

国立国語研究所学術情報リポジトリ

Aeroacoustic differences between the Japanese fricatives [ç] and [ç̚]

メタデータ	言語: English 出版者: 公開日: 2022-10-19 キーワード (Ja): キーワード (En): 作成者: Yoshinaga, Tsukasa, Maekawa, Kikuo, Iida, Akiyoshi メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/3681

This work is licensed under a Creative Commons Attribution 3.0 International License.



Aeroacoustic differences between the Japanese fricatives [ɕ] and [ç]

Tsukasa Yoshinaga, Kikuo Maekawa and Akiyoshi Iida

Citation: *The Journal of the Acoustical Society of America* **149**, 2426 (2021); doi: 10.1121/10.0003936

View online: <https://doi.org/10.1121/10.0003936>

View Table of Contents: <https://asa.scitation.org/toc/jas/149/4>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Global numerical simulation of fluid-structure-acoustic interaction in a single-reed instrument](#)

The Journal of the Acoustical Society of America **149**, 1623 (2021); <https://doi.org/10.1121/10.0003757>

[General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent](#)

The Journal of the Acoustical Society of America **149**, 2602 (2021); <https://doi.org/10.1121/10.0004240>

[Vocal fold kinematics and relative fundamental frequency as a function of obstruent type and speaker age](#)

The Journal of the Acoustical Society of America **149**, 2189 (2021); <https://doi.org/10.1121/10.0003961>

[Consonantal \$F_0\$ perturbation in American English involves multiple mechanisms](#)

The Journal of the Acoustical Society of America **149**, 2877 (2021); <https://doi.org/10.1121/10.0004239>

[Mechanics of human voice production and control](#)

The Journal of the Acoustical Society of America **140**, 2614 (2016); <https://doi.org/10.1121/1.4964509>

[Informational masking with speech-on-speech intelligibility: Pupil response and time-course of learning](#)

The Journal of the Acoustical Society of America **149**, 2353 (2021); <https://doi.org/10.1121/10.0003952>



**Advance your science and career
as a member of the**

ACOUSTICAL SOCIETY OF AMERICA

LEARN MORE



Aeroacoustic differences between the Japanese fricatives [ç] and [ç̥]

Tsukasa Yoshinaga,^{1,a)} Kikuo Maekawa,² and Akiyoshi Iida^{1,b)}

¹Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku, Toyohashi, Aichi 441-8580, Japan

²National Institute for Japanese Language and Linguistics, 10-2 Midoricho, Tachikawa, Tokyo 190-8561, Japan

ABSTRACT:

To elucidate the linguistic similarity between the alveolo-palatal sibilant [ç̥] and palatal non-sibilant [ç] in Japanese, the aeroacoustic differences between the two consonants were explored via experimentation with participants and analysis using simplified vocal tract models. The real-time magnetic resonance imaging (rtMRI) observations of articulatory movements demonstrated that some speakers use a nearly identical place of articulation for /si/ [ç̥i] and /hi/ [çi]. Simplified vocal tract models were then constructed based on the data captured by static MRI, and the model-generated synthetic sounds were compared with speaker data producing [ç̥] and [ç]. Speaker data demonstrated that the amplitude of the broadband noise of [ç̥] was weaker than that of [ç]; the characteristic peak amplitude at approximately 4 kHz was greater in [ç̥] than in [ç], although the mid-sagittal vocal tract profiles were nearly identical for three of ten subjects in the rtMRI observation. These acoustic differences were reproduced by the proposed models, with differences in the width of the coronal plane constriction and the flow rate. The results suggest the need to include constriction width and flow rate as parameters for articulatory phonetic descriptions of speech sounds.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0003936>

(Received 17 September 2020; revised 3 March 2021; accepted 6 March 2021; published online 7 April 2021)

[Editor: Susanne Fuchs]

Pages: 2426–2436

I. INTRODUCTION

This paper aimed to elucidate an important aspect of the phonetic differences between the sibilant and non-sibilant fricatives through the application of articulatory and acoustic analyses, including numerical simulations of simplified vocal tract models. Although not a part of the classification criteria of the International Phonetic Alphabet (IPA), the distinction between sibilant and non-sibilant sounds is largely recognized among researchers in phonetics/phonology and related fields. In distinctive feature theories, such a distinction is often represented by a feature called [\pm STRIDENT] (Jakobson *et al.*, 1952; Chomsky and Halle, 1968). According to Chomsky and Halle, consonants that bear this feature are marked acoustically by greater noisiness than their non-strident counterparts. Some researchers have used this feature to classify consonants in general; however, most researchers restrict its use to the classification of fricatives.

From the perspective of general phonetics, Ladefoged and Maddieson (1996) wrote that “fricative sounds may be the result of turbulence generated at the constriction itself, or they may be due to the high-velocity jet of air formed at a narrow constriction going on to strike the edge of some obstruction such as the teeth. We will call the latter type sibilants” (p. 138). According to Ladefoged and Maddieson

(1996), only coronal consonants can be sibilant. Similarly, in the system of Chomsky and Halle (1968), the feature [STRIDENT] is used to distinguish, for example, a palato-alveolar fricative, such as [ʃ], which is [+STRIDENT], from a palatal fricative, such as [ç], which is [–STRIDENT].

Owing to its importance to the phonetic description of speech sounds and their phonological classification, the articulatory and acoustic bases for the distinction between sibilant and non-sibilant fricatives have been the theme of numerous speech production studies. The articulatory differences causing the acoustic contrast between sibilance and non-sibilance have been investigated by measuring vocal tract configurations by applying medical imaging techniques. For example, one-dimensional area functions have been estimated using x-ray pictures (Badin, 1991), and three-dimensional vocal tract geometries have been reconstructed using magnetic resonance imaging (MRI) (Narayanan *et al.*, 1995). In addition, the detailed geometrical differences among the sibilants [s], [ʃ], [ç̥], and [ç] have been analyzed via MRI (Toda and Honda, 2003; Toda *et al.*, 2010).

The acoustic contrast between sibilants and non-sibilants can be caused by small differences in place of articulation (e.g., the alveolar ridge for [s] and the hard palate for [ç̥]); however, the resulting change in the spectral characteristics covers a wide frequency range of up to 20 kHz because of the nature of turbulence noise (Stevens, 1989). A sound associated with sibilance has been characterized by the formation of a narrow groove of the tongue tip, which is called the “sibilant groove,” as opposed to the other fricative

^{a)}Electronic mail: yoshinaga@me.tut.ac.jp, ORCID: 0000-0002-4655-4186.

^{b)}ORCID: 0000-0003-1370-6144.

consonants. It is believed that changes in the shape of this groove and its jet flow formation coincide with the acoustic contrast between sibilants and non-sibilants (Fletcher, 1989). This nature of fricatives has attracted the attention of numerous researchers (Jesus and Shadle, 2002; Perkell *et al.*, 2004; Reidy, 2016; Zharkova, 2016), but the articulatory–acoustic relationship of fricative production has not been elucidated in detail.

The remainder of this paper focuses on the contrast between the voiceless alveolo-palatal sibilant fricative [ç] and voiceless palatal non-sibilant fricative [ç̟] as they occur in Standard (or Tokyo) Japanese. Canonically, these consonants are contrastive vis-à-vis their places of articulation. However, as Sec. II reveals, such a contrast in place of articulation is nearly lost in some speakers, whereas the perceptual contrast between the produced fricatives is maintained. Thus, the concrete issue under examination in this study is the clarification of the articulatory mechanism by which the acoustic contrast between [ç] and [ç̟] is preserved under nearly identical places of articulation.

The paper is organized in the following manner: Sec. II presents a brief introduction to Japanese phonetics. Section III presents the results for the observations of the places of articulation using real-time magnetic resonance imaging (rtMRI) for two subjects: A and B. Section IV discusses the acoustic analysis of [ç] and [ç̟] of four other subjects—C, D, E, and F—as the sounds recorded for subjects A and B during the rtMRI operation were polluted by the large operational noise of the MRI apparatus. To investigate the cause of the acoustic contrast observed in Sec. IV, models of simplified vocal tract geometry were constructed based on the static three-dimensional MRI data of [ç] by subject C. The results of the aeroacoustic experiment and simulations are presented in Sec. V, and the overall results are discussed in Sec. VI. Finally, Sec. VII concludes the paper.

II. JAPANESE FRICATIVES /SI/ AND /HI/

This section comprises a short digressive introduction to the phonological characteristics of Japanese fricatives, for readers unfamiliar with the Japanese language. The phonology of Tokyo or Standard Japanese includes a skewed inventory of fricatives—three are voiceless (/s/, /h/, and /f/), and one is voiced (/z/). All these fricatives exhibit allophonic variations.

