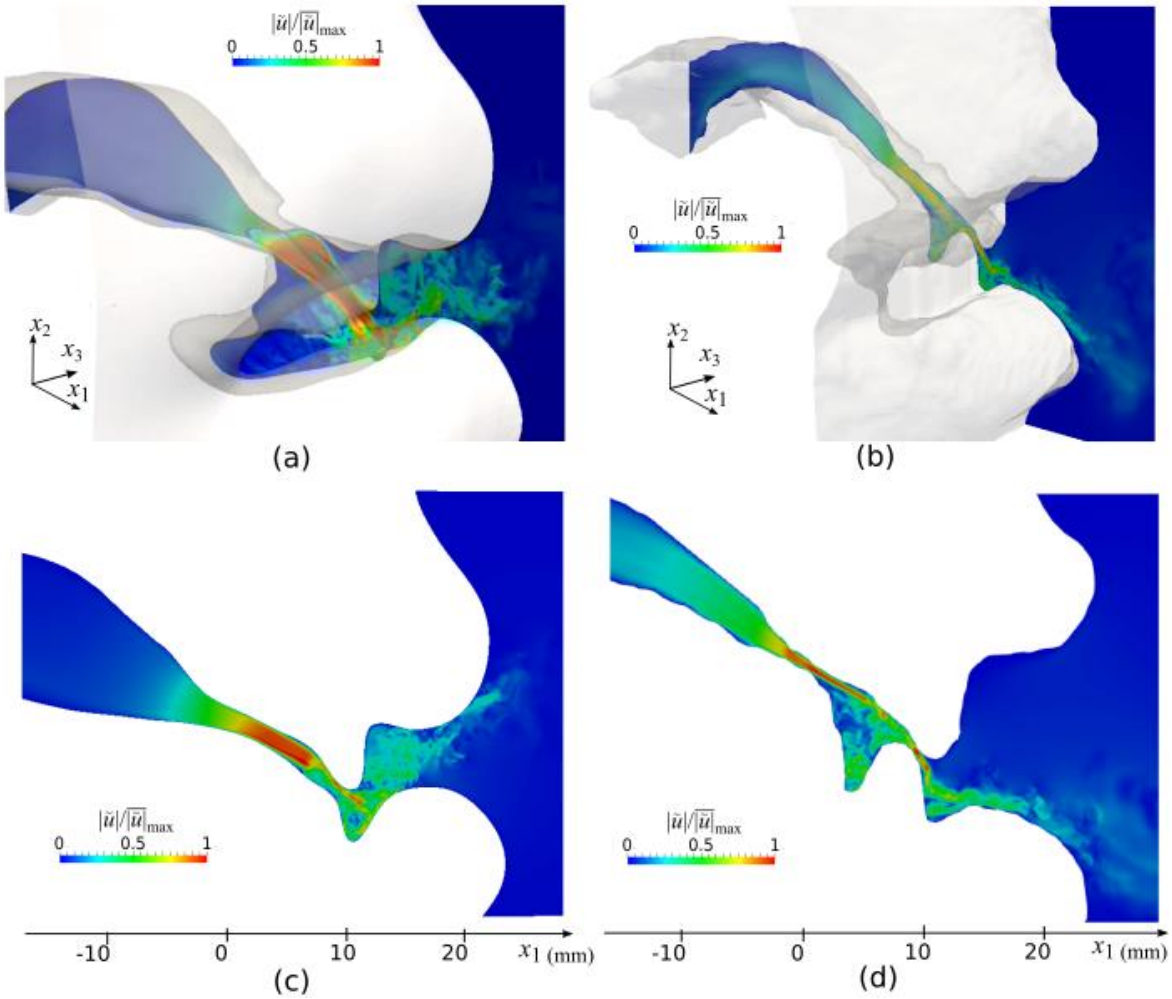


Fig. 4.23 Three-dimensional instantaneous velocity distributions (a-b) and mid-sagittal plane of instantaneous velocity field (c-d) in the realistic vocal tract geometry of /s/ and /ʃ/, respectively.

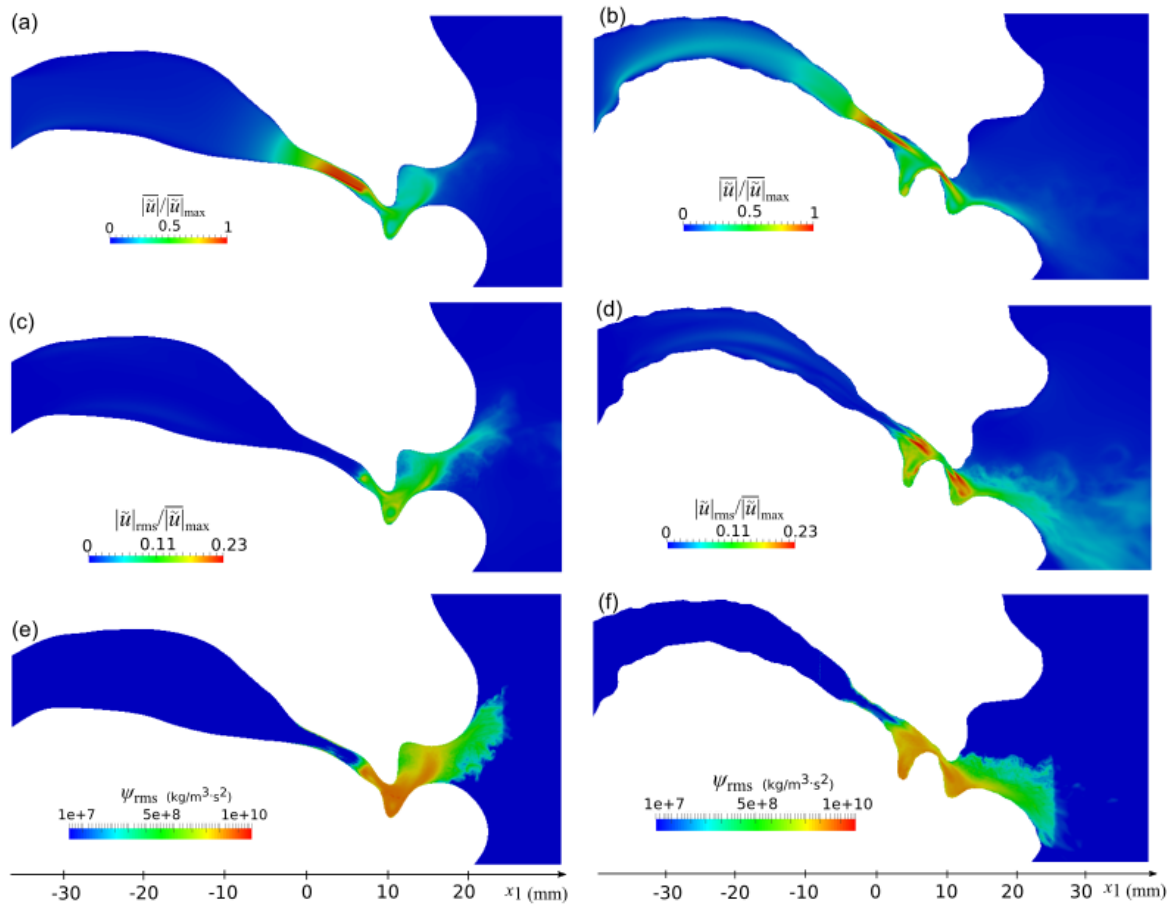


Fig. 4.24 Mean velocity (a-b), RMS of velocity (c-d), and RMS of sound source (e-f) on the mid-sagittal plane of realistic vocal tract geometry /s/ and /ʃ/, respectively.

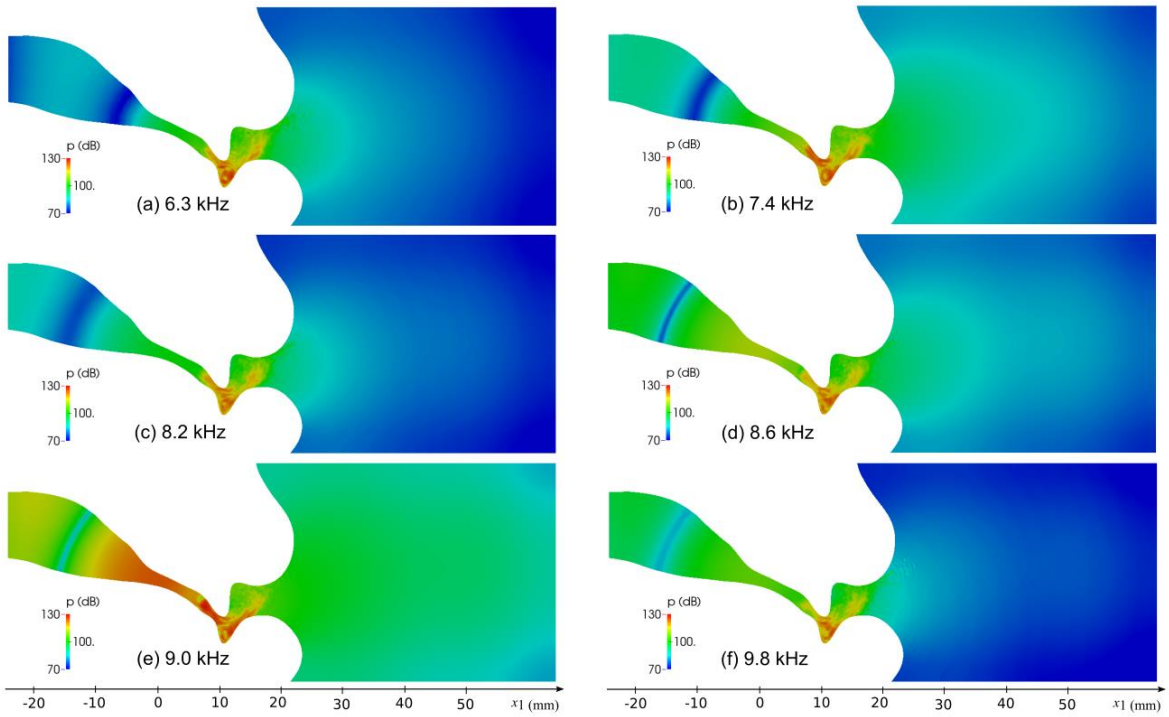


Fig. 4.25 Pressure amplitude inside and outside of the realistic vocal tract geometry for /s/.

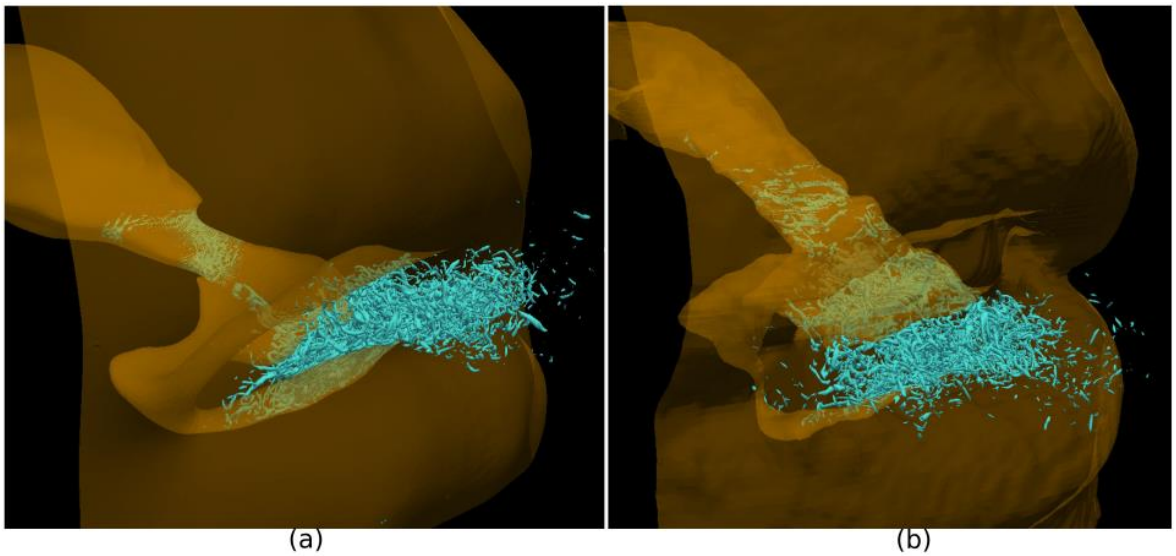


Fig. 4.26 Iso-surfaces of second invariant in the vocal tract of /s/ (a) and /ʃ/ (b).

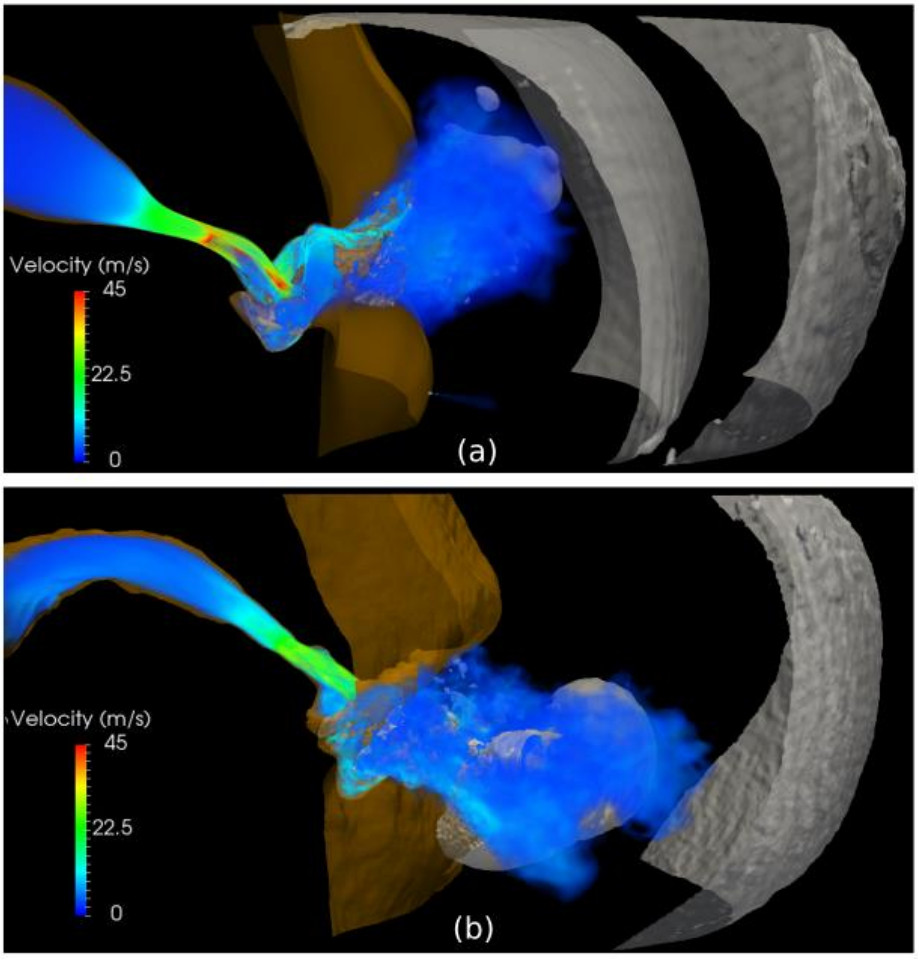


Fig. 4.27 Velocity fields and iso-surface of pressure (white) for the vocal tract of /s/ (a) and /ʃ/ (b).

4.3.3 Relationship between Simplified and Realistic Geometry

Instantaneous velocity magnitudes in the mid-sagittal plane of the realistic geometry of /s/ and the simplified geometry with the tongue position L_C/L_T = 0 are shown in Fig. 4.28. The velocity magnitude was normalized by the maximum mean velocity $\overline{|\tilde{u}|}_{\max}$= 50.1 m/s for the realistic geometry or $\overline{|\tilde{u}|}_{\max}$ = 43.6 m/s for the simplified geometry. The longitudinal coordinate x_1 was normalized by the length L = 10 mm between the maximum constriction (or the center of the constriction) and the lower teeth. In the realistic geometry of /s/, the maximum velocity appeared at the middle of the constriction x_1/L = 0.5, and the jet flow left from the constriction impinged on the upper teeth x_1/L = 0.7. Then, the impinged flow separated from the upper teeth surface x_1/L = 1, and formed a large velocity fluctuation nearby the lower lip surface x_1/L = 1.5. The flow nearby the lower lip surface traveled towards the upper lips and left for the lip cavity. In the simplified geometry, similar flow configuration was observed from the back cavity to the lip cavity.

In order to evaluate the flow field, the maximum mean velocity, RMS of the velocity fluctuation, and RMS of the sound source in the mid-sagittal plane of /s/ are plotted along the axis x_1 in Fig. 4.29. The mean velocity upstream from the constriction rapidly increased in the simplified geometry whereas the mean velocity gradually increased in the realistic geometry. The mean velocity of the flow left from the constriction once decreased near the upper teeth x_1/L = 1, and increased by the separation from the upper teeth surface x_1/L = 1.1, then again decreased towards the lip cavity in both geometries. RMS of the velocity fluctuation increased from the inlet of the constriction, and reached the maximum nearby the upper teeth surface x_1/L = 1, and then decreased towards the lip cavity in both geometry. Meanwhile, RMSs at the constriction inlet x_1/L = 0 and at the upper teeth edge x_1/L = 1.2 in the simplified geometry were larger than those in the realistic geometry. RMS of the sound source showed that the magnitude of the source increased at upstream from the constriction x_1/L = -0.5 and reached the maximum near the upper teeth surface x_1/L = 1, and decreased towards the far-field for the both geometries.

Instantaneous velocity magnitudes in the mid-sagittal plane of the realistic geometry of /ʃ/ and the simplified geometry with the tongue position L_C/L_T = 1.25 are shown in Fig. 4.30. The velocity magnitude was normalized by the maximum mean velocity in the same way as Fig. 4.28. The longitudinal coordinate x_1 was normalized by the length L = 10 mm between the maximum constriction (or the constriction exit) and the lower teeth. In the realistic geometry of /ʃ/, the maximum velocity appeared at the maximum constriction x_1/L = 0, and the jet flow left from the constriction recirculated in the cavity between tongue and lower teeth at x_1/L = 0.4. Then, the disturbed flow at the cavity accelerated at the gap between teeth x_1/L = 0.9, and formed a large velocity fluctuation nearby the lower lip surface x_1/L = 1.2. The flow nearby the lower lip surface traveled along the lower lip surface and left the lip cavity. In the simplified geometry, similar flow configuration was observed from the back cavity to the lower lip surface.

The maximum mean velocity, RMS of the velocity fluctuation, and RMS of the sound source in the mid-sagittal plane of /ʃ/ are plotted along the axis x_1 in Fig. 4.31. The mean velocity upstream from the constriction rapidly increased in both the realistic and

simplified geometries. The mean velocity of the flow left from the constriction once decreased near the lower teeth x_1/L = 0.5, and increased by the gap between teeth at x_1/L = 0.8 in the realistic geometry and at x_1/L = 1 in the simplified geometry. Then, the mean velocity decreased towards the lip cavity in the both geometries. RMS of the velocity fluctuation increased from the inlet of the constriction, and reached the maximum nearby the upper teeth surface x_1/L = 1, and then decreased towards the lip cavity in the both geometries. RMS of the sound source showed that the magnitude of the source increased at upstream from the constriction x_1/L = -0.5 and reached the maximum near the upper teeth surface x_1/L = 1, and decreased towards the far-field for the both geometries. The amplitude of the sound source in the vocal tract of /ʃ/ was smaller than that in the vocal tract of /s/.

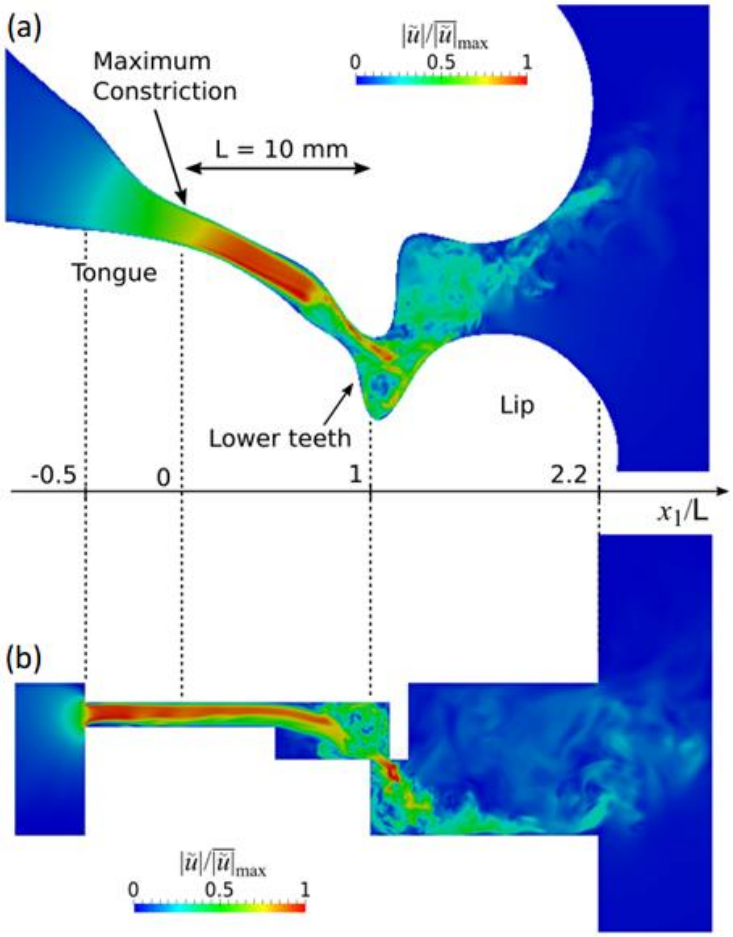


Fig. 4.28 Mid-sagittal planes of instantaneous flow velocity field in the realistic geometry of /s/ (a) and the simplified geometry L_C/L_T = 0 (b).