The /s/ is realized as a voiceless alveolar fricative when it is followed by vowels other than /i/ (i.e., [sa], [su], [se], and [so]) but is realized as a voiceless alveolo-palatal fricative [çi] due to the regressive palatalization triggered by the following vowel /i/. The /h/ consonant is more variable. When combined with the five vowels, it is realized either as glottal [ha] (/ha/), [he] (/he/), or [ho] (/ho/); bilabial [ɸu] (/hu/); or palatal [çi] (/hi/). In this case, again, the following vowel /i/ palatalizes the consonant. The phoneme /f/ only appears in loanwords and is invariably realized as a voiceless bilabial fricative [ɸ] (i.e., /fa/ [ɸa], /fi/ [ɸi], /fe/ [ɸe], and /fo/ [ɸo]), except when followed by /u/; /fu/ is realized

as [ɸu] and is identical to /hu/ (see above). The /z/ is realized either as fricative [z] or affricate [dz], depending mostly on the duration of the consonant (Maekawa, 2010), and it is also palatalized before /i/ as [dzi]. In addition to the phonetic palatalization caused by /i/, Japanese also has phonologically palatalized (hence, contrastive) fricatives, including /sj/ (/sja/ [ç̟a], /sju/ [ç̟u], and /sjo/ [ç̟o]), /hj/ (/hja/ [ç̟a], /hju/ [ç̟u], and /hjo/ [ç̟o]), and /zj/ (/zja/ [ç̟a], /zju/ [ç̟u], and /zjo/ [ç̟o]).

This study focuses on the contrast between the fricatives of /si/ [çi], which is sibilant, and that of /hi/ [çi], which is non-sibilant. There is a good reason to posit that in Japanese, the contrast between the two aforementioned fricatives is somewhat unstable. According to the linguistic geographical survey reported in the *Linguistic Atlas of Japan* (National Language Research Institute, 1974), the initial mora of words such as /higasi/ “east” and /hige/ “mustache/whiskers” is often realized as [çi] (i.e., the same sound as /si/) in Tokyo and a wide area of the northeast (*Tōhoku*) region of Japan. Similarly, the initial mora of /sici-gacu/ [çitçigatsu] “July (literally seventh month)” is variably realized as [çi] and [çi] throughout Japan, except in the northeast region, where it is realized as non-palatalized [si].

A similar phonological change is reported in the present-day urban dialect of Berlin, Germany (Jannedy and Weirich, 2014). The instability of the contrast between [ç] and [ç̟] is probably a universal trend, given the small number of languages incorporating this contrast. According to the PHOIBLE 2.0 database, which comprises the phonemic inventories of 2186 distinct languages (Moran and McCloy, 2019), 34 languages demonstrated this contrast.

III. OBSERVATIONS USING REAL-TIME MRI

Figure 1 presents the comparison of the vocal tract constriction for [ç] and [ç̟] in /si/ and /hi/ as uttered by two male speakers of Tokyo Japanese. These snapshots of vocal tracts in the mid-sagittal plane were obtained from the rtMRI articulatory movement database, which is under construction (Maekawa, 2019, 2021).

The articulatory movements were recorded in real time in the mid-sagittal plane by a 3T MRI system (MAGNETOM Prisma fit 3 T, Siemens, Munich, Germany), at the ATR Brain Activity Imaging Center in Kyoto, using a FLASH sequence with the acceleration factor 3. The spatial resolution was set to 256 × 256 pixels with a pixel size of 1 × 1 mm² and a slice thickness of 10 mm. The temporal reconstruction rate was 14 frames/s (fps). The database covered ten Tokyo Japanese speakers, providing speech samples of approximately 30 min each. Note that none of the speakers was aware of the aim of this study. [çi] and [çi] were recorded as parts of a long list of utterances covering the inventory of the 110 Japanese morae and 676 nonsense bimoraic words.

Figure 1 shows the rtMRI recording of the bimoraic nonsense word /hisi/ (usually realized as [çiçi] as a

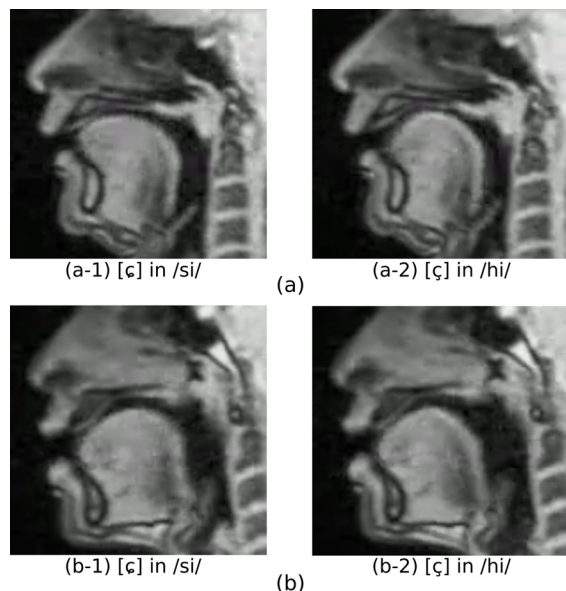


FIG. 1. (Color online) Mid-sagittal images of the vocal tract corresponding to [ç] in /si/ and [ç] in /hi/ in the bimoraic nonsense word /hisi/. Subjects A and B are presented in (a) and (b), respectively.

canonical form) being uttered in the carrier sentence /korega ___ gata/ (“this is the ___ type”). Frames corresponding to [ç] and [ç] were selected via visual and auditory inspection. Although the difference in place of articulation was clearly visible in subject A [Fig. 1(a)], nearly the same place of articulation was used for both fricatives in subject B [Fig. 1(b)]. In the aforementioned rtMRI database, three of ten subjects were found to demonstrate remarkably similar articulation for [ç] and [ç]. Subjects A and B also differed in the shapes of their tongue dorsum. In the articulation of [ç] in /hi/, subject A lifted the tongue dorsum, whereas subject B did not. Moreover, the shape and size of the vocal tract around the teeth were nearly identical between [ç] and [ç] in all the subjects. We examined the mid-sagittal plane of subject B using static three-dimensional MRI data to evaluate the potential deviation of the vocal tract midline from the mid-sagittal plane (Verhoeven *et al.*, 2019). No notable deviation was observed for the subject.

The articulatory variability of [ç] and [ç] seen in Fig. 1 might be one of the causes of the phonetic variability observed among Japanese dialects; however, the speech samples of /si/ and /hi/ produced by subject B in Fig. 1 are perceptually clearly distinct as sibilant and non-sibilant despite the nearly identical mid-sagittal vocal tract profiles.¹ This observation indicates that some articulatory–acoustic factors of speech production have not been fully described, neither by the IPA’s classification nor by research on the acoustic–articulatory relation. One possibility is that the subject changed the constriction width in the coronal plane to form the acoustic contrast between [ç] and [ç] with the same place of articulation. However, we could not observe changes in the constriction width in the coronal plane, as the rtMRI data only provide information on the sagittal plane, excluding information in the coronal plane. To solve this

problem, we constructed simplified models of vocal tract geometry by means of static MRI data and examined the effects of constriction width (see Sec. V).

IV. ACOUSTIC ANALYSIS

A detailed acoustic analysis of [ç] in /si/ and [ç] in /hi/ was conducted using separately recorded speech samples, as the speech recorded during the rtMRI recording session was heavily polluted by the operational noise of the machinery. Samples of [ç] and [ç] were separately recorded in an anechoic chamber (see Sec. V for details). Four male speakers of Standard Japanese pronounced the sustained [ç] in /si/ and [ç] in /hi/ for 3 s; the utterances were recorded using a condenser microphone (type 4939, Bruel & Kjaer, Naerum, Denmark) located 300 mm from the subjects’ lips, with a sampling frequency of 50 kHz. The sound spectrum was calculated using the discretized Fourier transform (DFT) with a 256-point time window, which was multiplied by a Hanning window. The magnitudes of spectra were calculated with 60 windows and an overlap ratio of 30%. The recording was repeated five times for each fricative. All the subjects were unaware that the focus of the recording was the contrast between /si/ and /hi/.

Figure 2 shows the comparison of the spectra of [ç] and [ç] uttered by these four new subjects (C, D, E, and F). In all the subjects, [ç] is characterized by a larger amplitude of the broad noise covering the frequency range of 3–20 kHz in comparison to [ç]. For subject C, the first characteristic peak of [ç] in /si/ was observed at 3.5 kHz, whereas the maximum amplitude was observed at 5.5 kHz. The spectra of [ç] in /hi/ by subject C were accompanied by a sharp resonance peak observed at 3.9 kHz; the amplitudes at higher frequencies were smaller than that of [ç], and the frequency range covered by the broad noise was identical in [ç] and [ç]. This tendency was observed in the other three subjects as well and was similar to the spectral difference observed in the Japanese sibilants [s] and [ç] (Yoshinaga *et al.*, 2017; Yoshinaga *et al.*, 2019). Moreover, the spectrum of Japanese [ç] was characterized by a relatively larger amplitude at approximately 3–4 kHz and a relatively smaller amplitude in the higher frequency range. This decrease in amplitude in the higher frequency range can be explained by the decrease in jet velocity with a change in flow rate or a wider constriction in the coronal plane (Shadle, 1985; Jesus and Shadle, 2002; Yoshinaga *et al.*, 2017). By contrast, however, a sharp peak for [ç] at approximately 4 kHz has not been observed in the literature. Thus, the mechanism by which the 4-kHz peak is derived is one of the main foci of this study. We investigate this topic using simplified tongue models in Sec. V.

V. STUDY WITH A SIMPLIFIED MODEL

A. Simplified model

Simplified vocal tract models of [ç] and [ç] for subject C were constructed based on the three-dimensional vocal tract geometry reconstructed from his MRI data. Using the

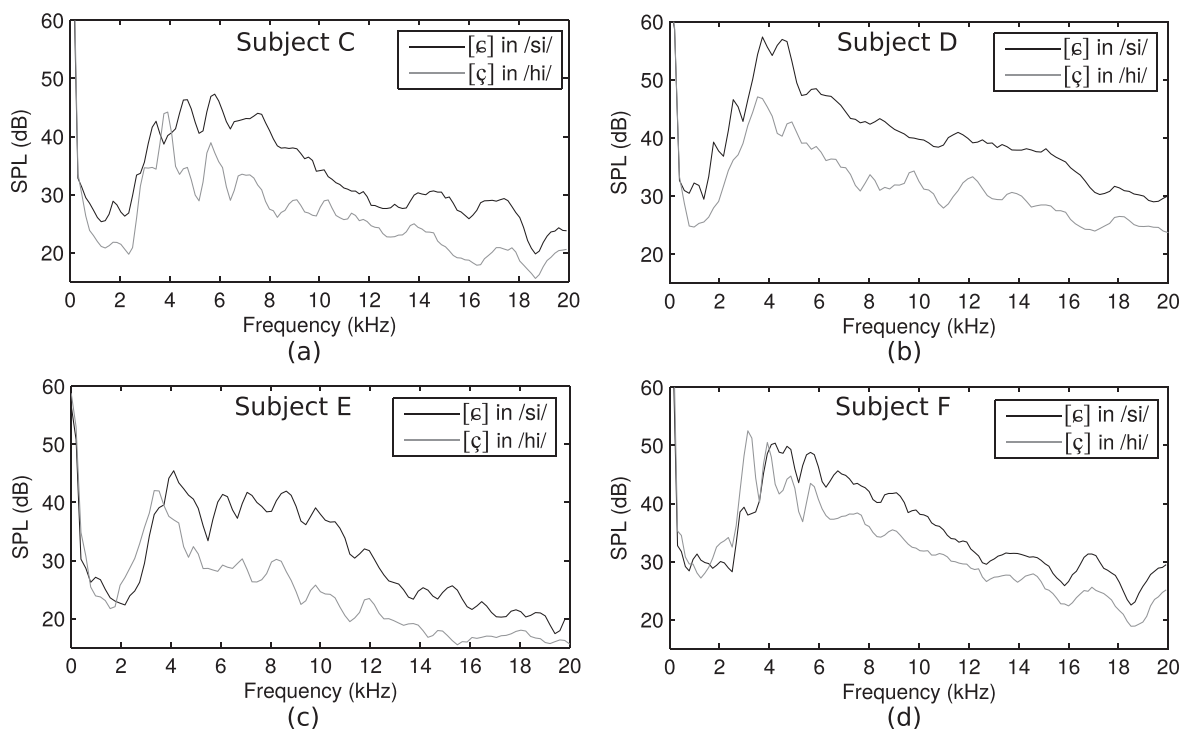


FIG. 2. Spectra of fricatives uttered by four subjects. The spectra of [ε] and [ç] for subjects C, D, E, and F are plotted in (a), (b), (c), and (d), respectively, and averaged from five repetitive recordings.

same MRI system used for rtMRI (see Sec. IV), the vocal tract images of the sustained fricative were obtained. The subject was asked to demonstrate the articulatory postures for the fricative [ç] in /si/ and retain it while the recording occurred. A total of 36 sagittal images of 512×512 pixels (slice thickness: 2 mm) were collected (voxel spacing: $0.5 \times 0.5 \times 2 \text{ mm}^3$). Based on the collected images, the vocal tract geometry was transformed into a simplified model of the pharynx, linguopalatal (tongue) constriction, space behind the incisors, and lip cavity. The dimensions of the model are presented in Fig. 3. The size of the teeth was based on the separately measured digital dental cast using MRI. Details of the data acquisition and vocal tract simplification processes were described by Yoshinaga *et al.* (2019).

When subject C uttered [ç], the size of the lip cavity was approximately $6 \times 33 \text{ mm}$ with a longitudinal length (x_1 direction) of 8 mm. The gap between the upper and lower incisors was 1.5 mm, and the distance between the lower incisor and the tongue constriction was approximately 5 mm. The constriction had a cross section of $2 \times 7 \text{ mm}$ in the coronal plane. Considering this geometry, the acoustic characteristics of the subject's sustained [ε] were reproduced by a mechanical replica (Yoshinaga *et al.*, 2019).

Three simplified tongue models—A, B, and C—were constructed and presented in Fig. 3(c) to examine the cause of the acoustic differences between [ε] and [ç] produced at the same place of articulation. The observed dimensions of the constriction of the sustained [ε] were replicated in model A. Based on the assumption that the place of articulation [i.e., the position of the constriction in the anterior–posterior direction; see Fig. 1(b)] in the mid-sagittal plane remained

the same and that the size of the constriction differed in the direction along the coronal plane, the width of the constriction W_C was increased from 7 mm in model A to 9.3 mm in model B and then 14 mm in model C. To maintain the same level of amplitude of the generated sounds, the height of the constriction H_C was decreased from 2 to 1 mm, which kept the maximum velocity at the constriction. Besides the constriction sizes, the flow rate of the model was also controlled in the experiment, as the amplitudes of [ç] in the frequency range above 4 kHz decreased more rapidly than those of [ε], as presented in the acoustic analysis in Sec. IV. In addition to the flow rate of $300 \text{ cm}^3/\text{s}$, which was the value observed in a study of Japanese sibilants (Yoshinaga *et al.*, 2019), a lower flow rate of $217 \text{ cm}^3/\text{s}$ was used.

Both flow rates were combined with each of the three tongue models in the experiment. The new flow rate ($217 \text{ cm}^3/\text{s}$) was determined based on trial and error to evaluate the flow rate value with which we could reproduce the observed acoustic amplitudes of the subject. Moreover, in a supplementary experiment, we confirmed that the intraoral pressure of the model decreased by approximately 200 Pa with the decrease in the flow rate from 300 to $217 \text{ cm}^3/\text{s}$, whereas the pressure increased by approximately 58 Pa by changing the tongue models from A to C.

B. Experiments with a physical setup

The simplified models we have described were constructed using acrylic boards, and the sound generated by the model was measured. Figure 4 presents the experimental setup. The airflow was supplied by an air compressor (SOL-

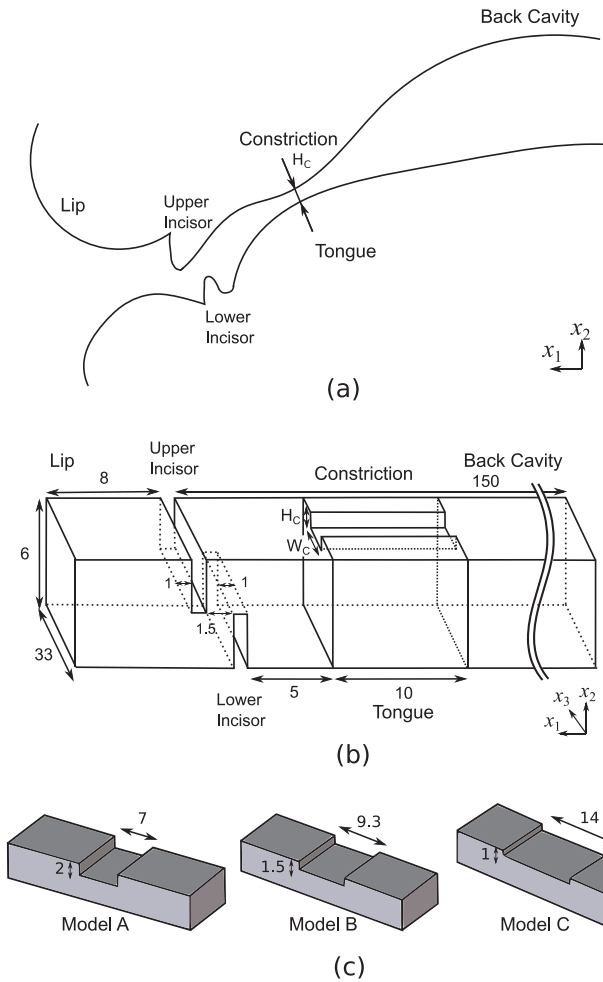


FIG. 3. (Color online) Simplified vocal tract model. The mid-sagittal plane of the vocal tract is presented in (a). The dimensions of the simplified model are presented in (b). The dimensions for three tongue models are presented in (c). The dimensions are expressed in mm.