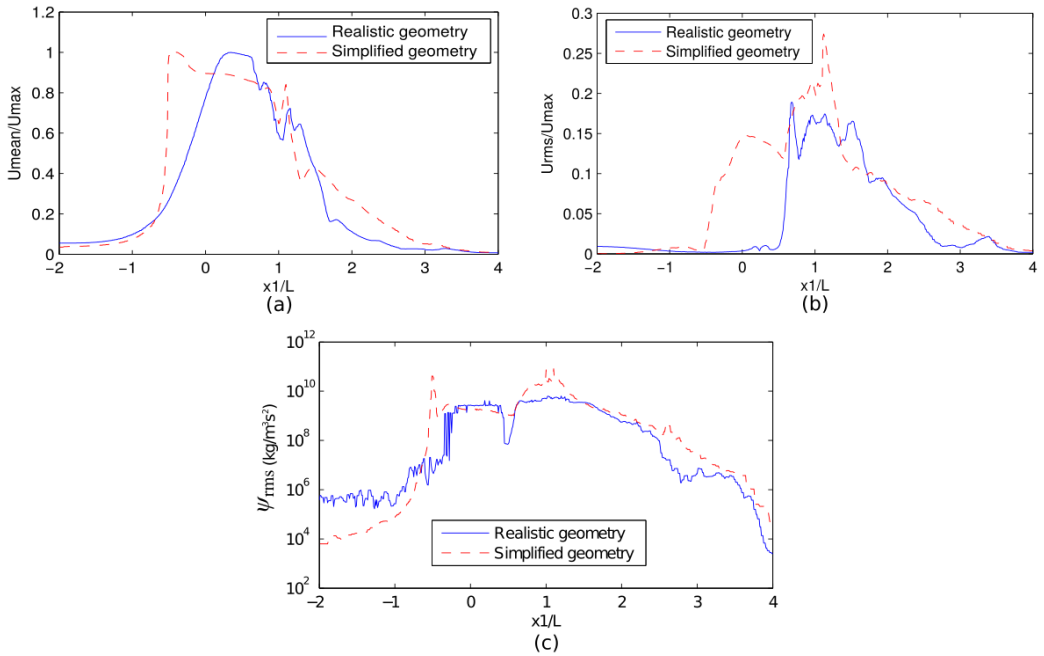


Fig. 4.29 The maximum mean velocity (a), RMS of velocity (b), and RMS of sound source (c) in the mid-sagittal plane of the realistic geometry of /s/ and simplified geometry L_C/L_T = 0. The maximum values are plotted along the axis x_1.

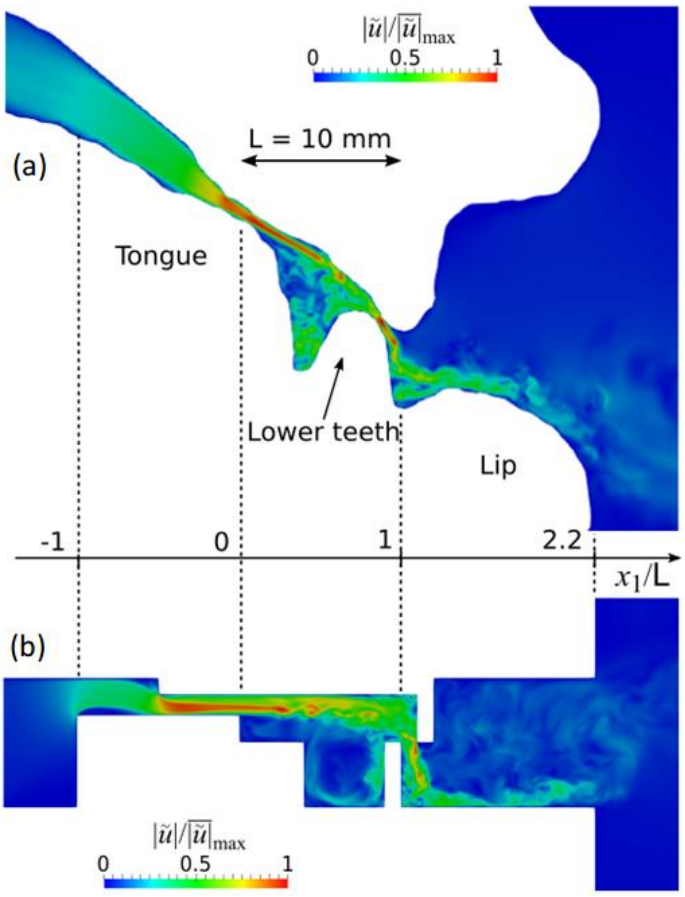


Fig. 4.30 Mid-sagittal planes of instantaneous flow velocity field in the realistic geometry of /ʃ/ (a) and the simplified geometry L_C/L_T = 1.25 (b).

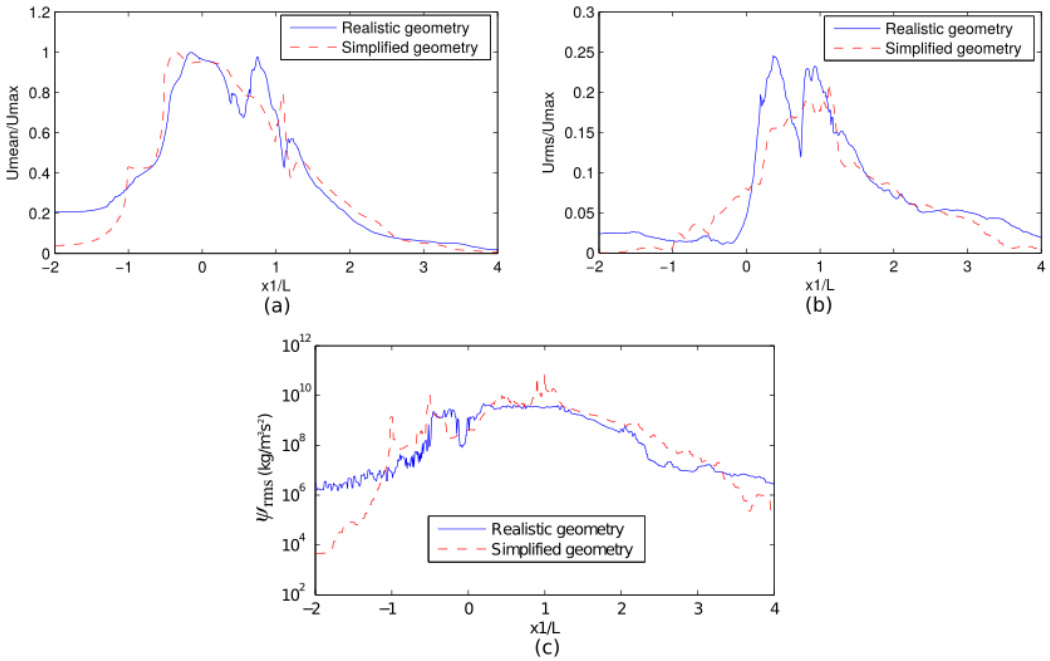


Fig. 4.31 The maximum mean velocity (a), RMS of velocity (b), and RMS of sound source (c) in the mid-sagittal plane of the realistic geometry of /ʃ/ and simplified geometry L_C/L_T = 1.25. The maximum values are plotted along the axis x_1.

4.4 Discussion

4.4.1 Simplified Vocal Tract Geometry

The RMS of the velocity fluctuation and the sound source in the simplified vocal tract geometry (Fig. 4.18 and 4.19) showed that the large values appeared near the upper and lower teeth ($3.8 < x_1/h < 5.4$) for $L_C/L_T = 0$, whereas the large values appeared above the cavity between the tongue and teeth ($x_1/h = 3.1$) and downstream from the gap between the teeth ($x_1/h = 8.5$) for $L_C/L_T = 1.25$. These results indicate that the large velocity disturbance of the jet flow at the gap between the teeth generates the main sound source of /s/ and at downstream from the constriction generates another sound source for /ʃ/. This is consistent with the assumption made in previous studies (Stevens, 1971; Shadle, 1985; Howe and McGowan, 2005). The magnitude of the source for $L_C/L_T = 0$ was larger than that for $L_C/L_T = 1.25$, especially near the upper and lower teeth wall. This result suggests that the larger source magnitude when $L_C/L_T = 0$ causes the larger magnitude of SPL in the spectrum of $L_C/L_T = 0$ compared with $L_C/L_T = 1.25$ in the frequency range above 4 kHz shown in Fig. 4.16 and 4.17.

The spatial mean of the sound source (Fig. 4.21) showed that the large magnitudes appeared at frequencies 8.2 kHz, 10.8 kHz, and 12.9 kHz within the constriction for $L_C/L_T = 0$, whereas the large magnitude appeared at frequency 12 kHz upstream from the constriction for $L_C/L_T = 1.25$. These periodic fluctuations were caused by the periodic vortexes generated in the small recirculating region near the inlet of the constriction. Yokoyama and Kato (2009) reported that a tonal noise with a Strouhal number 0.8 (calculated as $=fL/U$ where f is frequency of the tonal noise, L is the cavity length, and U is free-stream velocity) was produced by fluid-acoustic interactions in a cavity with depth-to-length ratio 0.5 and free stream of Mach number 0.3. By considering that the free stream velocity is the mean velocity at the constriction, the tonal noise is expected to be produced at 10.3 kHz when $L_C/L_T = 0$ and at 9.8 kHz when $L_C/L_T = 1.25$. These frequencies roughly matched the peaks that we observed in the source fluctuation (Fig. 4.21), indicating that the fluid-acoustic interaction occurred in the region between the constriction and downstream cavities.

In contrast, the large amplitudes around 4.7 kHz for $L_C/L_T = 0$ and around 3.5 kHz for $L_C/L_T = 1.25$, which correspond to the characteristic peak frequencies of the far-field sound spectrum in Fig. 4.16, were not observed in the source magnitude. This indicates that the pressure amplitude inside the simplified geometry was increased by the acoustic resonance in the frequency range of the characteristic peaks.