2039, Misumi, Tokyo, Japan) through a tube with a 16-mm inner diameter connected to a flow valve (IR2000-02, SMC, Tokyo, Japan) and a flow meter (PFM750-01, SMC). A silencer was set between the model and flow meter to reduce the noise emitted upstream from the flow meter. The model was positioned at the center of an anechoic chamber ($V=8.1\text{ m}^3$), and the sound generated by the model was measured using a microphone (type 4939, Bruel & Kjaer) at 300 mm (for comparison with the subjects) and 100 mm (for validation of the numerical simulation) from the model outlet. At the outlet of the model, a baffle board of $300 \times 300\text{ mm}$ was set to imitate the speaker's face.

A constant flow rate of 217 or $300\text{ cm}^3/\text{s}$ was set by the flow valve, and the signal from the microphone was recorded for 3 s by a data acquisition system (PXIe-6351, National Instruments, Austin, TX). The data were sampled with a sampling frequency of 50 kHz. The sound spectrum was calculated using the DFT with a 256-point time window that was multiplied by a Hanning window. The magnitudes of spectra were calculated with 60 windows and an overlap ratio of 30%. The frequency resolution of the spectra for the experiments was 196 Hz.

C. Numerical simulation

To elucidate the flow configuration and cause of the acoustic differences, a large eddy simulation (LES) of the compressible flow was conducted on the simplified models. The LES is one of the numerical simulations in computational fluid dynamics; it calculates the time variation of flow fields only in the computational grid scale, and the energy of smaller-scale eddies than the grid scale is dissipated by applying additional viscosity to the flow with a grid filter. The governing equations are the three-dimensional compressible Navier–Stokes equations (Liu and Vasilyev, 2007):

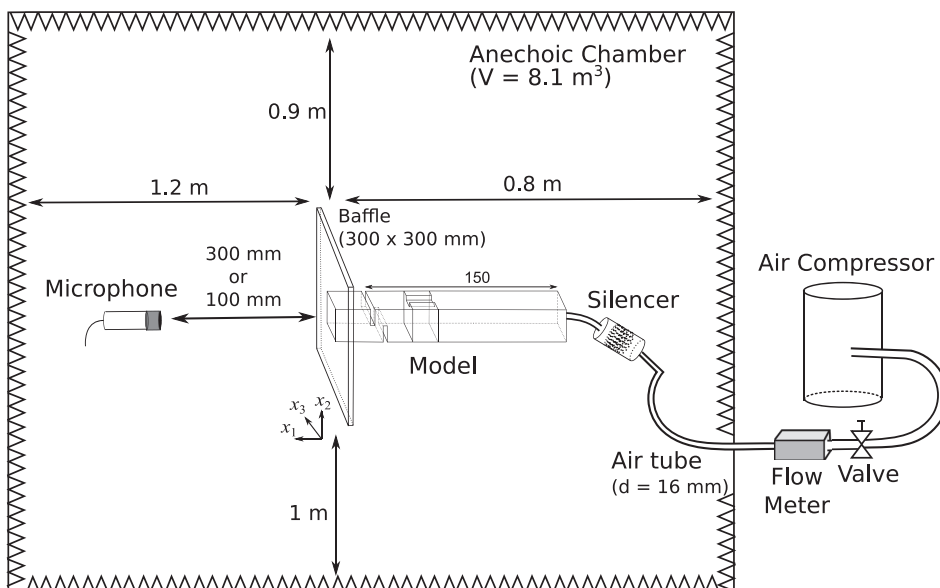


FIG. 4. Experimental setup.

$$\mathbf{Q}_t + (\mathbf{E} - \mathbf{E}_\nu)_{x_1} + (\mathbf{F} - \mathbf{F}_\nu)_{x_2} + (\mathbf{G} - \mathbf{G}_\nu)_{x_3} = \mathbf{V}, \quad (1)$$

$$\mathbf{V} = -\left(\frac{1}{\phi} - 1\right)\chi \begin{pmatrix} \partial\rho u_i/\partial x_i \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (2)$$

where \mathbf{Q} denotes the vector of the conservative variables; \mathbf{E} , \mathbf{F} , and \mathbf{G} denote the inviscid flux vectors; and \mathbf{E}_ν , \mathbf{F}_ν , and \mathbf{G}_ν denote the viscous flux vectors. To express the geometry of the model in computational grids, volume penalization (VP) was employed as an immersed boundary method. The penalization term \mathbf{V} was added to the right side of the Navier–Stokes equations as an external force, so that the complicated wall geometry could be easily expressed in the structured grids. In the VP method, the model wall was expressed as a porous medium, and the porosity of the medium ϕ was set as $\phi = 0.25$ to reflect the sound wave on the model’s wall. Through a preliminary simulation, we confirmed that the reflectivity of the wall was 99%. The mask function χ is expressed as

$$\chi = \begin{cases} 1 & \text{(inside object),} \\ 0 & \text{(outside object)} \end{cases} \quad (3)$$

to distinguish the wall region from the flow region.

The spatial derivatives were solved by a sixth-order-accuracy compact finite-difference scheme (fourth-order accuracy at the boundaries). Time integration was performed using the third-order-accuracy Runge–Kutta method. To reduce computational costs, the LES was applied by filtering turbulent energy from the grid scale to the subgrid scale as an implicit turbulence model. LES filtering was conducted using a tenth-order spatial filter:

$$\alpha_f \hat{\psi}_{i-1} + \hat{\psi}_i + \alpha_f \hat{\psi}_{i+1} = \sum_{n=0}^5 \frac{a_n}{2} (\psi_{i+n} + \psi_{i-n}), \quad (4)$$

where ψ denotes a conservative quantity, and $\hat{\psi}$ denotes the filtered quantity. The coefficient a_n had the same values as in Gaitonde and Visbal (2000), and α_f was set to 0.45. Details of this simulation methodology were described by Yokoyama *et al.* (2015).

In Fig. 5, the computational domain and boundary conditions of the LES are presented. The computational domain was divided into three regions: the flow region, acoustic region, and buffer region. In the flow region, the grid spacing was set to resolve the turbulent vortices around the constriction and teeth obstacles. In Fig. 6, the computational grids in the flow region for models A and C are presented. The minimum grid size was $\Delta x_1 = 1.94 \times 10^{-2}$ mm for both models. In the acoustic region, the grid spacing was set to capture the sound waves. The maximum grid size was set to $\Delta x_1 = 4.25$ mm, and the wavelength of a sound at 4 kHz was resolved by 20 points. The pressure at $x_1 = 100$ mm from the

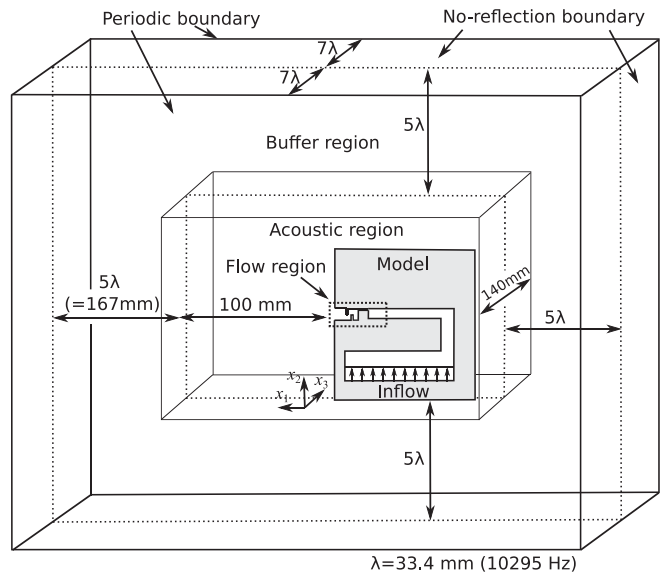


FIG. 5. Computational domains and boundary conditions of the numerical simulation.

outlet of the model (outer boundary of the acoustic region) was sampled for the sound spectrum. The buffer region was set to prevent acoustic reflections at the outlet of the computational domain which smoothly reduces the pressure fluctuations that propagate from the model. Among these three regions, the grid sizes were gradually altered to prevent acoustic disturbances at the boundaries. The total number of grid points was approximately 8.2×10^7 for both models.

As a boundary condition, the non-reflecting boundary was set in the x_1 and x_2 directions, whereas the periodic boundary condition was set in the x_3 direction. To reduce the computational cost, the upper airway of the model was bent perpendicularly, and the uniform flow velocity was set

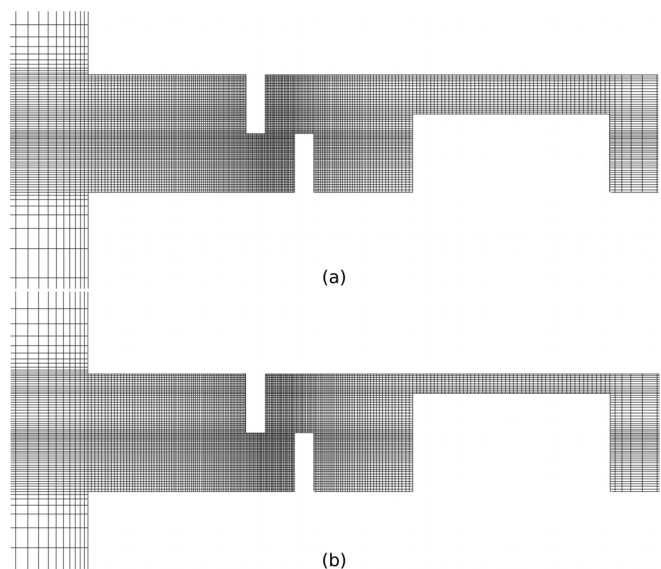


FIG. 6. Computational grids for the flow region of the simplified model. Grids for tongue models A and C are presented in (a) and (b), respectively. Every third grid line is presented for clarity.

at the inlet of the model to reproduce the experimental flow rates of 217 and 300 cm³/s. Inside the model region, the mask function χ was set to 1, and the velocity was set to 0. A time step for the time integration was set to 2.26×10^{-8} s, and 6×10^5 iterations were performed after 10^5 preliminary iterations. The sound spectrum was calculated using the DFT with a 256-point window, multiplied by a Hanning window. The magnitudes of spectra were calculated with five windows and the overlap ratio 30%. The frequency resolution of the spectra was 347 Hz.

D. Results of the experiment with the physical setup

In Fig. 7, the spectra of the sounds generated by the three models are presented. The spectra of sounds generated with a flow rate of 300 cm³/s and those of the sounds generated with a flow rate of 217 cm³/s are plotted in Figs. 7(a) and 7(b), respectively. The simplified vocal tract with the three models—A, B, and C—generated sound with frequencies ranging from 3 to 14 kHz for both flow rates. All the sounds had a characteristic peak at 4 kHz. The overall amplitudes of the sounds generated with models A and B were almost the same under the flow rates of 217 and 300 cm³/s. On the contrary, when model C with a wider constriction width was combined with the flow rate of 300 cm³/s, the amplitudes above 4 kHz were increased by 3 dB. Moreover, the peak amplitude at 4 kHz increased by 13 dB under the flow rate of 217 cm³/s. These results indicate that

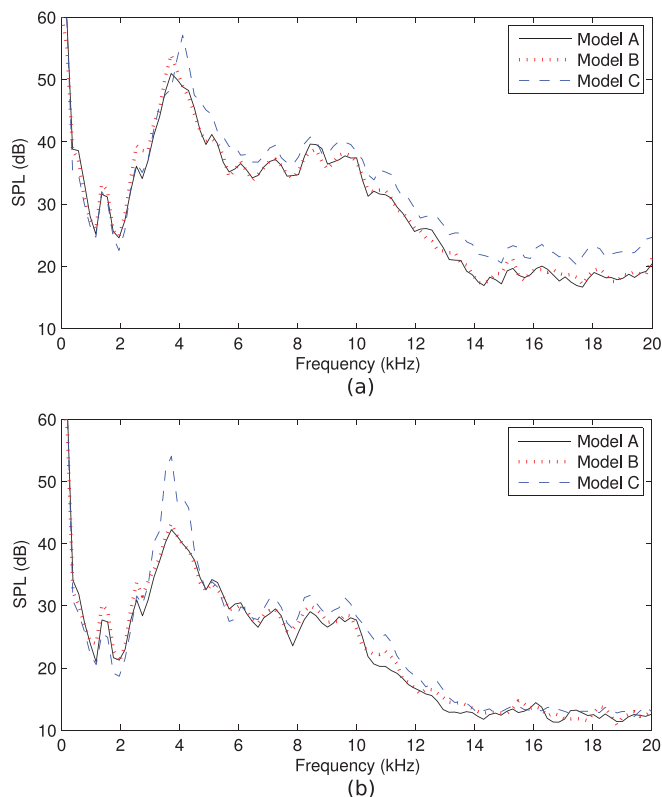


FIG. 7. (Color online) Comparison of the spectra of sounds generated by the three models A, B, and C. The sounds generated with the flow rates of 300 and 217 cm³/s are presented in (a) and (b), respectively.

the change in the flow rate caused not only the overall amplitude changes [as has been reported by Jesus and Shadle (2002) and Nozaki *et al.* (2014)] but also the changes in the shapes of the spectra at the characteristic peak.

Figure 8 presents the comparison of the spectra of sounds generated by the simplified tongue models and fricatives [ç] and [ç], as produced by human subject C. The measured spectra of model A with a flow rate 300 cm³/s and those of model C with a flow rate of 217 cm³/s are compared with the spectra of human-generated [ç] and [ç]. The sounds analyzed in this figure are attached as supplementary materials.² The first characteristic peaks at 4 kHz for both [ç] and [ç] were captured by the models. More importantly, a decrease in amplitudes from [ç] to [ç] in the frequency range above 4 kHz was reproduced by decreasing the flow rate from 300 to 217 cm³/s and changing the tongue model from A to C. By contrast, the peaks in the frequency range of 6–8 kHz, and above 12 kHz, were underestimated by approximately 10 dB in models A and C. In addition, the harmonic-like peaks at 6, 8, 15, and 17 kHz in human-generated [ç] were not observed in the sound generated by model A. The overall spectral differences between the human-generated [ç] and [ç] agreed well with the results of the simplified models A and C. To further elucidate this issue, the results of the numerical simulation using the combination of model A and the 300-cm³/s flow rate and

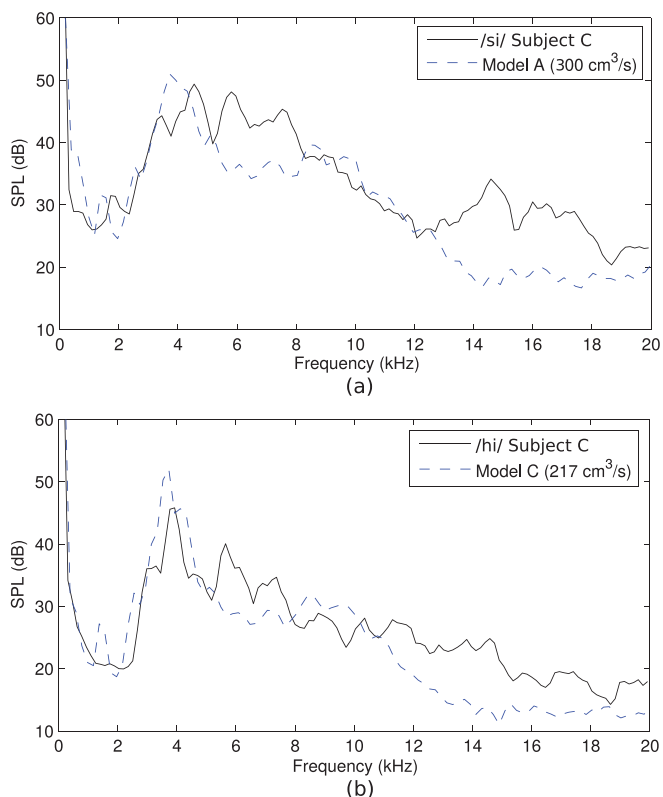


FIG. 8. (Color online) Comparison of the spectra of fricatives generated by subject C and sounds generated by the simplified models. (a) compares the spectra of human-generated [ç] and that of model A with the 300-cm³/s flow rate. (b) compares the spectra of human-generated [ç] and that of model C with the 217-cm³/s flow rate.