Considering the resonance length in the model geometry, 1/4 wavelengths of the frequencies 3.5 kHz and 4.7 kHz are 24.6 mm ($x_1/h = 18.9$) and 18.3 mm ($x_1/h = 14.1$), respectively, and these wavelengths match the distances between the large amplitude region in the lip cavity ($x_1/h = 9$ for $L_C/L_T = 0$, $x_1/h = 14$ for $L_C/L_T = 1.25$) and small amplitude regions that appeared near the constriction inlet ($x_1/h = -5$). The 1/4 wavelength and pressure contour in the simplified geometries are shown in Fig. 4.32. These results indicate that the small amplitude region near the constriction inlet and the large amplitude region in the lip cavity became the acoustic node and antinode, respectively, and this

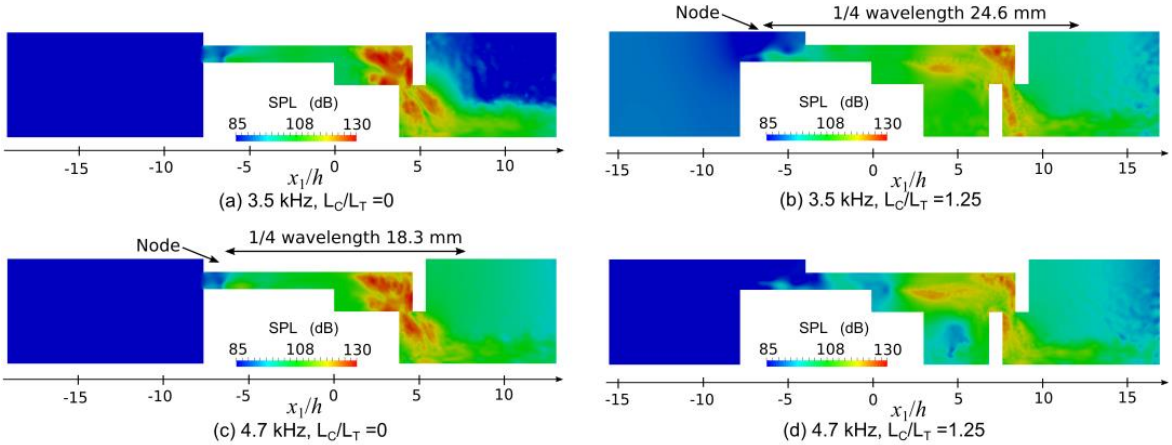


Fig. 4.32 Pressure amplitudes on vertical (x_1 – x_2) plane at the center x_3 = 0. The amplitudes at frequency 3.5 kHz (a-b), and 4.7 kHz (c-d) are shown for L_C/L_T = 0 and 1.25, respectively. The 1/4 wavelengths of corresponding frequencies are shown above the contour plot.

acoustic resonance increased the amplitude at the characteristic peak frequencies for L_C/L_T = 0 and 1.25. Consequently, by shifting the tongue position and changing the distance between the acoustic node and antinode inside the model, the characteristic peak frequency varied and the frequency characteristics of the generated sound changed from /s/ to /ʃ/.

4.4.2 Realistic Vocal Tract Geometry

The flow in the lip cavity left along the upper lip surface in the vocal tract of /s/ whereas the flow left along the lower lip surface of /ʃ/. This difference is probably caused by the lip shape of the subject during the measurement of CT images. Since /ʃ/ was pronounce in the word context, the lip was protruded forward compared to the sustained /s/. Thus, the flow impinged on the lower lip surface left along the lower lip in the vocal tract of /ʃ/ whereas the flow impinged on the lower lip surface of /s/ was deflected and left along the upper lip surface. For future work, it is of interest to investigate the effect of flow direction in the lip cavity on the sound directivity patterns.

The amplitude of the sound source in the vocal tract of /ʃ/ was smaller than that in the vocal tract of /s/. This is caused by the difference in the maximum velocity magnitudes at the constriction. Since the width of the constriction of /ʃ/ is larger than the width of /s/, the maximum velocity as well as the RMS of the velocity fluctuation in the vocal tract of /ʃ/ was smaller than those in the vocal tract of /s/.

The surface of the realistic geometry of /ʃ/ was rougher than the surface of /s/ because of the measurement precision of the CT scan. However, the difference of the sound between /s/ and /ʃ/ was produced by the simulation of the direct method. This indicates that the difference in the position of the acoustic nodes and antinodes in the pressure amplitude observed in the simplified geometry (Fig. 4.32) was also generated in the realistic vocal tract geometries of /s/ and /ʃ/. Meanwhile, the overall amplitude of the generated sound was larger than the measured sound. This indicates that the magnitude of the source fluctuation was increased by the rough surface in the simulation.

4.4.3 Relationship between Simplified and Realistic Geometry

The flow configuration on the vertical plane at the center of the simplified geometry with $L_C/L_T = 0$ was similar to the flow configuration observed in the mid-sagittal plane of the realistic geometry constructed based on the subject pronouncing /s/. Moreover, the flow configuration on the vertical plane of the simplified geometry with $L_C/L_T = 1.25$ was similar to the flow configuration observed on the mid-sagittal plane of the realistic geometry for the subject pronouncing /ʃ/. These results suggest that the simplified vocal tract geometry reproduced the source generation mechanisms inside the vocal tract as well as the far-field sound of both /s/ and /ʃ/.

The RMSs at the constriction inlet $x_1/L = 0$ and at the upper teeth edge $x_1/L = 1.2$ in the simplified geometry when $L_C/L_T = 0$ were larger than those in the realistic geometry of /s/. In contrast, the overall velocity distribution in the realistic geometry of /ʃ/ was reproduced by the simplified geometry when $L_C/L_T = 1.25$. This is because the flow channel upstream of the constriction was gradually narrowed near the maximum constriction by changing the tongue position backward. By gradually changing the dimension of flow channel, the flow configuration in the simplified geometry became similar to the flow in the realistic geometries. In contrast, the differences between the simplified geometry with $L_C/L_T = 0$ and the realistic geometry of /s/ were caused by the sharp edges of the tongue dorsum in the simplified geometry. By rounding the edges at the constriction inlet and the upper and lower teeth, further agreement on the flow configuration might be achieved.

The RMS of the sound source showed that the magnitude of the source increased at upstream from the constriction $x_1/L = -0.5$ and reached the maximum near the upper teeth surface $x_1/L = 1$, and decreased towards the far-field for the both geometries. These results indicate that the accelerated flow in the constriction impinged on the upper teeth surface and produces the maximum magnitude of the source at the separation region near the upper teeth surface in the both geometries. The comparison of the flow inside the geometries showed that the source generation mechanisms in the realistic geometry were well represented by the simplified vocal tract geometry. Moreover, these results are consistent with the assumptions made in the previous simplified models (Shadle, 1985; Howe and McGowan, 2005). In future work, it is desirable to construct the simplified vocal tract geometries for the other subjects pronouncing sibilant fricatives and investigate the difference of the flow and acoustic properties in the different people. Then, it is expected to construct the generally normalized simplified model and understand the fundamental phenomena of the sibilant fricative production in all languages.

4.5 Summary

The aeroacoustic mechanisms of the sibilant fricatives were numerically investigated using the simplified and realistic vocal tract geometries. The flow and acoustic fields predicted by the direct method considering the LES of compressible flow were validated by the experimental measurement. The spectral shapes predicted by the simplified geometry with L_C/L_T = 0 and 1.25 matched those of /s/ and /ʃ/ pronounced by the subject. The direct method on the simplified geometry showed that the large magnitude of the velocity fluctuation and its source appeared near the upper and lower teeth walls when L_C/L_T = 0, whereas it appeared above the tongue tip and downstream from the gap between the teeth when L_C/L_T = 1.25. The magnitude of the source was decreased by increasing L_C/L_T from 0 to 1.25 because the flow decelerated in the cavity between the tongue and teeth.

The magnitudes of the sound source decreased with increments of the frequency downstream from the constriction, and periodic fluctuation of the source was not observed in the frequency range of the characteristic peak frequencies. In contrast, SPL_Ω downstream from the constriction was increased at 4.7 kHz for L_C/L_T = 0 and at 3.5 kHz for L_C/L_T = 1.25. These results indicate that the characteristic peaks of the generated sounds are caused mainly by the acoustic resonance downstream from the constriction. Although the narrow flow channel is occupied by the turbulent flow, we found that the different frequency characteristics between /s/ and /ʃ/ were produced by changing the tongue position and the acoustic node position formed by the constricted channel.

The flow and acoustic fields in the simplified geometry were compared with those in the realistic geometries of /s/ and /ʃ/. Results suggested that the simplified vocal tract model reproduced the source generation mechanisms inside the vocal tract as well as the far-field sound of both /s/ and /ʃ/. By rounding the edges at the constriction inlet and the upper and lower teeth, further agreement on the flow and acoustic fields might be achieved.

Chapter 5. Conclusion and Perspective

The central goal of this thesis has been to understand the aeroacoustic mechanisms of the sibilant fricative production. The approaches of working in three domains: the mechanical replicas, the theoretical modeling, and the numerical simulations have allowed us to examine the physical phenomena in an appropriate context. The mechanical replicas and experimental measurements enabled to simplify the vocal tract geometry and explore the fundamental geometry for the pronunciation of /s/ and /ʃ/. The theoretical modeling allowed us to capture the nodes and antinodes in the acoustic pressure distribution inside the vocal tract as well as the amplitude and frequency of the characteristic peaks observed in the subject's /s/. The numerical simulation enabled to capture the source distribution and a role of the acoustic resonance on the frequency characteristics of /s/ and /ʃ/. The summary of results and merits obtained from the three domains is depicted in Fig. 5.1 with the table of previous studies. Further detailed conclusion and perspective for further work are presented below.