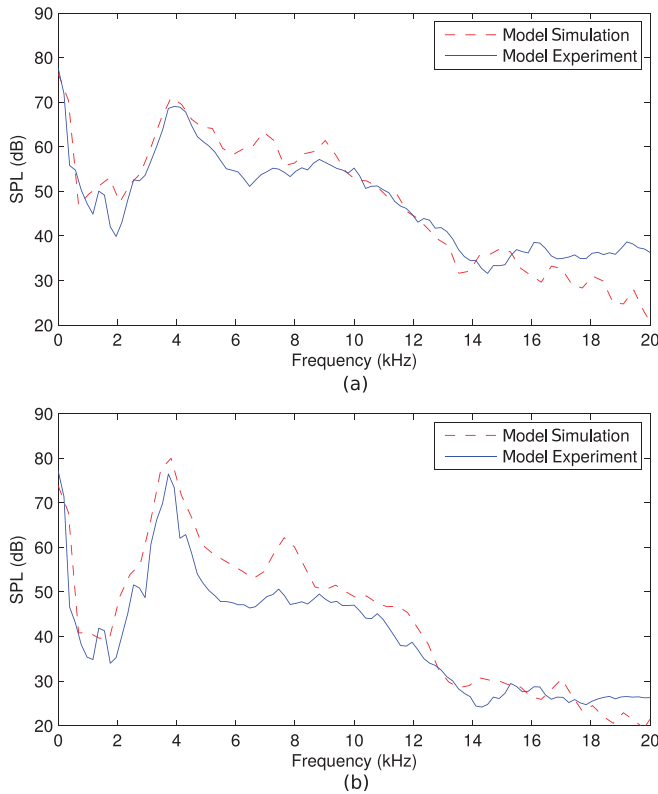


FIG. 9. (Color online) Comparison of spectra obtained by the physical experiment and numerical simulation. The spectra obtained by models A and C are plotted in (a) and (b), respectively. In the experiment and simulation, sounds were sampled at 10 cm from the lip outlet of the simplified model.

that of model C and the 217-cm³/s flow rate are discussed in Sec. V E.

E. Results of the numerical simulation

This section discusses the results of the numerical simulation conducted to elucidate the causes of the acoustical differences between [ç] and [ç̣]. To confirm the accuracy of

the simulation, Fig. 9 presents the comparison of the spectra obtained by the simulations with the spectra obtained by physical experiments. The sound pressures for both simulation and experiment were sampled at 10 cm from the lip outlets of the simplified models.

The overall spectral amplitudes of the simulation using model A matched well up to 15 kHz with those obtained by the experiment with the simplified models. A maximum discrepancy of approximately 7 dB was observed at approximately 7 kHz. The spectrum obtained by the simulation using model C overestimated the values obtained by the experiment in the frequency range of 2–14 kHz; however, the overall spectrum shape matched well with the measured spectrum. In addition, the resonance peak observed at 4 kHz in the experiment was replicated by the simulation. These results indicate the validity of using the results of the simulation as evidence for the examination of the causes of the acoustical differences between the two fricatives.

In Fig. 10, the instantaneous velocity magnitudes, root mean square (rms) values of velocity fluctuation, and vorticity in the x_3 direction are presented. The instantaneous velocity magnitudes presented in Figs. 10(a) and 10(b) reveal that the flow traveling from the back cavity caused the maximum flow velocity at the constriction in both models; the flow leaving from the constriction impinged on the teeth obstacle in both models. A part of the jet flow travelled to the gap between the lower incisor and tongue constriction, and the recirculated flow disturbed the jet flow. The recirculation is especially clear in model C. After leaving the gap between the teeth, the flow impinged on the lower lip surface and exited the model.

The Reynolds numbers Re , based on the maximum mean velocity at the constriction $|u|_{max}$ and the height of the constriction H_C , were 5107 and 1660 for models A and C, respectively. The rms of the velocity fluctuations presented in Figs. 10(c) and 10(d) indicate that large magnitudes of velocity fluctuation were observed near the backside of the

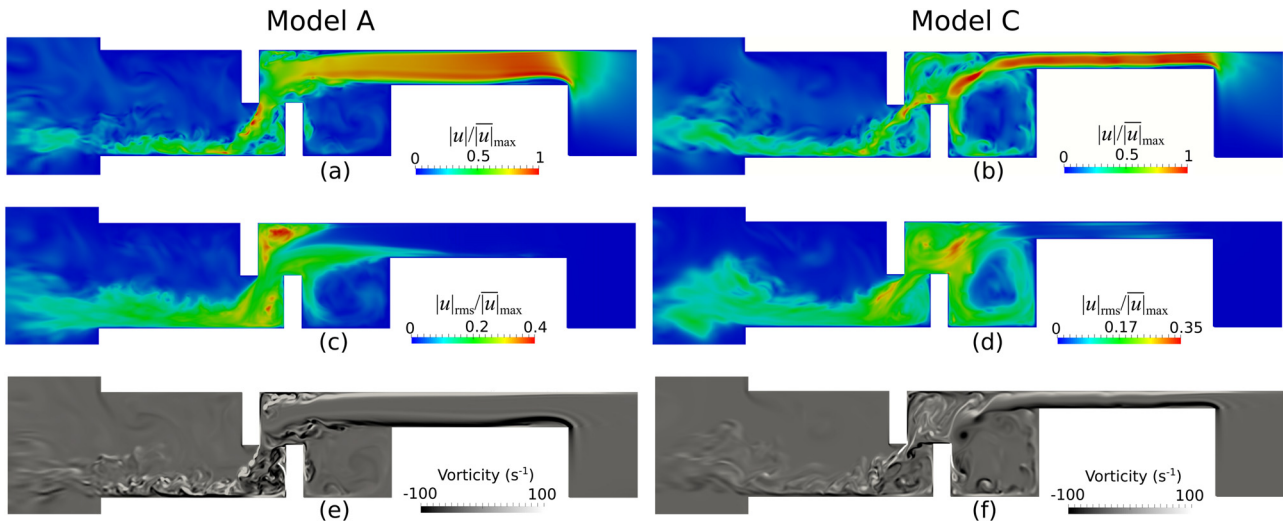


FIG. 10. (Color online) Comparison of models A and C with respect to the contour of the flow field in the mid-sagittal plane (x_1-x_2) at the center ($x_3 = 0$). (a) and (b) present the instantaneous velocity magnitude; (c) and (d) show the rms of the velocity fluctuation; and (e) and (f) show the vorticity in the x_3 direction.

upper incisor and the front side of the lower incisor in model A; in model C, a large fluctuation was observed in the space between the upper incisor and tongue constriction. As the flow rate was decreased from model A to C, the maximum rms value $|u|_{rms}/|u|_{max}$ decreased from 0.40 to 0.35.

Vorticity in the x_3 direction is plotted in Figs. 10(e) and 10(f) to elucidate the vortex structures in the jet flow presented in Figs. 10(a) and 10(b). When the jet flow impinged on the upper incisor, small vortices were observed in the mixing layer of the jet flow generated by the constriction. In model A with a lower tongue (hence a wider pathway of air at the constriction), a thick jet directly impinged on the upper incisors; in model C with a higher tongue (hence a narrower pathway at the constriction), a thinner jet formed small vortices before it impinged on the teeth obstacles. Moreover, in model C, small disturbances were observed inside the flow in the constriction.

To elucidate the three-dimensional flow configurations at the constriction, the second invariant of the velocity gradient tensor $Q = \|\Omega\|^2 - \|\mathbf{S}\|^2$ was calculated, where Ω and \mathbf{S} denote the asymmetric and symmetric parts of the velocity gradient tensor, respectively. The regions in $Q > 0$ represent vortex tubes. In Fig. 11, the iso-surfaces for $Q/(|u|_{max}/H_C)^2 = 0.65$ for tongue models A and C are presented. In the constriction of model A, the flow separated from the walls of the constriction inlet became turbulent and began to form vortex tubes near the sidewalls of the constriction. On the

contrary, the flow was smooth at the center of the constriction of model C, although the separated flow formed small vortex tubes near the sidewalls of the constriction. In addition, the periodic vortex tubes in the span-wise (x_3) direction were observed only for model C (see also the supplemental video files).²

To determine the frequency at which the periodic vortices in the constriction emerge, the power spectrum density of the velocity fluctuations sampled at the constriction inlet (where the periodic vortices were observed in Fig. 11) was calculated and plotted in Fig. 12. The sampling point was 2 mm downstream of the constriction inlet on the mid-sagittal plane. Although the frequency characteristics of the velocity fluctuation for model A were nearly flat up to 20 kHz, those of model C peaked at 4 kHz. This peak frequency coincided with that observed in the far-field sound of the model and human-generated [ç] (Fig. 8).

VI. DISCUSSION

Real-time MRI observation demonstrated that some Japanese speakers produce perceptually distinct [ç] and [ç̥] with nearly the same place of articulation in the mid-sagittal plane. Although numerous researchers have investigated the acoustic characteristics of fricatives [s], [ʃ], and [f], and so forth (Badin, 1991; Shadle, 1985; Jesus and Shadle, 2002), the acoustic contrast between sibilant and non-sibilant fricatives issued from a similar place of articulation has not been investigated.

In this study, it was found that the acoustic contrast between [ç̥] and [ç] is characterized by the differences in the peak frequency of approximately 4 kHz and the relative amplitudes of the broadband noise to the resonance peak of 4 kHz by means of the acoustic measurement of sustained fricatives recorded in an anechoic room. It is important to note here that the acoustic contrast between [ç̥] and [ç] in Fig. 2 may have been more extreme than that of naturally produced fricatives, as these sounds were recorded as sustained consonants. We have confirmed in a supplementary study, however, that the similar tendencies were observed in both fricatives produced as a part of words in an anechoic chamber and fricatives produced in rtMRI recordings under

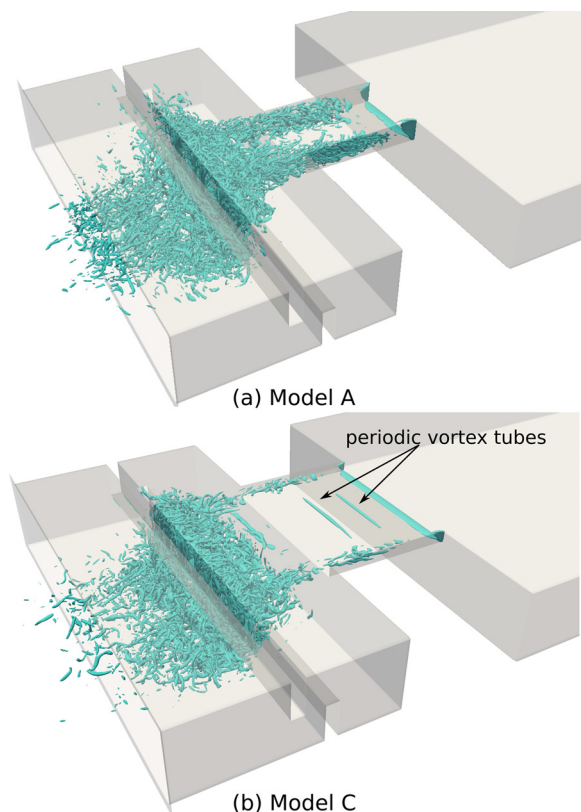


FIG. 11. (Color online) The iso-surfaces of the second invariant $Q/(|u|_{max}/H_C)^2 = 0.65$ in the simplified model. The iso-surfaces for models A and C are presented in (a) and (b), respectively.

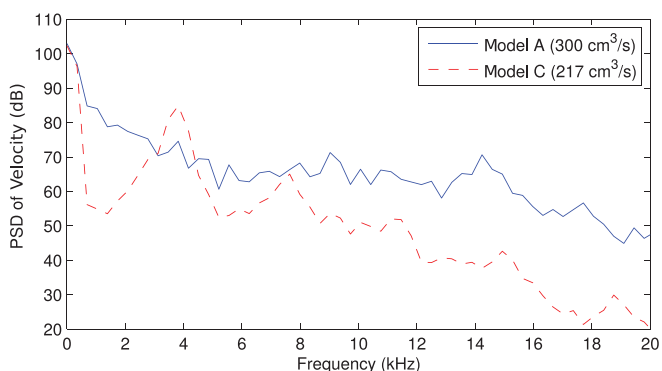


FIG. 12. (Color online) Comparison of models A and C with respect to the spectra of velocity fluctuation at 2 mm from the constriction inlet.

operational machine noise.³ According to systematic studies of simplified vocal tract models, the change in peak frequency of [ç] can be explained as a variation in the constriction position in the anterior–posterior (x_1) direction (Shadle, 1985; Toda *et al.*, 2010; Yoshinaga *et al.*, 2017). Further study is needed to elucidate whether these characteristics are common in other languages.

Using simplified physical models, we reproduced the acoustic differences between [ç] and [ç] by changing the constriction width and flow rate. Such differences were also reproduced by employing numerical simulation on the simplified models. The results strongly suggest the possibility that some Japanese speakers differentiate sibilant [ç] from non-sibilant [ç] by controlling their constriction width and flow rate, given that the place of articulation in the mid-sagittal plane is the same. The differences in amplitudes in the frequency ranges of 6–8 and 14–20 kHz were caused by the flow channel's geometrical differences between the rectangular simplified model and rounded vocal tracts of the subjects [details are compared by Yoshinaga *et al.* (2018)]. Moreover, the overestimated peaks in the simulation were probably caused by the insufficient grid resolution of the turbulent flow. This discrepancy is reduced by improving the grid resolution, as presented by Yoshinaga *et al.* (2020).

When the tongue model was changed from A to C, the amplitudes were increased by 3 dB at 4 kHz with a constant flow rate of 300 cm³/s and by 13 dB at 4 kHz with a flow rate of 217 cm³/s (Fig. 7). This finding indicates that the turbulent intensity decreased due to the decrease in the flow rate, and periodic vortex structures were strongly observed at the constriction (Fig. 11). Yokoyama and Kato (2009) reported that a tonal noise was produced by fluid–acoustic interactions in a rectangular cavity along a uniform flow. The periodic sound of the simplified model was also produced by the interaction between the acoustic resonance of the model's geometry (4 kHz) and periodic vortex generation at the constriction inlet (Figs. 11 and 12).⁴

In speech sciences, it has been thought that sibilant fricatives are characterized by the sibilant groove, uniquely formed by the tongue tip's contact with the hard palate (Stevens, 1989; Perkell *et al.*, 2004). On the contrary, non-sibilant fricatives are known to be generated by the wider constriction in the coronal plane (Shadle, 1985; Narayanan *et al.*, 1995). Similarly, the feature [Shape] has been proposed by Ladefoged and Maddieson (1996). Such a feature corresponds to a flat and grooved tongue for alveolar fricatives, while it can be associated with a domed and palatalized tongue for post-alveolar fricatives. Our study revealed that the acoustic contrast between [ç] and [ç] is partly accompanied by a difference in the constriction width; this result is in agreement with those in the cited literature.⁵

In contrast, a distinction based on the presence versus absence of the high-velocity jet of air striking the edge of the teeth was not fully supported between the sibilants and non-sibilant fricatives by this study. As can be seen in Fig.

10, the jet of air produced by the constriction reached the edge of the upper incisor in both models (i.e., model A approximating sibilant [ç] and model C approximating non-sibilant [ç]). The presence of airflow turbulence around the upper incisor was also observed for both models in Fig. 11. Theoretically, the physical characteristics of the vortices determine the spectral characteristics of the fricatives: fine vortex tubes result in larger amplitude in the high-frequency region, whereas coarse vortex tubes result in a lower frequency region (Krane, 2005). Therefore, the finer vortices around the upper incisor of model A might result in larger amplitudes in the high-frequency region. However, a more detailed examination of the physical properties of the vortex tubes should be the theme of further research.

Finally, the difference in the flow rate was important to produce the relatively large amplitude of the resonance peak. The relevance of flow rate for the description of fricatives has attracted the attention of some scholars (e.g., Isshiki and Ringel, 1964; Catford, 1977; Laver, 1994) but has not been incorporated into the mainstream classificatory theories in phonetics and phonology (International Phonetic Association, 1999; Ladefoged and Maddieson, 1996; Laver, 1994). In addition, the whistling effect in fricative production is also evident in the data of other experiments with physical setups at certain flow rates (Birkholz *et al.*, 2020).

VII. CONCLUSION

In this study, the causes of the acoustic difference between the Japanese fricatives [ç] and [ç] produced with a nearly identical place of articulation in the mid-sagittal plane were investigated using (1) a real-time MRI movie, (2) spectral measurements in speech of sustained fricatives produced in an anechoic room, (3) an aerodynamic experiment with simplified vocal tract models combining three constriction widths (7, 9.3, and 14 mm) and two flow rates (300 and 217 cm³/s), and (4) a numerical simulation of the airflow in simplified vocal tracts. Based on these analyses, we reproduced the acoustic differences between [ç] and [ç] by controlling the coronal width of the constriction and the flow rate without changing the place of articulation in the mid-sagittal plane. According to the numerical flow simulation, the sharp spectral peak at 4 kHz characterizing [ç] in the subjects was the result of the fluid–acoustic interactions caused by the generation of periodic vortex tubes at the constriction.