5.1 Summary of Results and Conclusion

In Chapter 2, the experimental measurements were conducted by using the realistic and simplified replicas. By using the realistic replica, acoustic pressure distribution patterns were measured on a vocal tract replica of sibilant /s/ with and without lips. It was found that complex pressure patterns with differences in amplitude of approximately 10 dB occur with and without lips for frequencies above 4 kHz. The lip horn enhances the pressure amplitude up to 15 dB at the center of the lips in both transverse and sagittal plane in the frequency range above 5 kHz. These tendencies were observed in the near-field and

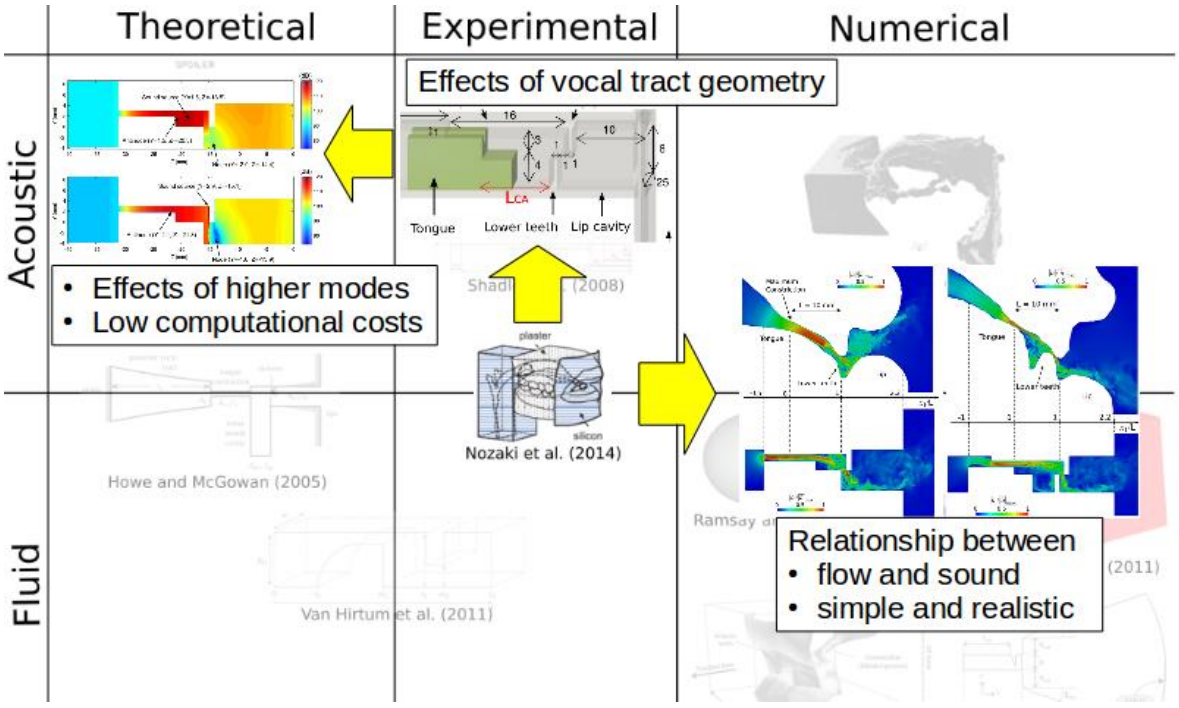


Fig. 5.1 Summary of results and merits for each approach.

far-field measurements with the acoustic source, and in the measurements with flow supply. The comparison between the near-field and far-field measurements showed that more precise directivity pattern can be achieved by the near-field measurement compared to the far-field measurement. The comparison between the compression driver and flow source showed that the pressure distribution pattern is affected by the vocal tract geometry rather than by the source characteristics. The presented experimental results motivate further studies involving spatially detailed directivity pattern measurements for different phonemes using the vocal tract replicas in combination with an acoustic source in order to further study the effect of the lip horn as well as to study the acoustic pressure patterns for different phoneme geometries. Furthermore, the perceptual relevance of these findings needs to be further investigated.

By using the simplified replica, the effects of tongue position as well as the tongue shapes on the acoustic properties were assessed. The simplified replica reproduced the change in main peak frequency of /s/ and /ʃ/ by changing the tongue position L_{CA} = 4 to 6 and removing the tongue tip, whereas the simplified replica reproduced the change in OASPL of /s/ and /ʃ/ by widening the constriction width. These geometries were consistent with the dimensions estimated in the subjects CT images, indicating that the dimensions of the replica with tongue model 1 at L_{CA} = 4 mm and tongue model 4 at L_{CA} = 6 mm are physiologically reasonable and represent the geometric features of the vocal tract pronouncing /s/ and /ʃ/, respectively. In future work, based on the effect of geometrical factors observed in this experiment, the tongue shapes and positions for the different subjects' sustained /s/ and /ʃ/ can be predicted through employing the mechanical experiments.

In Chapter 3, the 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. For the experimental validation, the acoustic driver and flow supply were applied to the inlet of the simplified geometry. 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 were 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 (Howe and McGowan, 2005). 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/ (Jesus and Shadle, 2002) and previous simplified model (Shadle, 1985), 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 positions of the node and antinode were 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/.

In Chapter 4, the aeroacoustic mechanisms of the sibilant fricatives were numerically investigated using the simplified and realistic vocal tract geometries. The flow and acoustic fields predicted by the direct method considering the LES of compressible flow were validated by the experimental measurement. The spectral shapes predicted by the simplified geometry with L_C/L_T = 0 and 1.25 matched those of /s/ and /ʃ/ pronounced by the subject. The direct method on the simplified geometry showed that the large magnitude of the velocity fluctuation and its source appeared near the upper and lower teeth walls when L_C/L_T = 0, whereas it appeared above the tongue tip and downstream from the gap between the teeth when L_C/L_T = 1.25. The magnitude of the source was decreased by increasing L_C/L_T from 0 to 1.25 because the flow decelerated in the cavity between the tongue and teeth.

The magnitudes of the sound source decreased with increments of the frequency downstream from the constriction, and periodic fluctuation of the source was not observed in the frequency range of the characteristic peak frequencies. In contrast, the special mean of the pressure field downstream from the constriction was increased at 4.7 kHz for L_C/L_T = 0 and at 3.5 kHz for L_C/L_T = 1.25. These results indicate that the characteristic peaks of the generated sounds are caused mainly by the acoustic resonance downstream from the constriction. Although the narrow flow channel is occupied by the turbulent flow, we found that the different frequency characteristics between /s/ and /ʃ/ were produced by changing the tongue position and the acoustic node position formed by the constricted channel.

The flow and acoustic fields in the simplified geometry were compared with those in the realistic geometries of /s/ and /ʃ/. Results suggested that the simplified vocal tract model reproduced the source generation mechanisms inside the vocal tract as well as the far-field sound of both /s/ and /ʃ/. By rounding the edges at the constriction inlet and upper and lower teeth, further agreement on flow and acoustic fields might be achieved.

In conclusion, the aeroacoustic mechanisms of the sibilant fricative production can be described by three factors: the jet flow configuration, the source position, and the acoustic resonance in the vocal tract. The jet flow configuration determines the position and amplitude of the sound source. The source position and the acoustic resonance affect the frequency characteristics of generated sound. Therefore, the vocal tract geometry has to be formed considering the interactions among those three factors to produce the sibilant fricatives.

5.2 Perspective for Future Work

In the current study, the aeroacoustic mechanisms of the sibilant fricative production were investigated by using the realistic and simplified vocal tract geometries. First of all procedure, the flow velocity distribution in the actual vocal tract of /s/ was observed by constructing the realistic replica (Nozaki *et al.*, 2014). However, we could not explain the aeroacoustic mechanisms in the vocal tract only using the realistic replica and the experimental measurement. Thus, we constructed the simplified replica by using the knowledge obtained from the realistic replica. By changing the position and dimensions of the tongue model, the frequency characteristics of /s/ and /ʃ/ were reproduced. Moreover, the multimodal modeling and the aeroacoustic simulations on the simplified geometry enable to observe the aeroacoustic phenomena in the vocal tract. Through the analysis on the simplified geometry, the phenomena observed in the realistic geometry (by the both experiment and simulation) could be explained in detail. This process suggests that the feedback loop of results between the realistic and simplified models (Fig. 5.2) was the most important part to understand the aeroacoustic mechanisms of sibilant fricative production. In the next step, it is expected to construct the models for different subjects and understand the fundamental phenomena of the sibilant fricative production in all languages.

In addition, further understanding of the speech production of the sibilant fricatives in word contexts can be expected. In the word context, the fricative consonants are pronounced using co-articulation between the vowel and consonant production. For this problem, the inverse analysis of the tongue muscle stress was proposed by our group (Koike *et al.*, 2017) to estimate the muscle activation in the pronunciation (Fig. 5.3). Through this analysis, the vocal tract geometry during the word pronunciation can be estimated, and further analysis of the aeroacoustic mechanisms in the word pronunciation can be analyzed by combining the estimated geometry and aeroacoustic simulations.

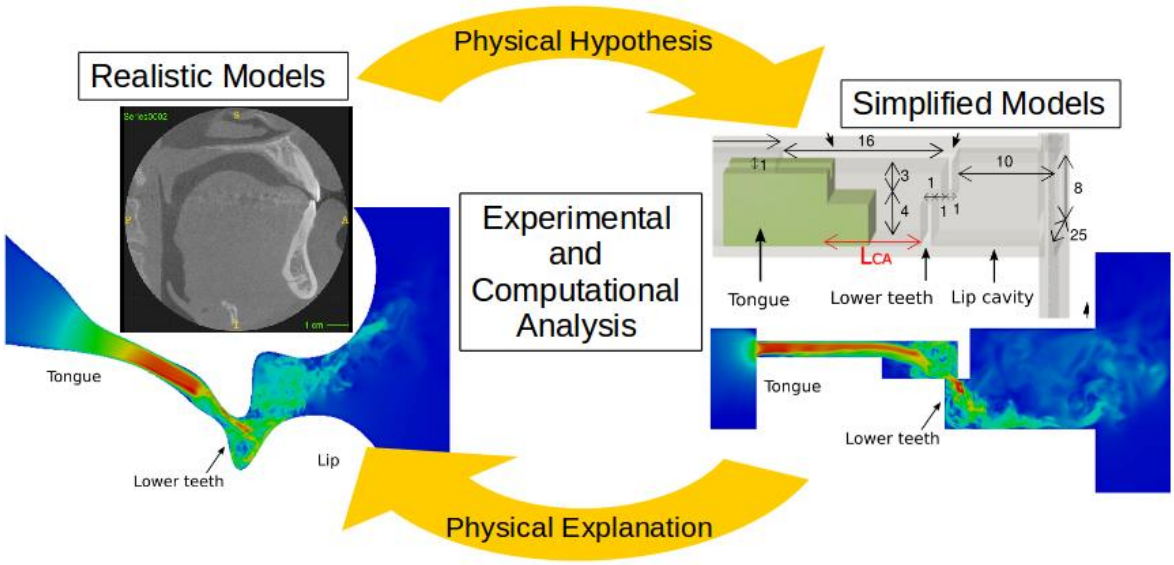


Fig. 5.2 Feedback loop between the realistic and simplified models.

And also, by applying the simulation to jaw bones of ancient people, origin of speech ability can be examined in the field of Evolutionary Linguistics.

Moreover, visualization of the aeroacoustic phenomena using the large-scale visualization system is expected to help the understanding of the speech production and support the speech therapy. By using the large-scale visualization system and 3D glasses, the turbulent flow structure in the vocal tract can be observed in a 3D field and discussed with many people including speech therapists and dentists. The state of a meeting with the visualization system projecting the detailed vortex structure of Fig. 4.23 is shown in Fig. 5.4. Further development of the visualization system for the speech production as well as the supports for the speech therapy using the computational analysis of aeroacoustic mechanisms are expected in future work.