From the viewpoint of phonetics and phonology, these results indicate that constriction width and flow rate play crucial roles in the scientific description of the difference between sibilants and non-sibilants in Japanese. Further studies must clarify whether these aeroacoustic differences are also relevant in the production of sibilant and non-sibilant fricatives in languages other than Japanese. If similar phenomena are observed in numerous languages, flow rate and/or constriction width should be integrated into a general classificatory theory of articulatory phonetics.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) as Program for Promoting Research on the Supercomputer “Fugaku” (Grant Nos. hp200123 and hp200134), Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Nos. JP17H02339, JP19H03976, JP19K21641, and JP20H01265), JSPS Grant-in-Aid for Scientific Research on Innovative Areas (Grant No. JP20H04999), and the research budget of the Center for Corpus Development, National Institute for Japanese Language and Linguistics. We acknowledge Dr. Hiroshi Yokoyama for his helpful comments on the numerical simulation.

¹Both speakers A and B devoiced the vowel /i/ in the first mora of the bimoraic word /hisi/. Devoicing of close vowels between the adjacent voiceless consonants is frequent in Tokyo Japanese (Maekawa and Kikuchi, 2005).

²See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0003936> for recording data of the subject /si/ and /hi/ (SuppPubmm1.wav and SuppPubmm2.wav) and models A and C (SuppPubmm3.wav and SuppPubmm4.wav), respectively, and for video files of the vortex tubes in models A and C (SuppPubmm5.mp4 and SuppPubmm6.mp4).

³This supplementary study was conducted based on the comment given by an anonymous referee to an earlier draft of this paper.

⁴An anonymous referee of this paper pointed out the possibility that in the case of human speech production, the rugged surface of the palate (caused by the presence of palatal rugae) suppresses, at least partly, the generation of periodic vortices observed in the numerical simulation. We will focus on this effect in further research.

⁵Note that the change in the constriction width is not identical to the change in the constriction degree, described by the cross-sectional areas of the constriction (Clements and Ridouane, 2006), because the cross-sectional area of the three tongue models is the same.

Badin, P. (1991). “Fricative consonants: Acoustic and x-ray measurements,” *J. Phon.* **19**, 397–408.

Birkholz, P., Kürbis, S., Stone, S., Häsner, P., Blandin, R., and Fleischer, M. (2020). “Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties,” *Sci. Data* **7**(1), 1–16.

Catford, J. C. (1977). *Fundamental Problems in Phonetics* (Edinburgh University Press, Edinburgh).

Chomsky, N., and Halle, M. (1968). *The Sound Patterns of English* (Harper & Row, New York).

Clements, G. N., and Ridouane, R. (2006). “Quantal phonetics and distinctive features,” in *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, August 28–30, Athens, Greece, pp. 17–27.

Gaitonde, D. V., and Visbal, M. R. (2000). “Pade-type higher-order boundary filters for the Navier–Stokes equations,” *AIAA J.* **38**(11), 2103–2112.

Fletcher, S. G. (1989). “Palatometric specification of stop, affricate, and sibilant sounds,” *J. Speech Hear. Res.* **32**(4), 736–748.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (Cambridge University Press, Cambridge).

Isshiki, N., and Ringel, R. (1964). “Air flow during the production of selected consonants,” *J. Speech Hear. Res.* **7**, 233–244.

Jakobson, R., Fant, G., and Halle, M. (1952). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates* (MIT, Cambridge, MA).

Jannedy, S., and Weirich, M. (2014). “Sound change in an urban setting: Category instability of the palatal fricative in Berlin,” *Lab. Phonol.* **5**(1), 91–122.

Jesus, L. M. T., and Shadle, C. H. (2002). “A parametric study of the spectral characteristics of European Portuguese fricatives,” *J. Phon.* **30**(3), 437–464.

Krane, M. H. (2005). “Aeroacoustic production of low-frequency unvoiced speech sounds,” *J. Acoust. Soc. Am.* **118**, 410–427.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World’s Languages* (Blackwell, London).

Laver, J. (1994). *Principles of Phonetics* (Cambridge University, Cambridge, England).

Liu, Q., and Vasilyev, O. V. (2007). “A Brinkman penalization method for compressible flows in complex geometries,” *J. Comput. Phys.* **227**(2), 946–966.

Maekawa, K. (2010). “Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese,” *J. Phon.* **38**(3), 360–374.

Maekawa, K. (2019). “A real-time MRI study of Japanese moraic nasal in utterance-final position,” in *Proceedings of the International Congress of Phonetic Sciences (ICPhS2019)*, August 5–9, Melbourne, Australia, pp. 1987–1991.

Maekawa, K. (2021). “Production of the utterance-final moraic nasal in Japanese: A real-time MRI study,” *J. Int. Phonetic Assoc.* (in press).

Maekawa, K., and Kikuchi, H. (2005). “Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report,” in *Voicing in Japanese*, edited by J. van de Weijer, K. Nanjo, and T. Nishihara (Mouton de Gruyter, Berlin), pp. 205–228.

Moran, S., and McCloy, D. (2019). PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History, <http://phoible.org> (Last viewed 7/18/2020).

Narayanan, S. S., Alwan, A. A., and Haker, K. (1995). “An articulatory study of fricative consonants using magnetic resonance imaging,” *J. Acoust. Soc. Am.* **98**(3), 1325–1347.

National Language Research Institute (1974). *Linguistic Atlas of Japan I–V* (National Printing Bureau, Tokyo).

Nozaki, K., Yoshinaga, T., and Wada, S. (2014). “Sibilant /s/ simulator based on computed tomography images and dental casts,” *J. Dent. Res.* **93**(2), 207–211.

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 and use of an articulatory saturation effect,” *J. Speech Lang. Hear. Res.* **47**(6), 1259–1269.

Reidy, P. F. (2016). “Spectral dynamics of sibilant fricatives are contrastive and language specific,” *J. Acoust. Soc. Am.* **140**(4), 2518.

Shadle, C. H. (1985). “The acoustics of fricative consonants,” Ph.D. dissertation (MIT, Cambridge, MA).

Stevens, K. N. (1989). “On the quantal nature of speech,” *J. Phon.* **17**(1–2), 3–45.

Toda, M., and Honda, K. (2003). “An MRI-based cross-linguistic study of sibilant fricatives,” in *Proceedings of the 6th International Seminar on Speech Production*, December 7–10, Sydney, Australia, pp. 1–6.

Toda, M., Maeda, S., and Honda, K. (2010). “Formant-cavity affiliation in sibilant fricatives,” in *Turbulent Sounds: An Interdisciplinary Guide*, edited by F. Fuchs, M. Toda, and M. Zygis (De Gruyter Mouton, Berlin) pp. 343–374.

Verhoeven, J., Mariën, P., De Clerck, I., Daems, L., Reyes-Aldasoro, C. C., and Miller, N. (2019). “Asymmetries in speech articulation as reflected on palatograms: A meta-study,” in *Proceedings of the International Congress of Phonetic Sciences (ICPhS2019)*, August 5–9, Melbourne, Australia, pp. 2821–2825.

Yokoyama, H., and Kato, C. (2009). “Fluid-acoustic interactions in self-sustained oscillations in turbulent cavity flows in fluid dynamic oscillations,” *Phys. 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,” *J. Acoust. Soc. Am.* **138**(2), 858–873.

Yoshinaga, T., Nozaki, K., and Iida, A. (2020). “Hysteresis of aeroacoustic sound generation in the articulation of [s],” *Phys. Fluid* **32**(10), 105114.

Yoshinaga, T., Nozaki, K., and Wada, S. (2017). “Effects of tongue position in the simplified vocal tract model of Japanese sibilant fricatives /s/ and /ʃ/,” *J. Acoust. Soc. Am.* **141**(3), EL314–EL318.

Yoshinaga, T., Nozaki, K., and Wada, S. (2018). “Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract model,” *Phys. Fluids* **30**(3), 035104.

Yoshinaga, T., Nozaki, K., and Wada, S. (2019). “Aeroacoustic analysis on individual characteristics in sibilant fricative production,” *J. Acoust. Soc. Am.* **146**(2), 1239.

Zharkova, N. (2016). “Ultrasound and acoustic analysis of sibilant fricatives in preadolescents and adults,” *J. Acoust. Soc. Am.* **139**(5), 2342.