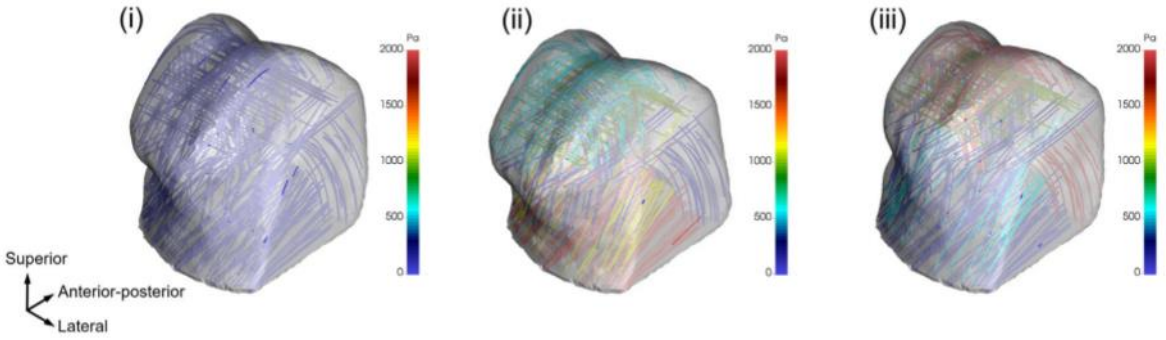


Fig. 5.3 Reference shape of the tongue model (i) and estimated shape and muscle contraction stresses for the forward protrusion (ii) and upward bending (iii) (Koike *et al*., 2017).

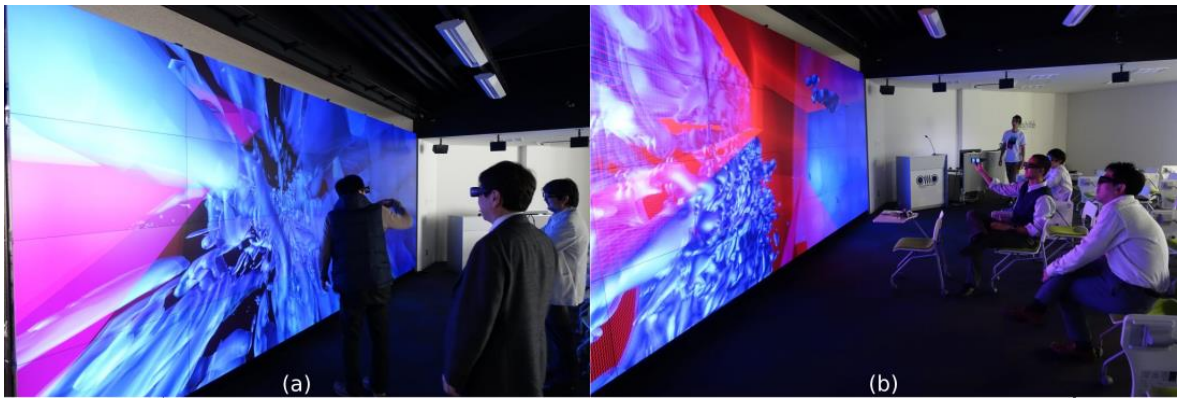


Fig. 5.4 Visualization using large-scale visualization system. Red indicates the vocal tract geometry and blue indicates the vortex structure (http://vis.cmc.osaka-u.ac.jp).

Acknowledgements

It is great pleasure to acknowledge those who helped me in a variety of ways at the end of writing this thesis.

First, I would like to thank Prof. Shigeo Wada for teaching and advising me patiently and thoughtfully. His guidance helped me in all the time of the research, taking a fellowship, and writing of this thesis. He generously shared his extraordinary insights and enthusiasm, while encouraging independence of ideas.

I am deeply grateful to Dr. Kazunori Nozaki, for helping and advising me from the beginning. Discussions with him always gave me productive and stimulating ideas. I could not have imagined having a better advisor and mentor for my study.

I would like to thank Dr. Annemie Van Hirtum for teaching and helping me in various ways from the experiments to the theoretical works in Grenoble. She was always available for discussing how to formulate the ideas, and giving me so many thoughtful suggestions.

I would like to acknowledge my thesis committee, Prof. Kazuyasu Sugiyama and Prof. Shinji Deguchi for their insightful comments and suggestions.

I wish to thank Dr. Satoshi Ii for giving me many helpful advices about the computational fluid simulation and usage of the super computer systems. And also, I would like to thank for making me nice coffee.

I would like to thank Dr. Kenichiro Koshiyama for helping me how to write and think as a physicist. He gave me a really critical view and was always helpful for me to discuss about the ideas.

Many thanks to the members of Wada Laboratry, especially Saptra P. Gabriel, Takuya Imamura, and Narihiro Koike for the stimulating discussions, for the days we were working before deadlines, and for all the fun we had. Thanks for giving me a nice environment to continue the study and finish this thesis.

Special thanks to Prof. Akiyoshi Iida and Prof. Chisachi Kato for giving me a lot of advices for the large-scale computational analysis, especially how to use the software Front Flow/blue.

I would like to thank Prof. Shinji Shimojo, Dr. Kensuke Yasufuku, Dr. Yoshiyuki Kido, and Cyber Media Center, Osaka University for the usage of the large-scale visualization systems.

This work was supported by a JSPS Grant-in-Aid for JSPS Research Fellows (Grant number: JP15J00413), a Grant-in-Aid for Scientific Research (Grant number: T15K013660), MEXT as “Priority Issue on Post-K computer” (Project ID: hp160218, hp170265), and the Program for Leading Graduate Schools of MEXT (Humanware Innovation Program).

Appendix A

Examples of Sound Spectra of Japanese Sibilant Fricatives

Figure A1 shows the examples of sound spectra of /s/ and /ʃ/ for five Japanese male subjects. The fricatives are sustained in the word contexts /usui/ or /mʲisoʃirɯ/ (misoshiru), and sounds are measured 30 cm from the lips. Details of experimental setups are described in section 2.1.4.

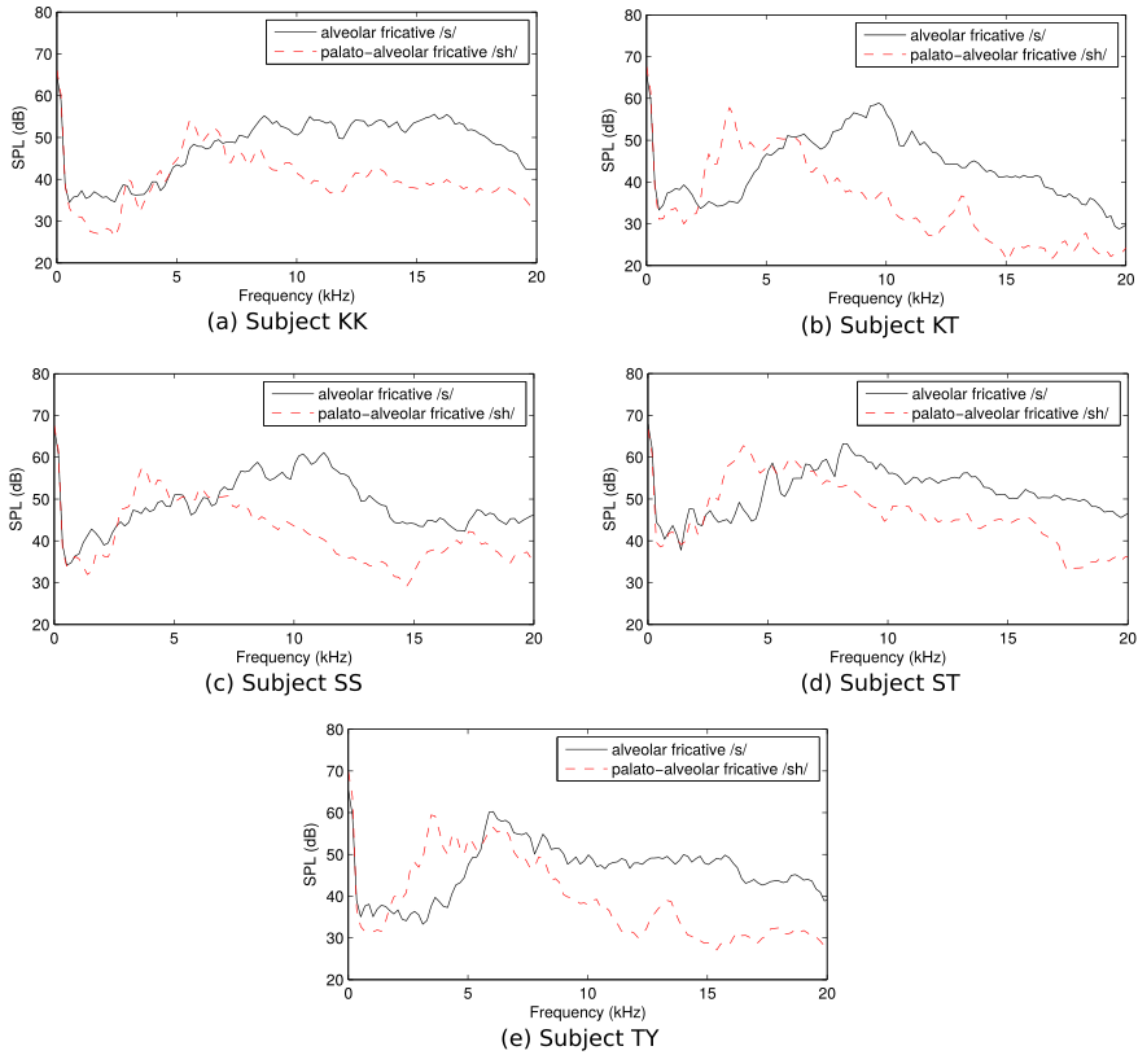


Fig. A1. Examples of sound spectra of Japanese sustained sibilant fricatives.

Appendix B

Validation of the Experimental Setup and Simulation Condition

To confirm that the experimental room was sufficiently quiet, we compared the spectrum of the back ground noise (BGN) with that of the sound generated by the model (Fig. 14). The high SPL values around 172 Hz were probably caused by noise in the data acquisition system. We also confirmed that the noise was not generated by the signal sampling, using the anti-aliasing filter (low-pass filter 22.4 kHz). Since SPLs above 1 kHz were less than 25 dB, and the differences between BGN and characteristic peak amplitudes were over 40 dB, we conclude that the experimental room and measurement equipment were sufficiently quiet for measuring sibilant fricatives.

To confirm that the flow and acoustic fields were developed in the simulation, the number of iterations was increased from 13.5×10^4 to 17.5×10^4 (time from 0.016 s to 0.02 s) when $L_C/L_T = 0$. The spectra predicted by the simulations are shown in Fig. 15. The amplitude of the sound did not significantly increase, and the maximum discrepancy was less than 1 dB in the frequency range 0–16 kHz. Therefore, the flow and acoustic fields were developed in the frequency range in which sibilant fricatives occur. The amplitude at 391 Hz was slightly lower for the 0.003–0.02 s period (59.7 dB) than for the 0.003–0.016 s period (60.6 dB). This indicates that amplitudes for frequency below 2 kHz will probably decrease if the sampling time is increased up to 1 s.

To confirm that boundary reflection was negligible, we examined a movie of the pressure fluctuations ($\bar{p} - \bar{p}_{\mathrm{Mean}}$) in the time series (Fig. 16). The frame rate of the movie was 10 fps and each frame was sampled at 100 kHz in the simulation. Thus, the total duration of the movie was 0.001 s in simulation time (10 s in real time). The movie showed that large pressure amplitude propagated mainly around 10 kHz through the far-field region, and the reflection was almost negligible at frequencies above 1 kHz.

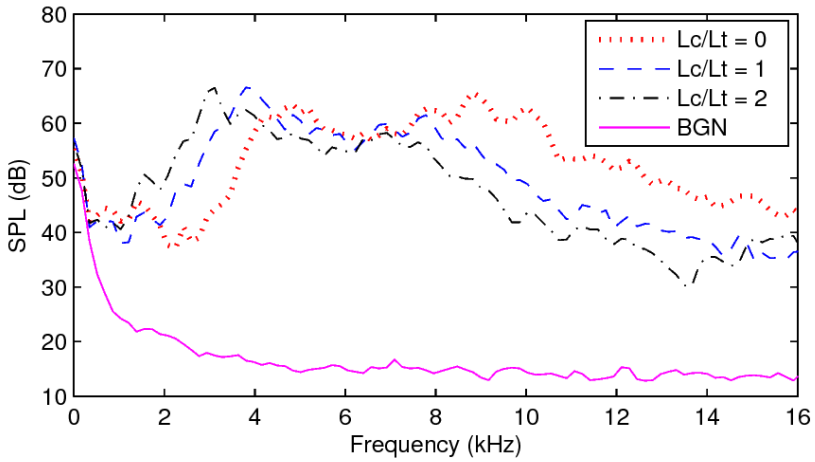


Fig. B1. Spectra of sound measured at 90 mm from the model and back ground noise (BGN).

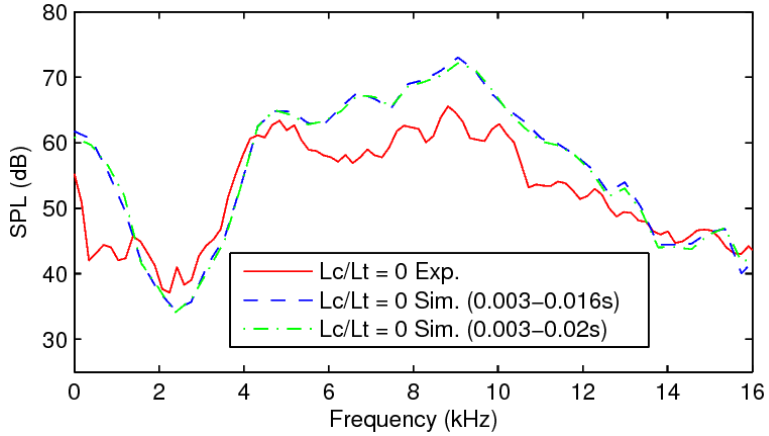


Fig. B2. Spectra of sound pressure at x_1/h = 82.3 when L_C/L_T = 0. The spectrum predicted in the 0.003–0.016 s period was compared with that predicted in the 0.003–0.02 s period.

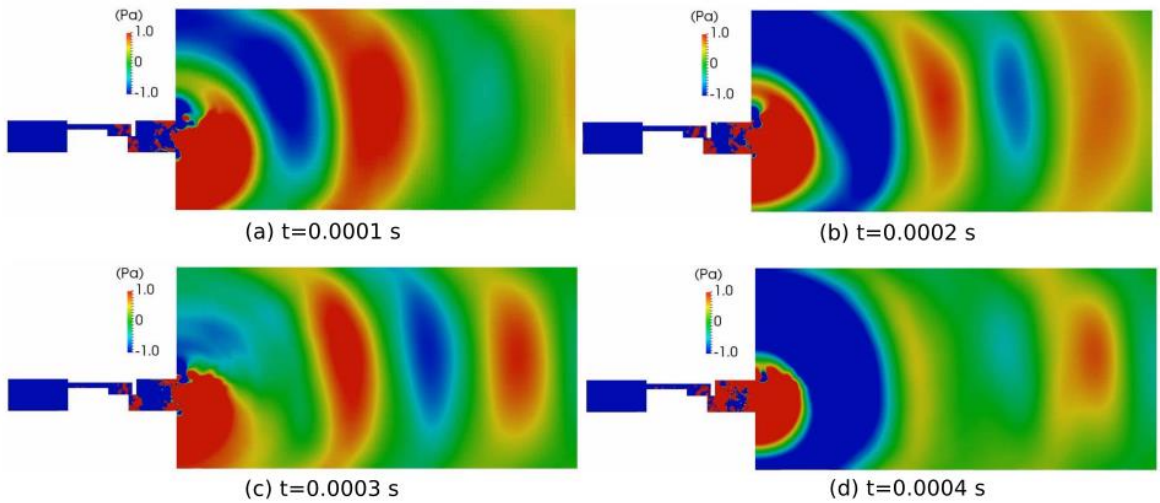


Fig. B3. The pressure fluctuation ($\bar{p} - \bar{p}_{\mathrm{Mean}}$) in time series when L_C/L_T = 0.

References

Amir, N., Pagneux, V., and Kergomard, J. (1997) A study of wave propagation in varying cross-section waveguides by modal decomposition. Part II. Results. *Journal of the Acoustical Society of America* 101 2504–2517.

Badin, P. (1991). Fricative consonants-acoustic and x-ray measurements. *Journal of phonetics*, 19(3-4), 397-408.

Blandin, R., Arnela, M., Laboissière, R., Pelorson, X., Guasch, O., Hirtum, A. V., and Laval, X. (2015). Effects of higher order propagation modes in vocal tract like geometries. *The Journal of the Acoustical Society of America*, 137(2), 832-843.

Blandin, R., Van Hirtum, A., Pelorson, X., and Laboissière, R. (2016). Influence of Higher Order Acoustical Propagation Modes on Variable Section Waveguide Directivity: Application to Vowel [α]. *Acta Acustica united with Acustica*, 102(5), 918-929.

Buchaillard, S., Perrier, P., and Payan, Y. (2009). A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America*, 126(4), 2033-2051.

Cabrera, D., Davis, P. J., and Connolly, A. (2011). Long-term horizontal vocal directivity of opera singers: Effects of singing projection and acoustic environment. *Journal of Voice*, 25(6), e291-e303.

Cisonni, J., Nozaki, K., Van Hirtum, A., and Wada, S. (2011). A parameterized geometric model of the oral tract for aero acoustic simulation of fricatives. *International Journal of Information and Electronics Engineering*, 1(3), 223-228.

Cisonni, J., Nozaki, K., Van Hirtum, A., Grandchamp, X., and Wada, S. (2013). Numerical simulation of the influence of the orifice aperture on the flow around a teeth-shaped obstacle. *Fluid Dynamics Research*, 45(2), 025505.

Curle, N. (1955). The influence of solid boundaries upon aerodynamic sound. *Proceedings of the Royal Society A* 231, 505-514.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton, The Hague, Paris, 328 pp.

Favre, A. (1969). Statistical Equations of turbulent gases. *Problems of Hydrodynamics and Continuum Mechanics*, SIAM, Philadelphia, pp. 231-266.

Fujiso, Y., Nozaki, K., and Van Hirtum, A. (2015). Estimation of minimum oral tract constriction area in sibilant fricatives from aerodynamic data. *The Journal of the Acoustical Society of America*, 138(1), EL20-EL25.

Fureby, C. (1996). On subgrid scale modeling in large eddy simulations of compressible fluid flow. *Physics of Fluids*, 8(5), 1301-1311.

Geoghegan, P. H., Spence, C., Ho, W. H., Jermy, M., Hunter, P., and Cater, J. E. (2012). Stereoscopic PIV measurement of airflow in human speech during pronunciation of fricatives. *Proceedings of the 16th International Symposium of Laser Techniques to Fluid Mechanics*.

Germano, M., Piomelli, U., Moin, P., and Cabot, W. H. (1991). A dynamic subgrid-scale eddy viscosity model. *Physics of Fluids A: Fluid Dynamics*, 3(7), 1760-1765.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F. H., Lane, H. and Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America*, 128(5), 3079-3087.

Gloerfelt, X., and Lafon, P. (2008). Direct computation of the noise induced by a turbulent flow through a diaphragm in a duct at low Mach number. *Computers & fluids*, 37(4), 388-401.

Guo, Y., Kato, C., and Yamade, Y. (2006). Basic features of the fluid dynamics simulation software "FrontFlow/Blue". *Seisan Kenkyu*, 58(1), 11-15.

Haley, K. L., Seelinger, E., Mandulak, K. C., and Zajac, D. J. (2010). Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics*, 38(4), 548-554.

Heinz, J. M., and Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 33(5), 589-596.

Howe, M. S. (2003). *Theory of Vortex Sound*, Cambridge University Press.

Howe, M. S. and McGowan, R. S. (2005). Aeroacoustics of [s]. *Proceedings of Royal Society A* 461, 1005-1028.

Ingard, U. (1953). On the theory and design of acoustic resonators. *The Journal of the Acoustical Society of America*, 25(6), 1037-1061.

Iskarous, K., Shadle, C. H., and Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129(2), 944-954.

Jesus, L. M. T., and Shadle, C. H., (2002). A parametric study of the spectral characteristics of european portuguese fricatives. *Journal of Phonetics* 30, 437–464.

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252-1263.

Kato, C., Yamade, Y., Wang, H., Guo, Y., Miyazawa, M., Takaishi, T., Yoshimura, S., and Takano, Y. (2007). Numerical prediction of sound generated from flows with a low Mach number. *Computers & Fluids*, 36(1), 53-68.

Kergomard, J., Garcia, A., Tagui, G., and Dalmont, J. P. (1989). Analysis of higher order mode effects in an expansion chamber using modal theory and equivalent electrical circuits. *Journal of Sound and Vibration*, 129(3), 457-475.

Koike, N., Ii, S., Yoshinaga, T., Nozaki, K., and Wada, S. (2017). Model-based inverse estimation for active contraction stresses of tongue muscles using 3D surface shape in speech production. *Journal of Biomechanics*, 64, 69-76.

Lighthill M. J. (1952). On sound generated aerodynamically. I. General theory. *Proceedings of Royal Society A* 211, 564-587.

Lilly, D. K. (1992). A proposed modification of the Germano subgrid‐scale closure method. *Physics of Fluids A: Fluid Dynamics*, 4(3), 633-635.

McGowan, R. S., and Howe, M. S. (2007). Compact Green's functions extend the acoustic theory of speech production. *Journal of Phonetics*, 35(2), 259-270.

Monson, B. B., Hunter, E. J., and Story, B. H. (2012). Horizontal directivity of low-and high-frequency energy in speech and singing. *The Journal of the Acoustical Society of America*, 132(1), 433-441.

Monson, B. B., Lotto, A. J., and Story, B. H. (2014). Detection of high-frequency energy level changes in speech and singing. *The Journal of the Acoustical Society of America*, 135(1), 400-406.

Motoki, K., Badin, P., Pelorson, X., and Matsuzaki, H. (2000). A modal parametric method for computing acoustic characteristics of three-dimensional vocal tract models. *Proceedings of 5th Seminar on Speech Production: Models and Data*, pp. 325-328.

Motoki, K. (2013). A parametric method of computing acoustic characteristics of simplified three-dimensional vocal-tract model with wall impedance. *Acoustical Science and Technology*, 34(2), 113-122.

Muehleisen, R. T. (1996). *Reflection, radiation, and coupling of higher order modes at discontinuities in finite length rigid walled rectangular ducts*, Ph.D. thesis, Pennsylvania State Univ.

Narayanan, S. S., Alwan, A. A., and Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging, *The Journal of the Acoustical Society of America*, 98(3), 1325-1347.

Narayanan, S., and Alwan, A. (2000). Noise source models for fricative consonants. *IEEE transactions on speech and audio processing*, 8(3), 328-344.

Nozaki, K., Akiyama, T., Tamagawa, H., Kato, S., Mizuno-Matsumoto, Y., Nakagawa, M., Maeda, Y., and Shimojo, S. (2005). The first grid for the oral and maxillofacial region and its application for speech analysis. *Methods of Information in Medicine*, 44(2), 253-256.

Nozaki, K. (2010). Numerical simulation of sibilant [s] using the real geometry of a human vocal tract. *High Performance Computing on Vector Systems 2010*, Springer, Berlin, Heidelberg, 137-148.

Nozaki, K., Nakamura, M., Takimoto, H., and Wada, S. (2012). Effect of expiratory flow rate on the acoustic characteristics of sibilant/s. *Journal of Computational Science*, 3(5), 298-305.

Nozaki, K., Yoshinaga, T., and Wada, S. (2014). Sibilant /s/ simulator based on computer tomography images and dental casts. *Journal of Dental Research* 93(2), 207-211.

Oberai, A. A., Roknaldin, F., and Hughes, T. J. (2000). Computational procedures for determining structural-acoustic response due to hydrodynamic sources. *Computer Methods in Applied Mechanics and Engineering*, 190(3), 345-361.

Pagneux, V., Amir, N., and Kergomard, J. (1996). A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation. *The Journal of the Acoustical Society of America*, 100(4), 2034-2048.

Patgiri, C., Sarma, M., and Sarma, K. K. (2013) Recurrent neural network based approach to recognize assamese fricatives using experimentally derived acoustic-phonetic features. *Proceedings of Emerging Trends and Applications in Computer Science (ICETACS), IEEE*, 33-37.

Patri, J. F., Diard, J., and Perrier, P. (2015). Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biological cybernetics*, 109(6), 611-626.

Perkell, J. S., Boyce, S. E., and Stevens, K. N. (1979). Articulatory and acoustic correlates of the /s-sh/ distinction. *Speech Communication Papers, 97th Meeting of the Acoustical Society of America*, J.J. Wolf and D.H. Klatt eds., pp. 109-113.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination & use of an articulatory saturation effect. *Journal of Speech Language Hearing Research* 47, 1259–1269.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25, 382-407.

Pierce, A. D. (1981). *Acoustics: an introduction to its physical principles and applications*. New York: McGraw-Hill.

Poinsot, T. J. A., and Lelef, S. K. (1992). Boundary conditions for direct simulations of compressible viscous flows. *Journal of Computational Physics*, 101(1), 104-129.

Powell, A. (1964). Theory of vortex sound, *The Journal of the Acoustical Society of America*, 33, 177–195.

Ramsay, G., and Shadle, C. (2006). The influence of geometry on the initiation of turbulence in the vocal tract during the production of fricatives. *Proceedings of 7th International Seminar on Speech Production*, 581-588.

Reidy, P. F. (2016). Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, 140(4), 2518-2529.

Shadle, C. H. (1985). *The acoustics of fricative consonants*. Ph. D thesis, Massachusetts Institute of Technology, Cambridge, MA.

Shadle, C. H., and Scully, C. (1995). An articulatory-acoustic-aerodynamic analysisof [s] in VCV sequences. *Journal of Phonetics*, 23(1), 53-66.

Shadle, C. H., Berezina, M., Proctor, M., and Iskarous, K. (2008). Mechanical models of fricatives based on MRI-derived vocal tract shapes. *Proceedings of 8th International Seminar on Speech Production*, 417-420.

Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2), 1103-1113.

Stavness, I., Gick, B., Derrick, D., and Fels, S. (2012). Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America*, 131(5), EL355-EL360.

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: static considerations. *The Journal of the Acoustical Society of America* 50(4), 1180-1192.

Stevens, K. N. (1998). *Acoustic phonetics*. MIT press.

Takemoto, H., Kitamura, T., Nishimoto, H., and Honda, K. (2004). A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. *Acoustical science and technology*, 25(6), 468-474.

Toda, M., and Honda, K. (2003). An MRI-based cross-linguistic study of sibilant fricatives. *Proceedings of the 6th International Seminar on Speech Production*, 1–6.

Toda, M., Kitamura, T., Honda, K., and Maeda, S. (2003). Vocal tract shape of sibilant fricatives derived from MRI and their acoustic modeling, *Technical Report of IEICE*, 103(219), 7-12 (in Japanese).

Toda, M., and Maeda, S. (2006). Quantal aspects of non anterior sibilant fricatives: a simulation study. *Proceedings of 7th International Seminar on Speech Production*, 573-580.

Van Hirtum, A., Grandchamp, X., and Pelorson, X. (2009) Moderate Reynolds number axisymmetric jet development downstream an extended conical diffuser: Influence of extension length. *European Journal of Mechanics B Fluid*, 28, 753-760.

Van Hirtum, A., Grandchamp, X., Pelorson, X., Nozaki, K., and Shimojo, S. (2010). Les and" in Vitro" Experimental Validation of Flow around a Teeth-Shaped Obstacle. International *Journal of Applied Mechanics*, 2(02), 265-279.

Van Hirtum, A., Pelorson, X., Estienne, O., and Bailliet, H. (2011). Experimental validation of flow models for a rigid vocal tract replica. *The Journal of the Acoustical Society of America* 130(4), 2128-2138.

Westbury, J. (1994). *X-ray Microbeam Speech Production Database User's Handbook* (University of Wisconsin, Madison, WI), pp. 17–27.

Wood, S., Wishart, J., Hardcastle, W., Cleland, J., and Timmins, C. (2009). The use of electropalatography (EPG) in the assessment and treatment of motor speech disorders in children with Down's syndrome: evidence from two case studies. *Developmental Neurorehabilitation*, 12(2), 66-75.

Yokoyama, H., and Kato, C. (2009). Fluid-acoustic interactions in self-sustained oscillations in turbulent cavity flows. I. Fluid-dynamic oscillations. Physics of Fluids, 21(10), 105103.

Yokoyama, H., Miki, A., Onitsuka, H., and Iida, A. (2015). Direct numerical simulation of fluid–acoustic interactions in a recorder with tone holes. *The Journal of the Acoustical Society of America*, 138(2), 858-873.

Zharkova, N. (2016). Ultrasound and acoustic analysis of sibilant fricatives in preadolescents and adults. *The Journal of the Acoustical Society of America*, 139(5), 2342-2351.

List of Publication

Original Papers

1) Yoshinaga, T., Nozaki, K., and Wada, S. Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract model. *Physics of Fluids* (accepted).

2) Yoshinaga, T., Van Hirtum, A., Nozaki, K., and Wada, S. (2018). Influence of the lip horn on acoustic pressure distribution pattern of sibilant /s/. *Acta Acustica united with Acustica*, 104, 145-152.

3) Koike, N., Ii, S., Yoshinaga, T., Nozaki, K., and Wada, S. (2017). Model-based inverse estimation for active contraction stresses of tongue muscles using 3D surface shape in speech production. *Journal of Biomechanics*, 64, 69-76.

4) Yoshinaga, T., Van Hirtum, A., and Wada, S. (2017). Multimodal modeling and validation of simplified vocal tract acoustics for sibilant /s/. *Journal of Sound and Vibration*, 411, 247-259.

5) Yoshinaga, T., Nozaki, K., and Wada, S. (2017). Effect of tongue position in the simplified vocal tract model of sibilant fricatives /s/ and /ʃ/. *The Journal of the Acoustical Society of America*, 141(3), EL314-EL318.

6) Nozaki K., Yoshinaga T., and Wada S., (2014). Sibilant /s/ simulator based on computed tomography images and dental casts. *Journal of Dental Research* 93(2), 207-211.

Conference Proceedings

1) Yoshinaga T., Nozaki K., and Wada S. (2017). A relationship between simplified and realistic vocal tract geometries for Japanese sibilant fricatives. *Proceedings of 11th International Seminar on Speech Production*.

2) Yoshinaga T., Nozaki K., and Wada S. (2017). Effects of tongue position on flow and sound in a simplified vocal tract model of sibilant fricatives. *Proceedings of 5th Japan-Switzerland Workshop on Biomechanics*.

3) Yoshinaga, T., Nozaki, K., and Wada, S. (2016) Effect of tongue position in the simplified vocal tract model of sibilant fricatives /s/ and /ʃ/. *The Journal of the Acoustical Society of America*, 140(4), 3221.

4) Yoshinaga T., Nozaki K., and Wada S. (2016). Experimental validation of sound generated from flow in simplified vocal tract model of sibilant /s/. *Proceedings of INTERSPEECH 2016*, 3584-3587.

5) Yoshinaga, T., Koike, N., Nozaki, K., and Wada, S. (2015). Study on production mechanisms of sibilants using simplified vocal tract model. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 250(1), 5662-5669.

6) Yoshinaga, T., Nozaki, K., and Wada, S. (2014). Effects of tongue elevation on the airflow in oral cavity and the sound generation of sibilant /s/. *Proceedings of 7th World Congress of Biomechanics.*

7) Yoshinaga, T., Nozaki, K., and Wada, S. (2014). Effects of the Position of Tongue on the Sound Generation of Sibilant /s/. *Proceedings of the 15th International Conference on Biomedical Engineering*, 364-367.