



Variability in Production of Non-Sibilant Fricative [ç] in /hi/

Tsukasa Yoshinaga¹, Kikuo Maekawa², Akiyoshi Iida¹

¹Toyohashi University of Technology, Japan

²National Institute for Japanese Language and Linguistics, Japan

yoshinaga@me.tut.ac.jp, kikuo@ninjal.ac.jp, iida@me.tut.ac.jp

Abstract

The alveolo-palatal sibilant fricative [ç] in /si/ and palatal non-sibilant fricative [ç] in /hi/ are known to be distinguished by the geometrical change of the vocal tract only in the coronal direction and acoustically unstable in some Japanese speakers. In this study, we reconstructed the vocal tract geometry with three repetitive magnetic resonance imaging (MRI) on the same subject and investigated the effects of coronal vocal tract shapes on the airflow and sound characteristics in [ç]. The computational grids were constructed on each vocal tract and a numerical airflow simulation was conducted to predict the turbulent airflow and aeroacoustic sound generation. The predicted sound properties were in good agreement with the measurement of Japanese subjects. The comparison of the airflow and acoustic characteristics among three vocal tracts showed that the slight changes of the constriction area and flow channel downstream from the constriction influenced the sound source generation and peak amplitudes at around 3 kHz, indicating that the characteristic peak of [ç] was variable due to the constriction shape at the middle part of the hard palate.

Index Terms: speech production, magnetic resonance imaging, aeroacoustics, numerical flow simulation, vocal tract

1. Introduction

The fricative consonants are known to be pronounced by the turbulent airflow formed by the narrow constriction in a part of the vocal tract. In the fricatives, sibilant fricatives are characterized as a sound generated by the jet airflow striking on the teeth obstruction, whereas non-sibilant fricatives are characterized by a sound produced by the turbulent jet itself [1]. The acoustic contrast between sibilants and non-sibilants can be caused by small differences in a place of articulation (e.g., the alveolar ridge for [s] and the hard palate for [ç]); however, the resulting change in the spectral characteristics covers a wide frequency range of up to 20 kHz because of the nature of the turbulence noise [2]. Therefore, it is believed that changes in the shape of the constriction and its jet flow formation (not the place of articulation) coincide with the acoustic contrast between sibilants and non-sibilants.

In the Japanese language, the constriction is formed at the hard palate for both alveolo-palatal sibilant fricative [ç] in /si/ and palatal non-sibilant fricative [ç] in /hi/. Since the place of articulation is very close, the difference between [ç] and [ç] is often confusing, and it has been argued that the same place of articulation may be one of the causes for Japanese dialects [3]. For example, the initial mora of /sicigacu/ “July” is variably realized as [çi] and [ç] in many dialects, according to the linguistic geographical survey reported in the Linguistic Atlas of Japan [4]. A similar ongoing phonological change is also reported in Germany [5-6].

Because of its importance to the phonetic description of speech sounds and their phonological classification, the articulatory differences causing the acoustic contrast have been investigated by measuring vocal tract configurations applying medical imaging techniques. For example, one-dimensional area functions have been estimated using X-ray pictures [7], and three-dimensional vocal tract geometries have been reconstructed using magnetic resonance imaging (MRI) [8]. In addition, the detailed geometrical differences among the sibilants [s], [ʃ], [ç], and [ç] have been analyzed *via* MRI [9-10].

In our group, the vocal tract geometry of [ç] in /si/ and [ç] in /hi/ was measured using real-time MRI technique, and we found that the places of articulation for [ç] and [ç] were nearly identical for three out of ten Japanese subjects [3]. The separate acoustic measurements in an anechoic chamber revealed that the amplitude of the broadband noise of [ç] was weaker than that of [ç], while the characteristic peak amplitude at approximately 3.5 kHz was greater in [ç] than in [ç]. The characteristic peaks in the spectra are shown in Fig. 1. The additional analysis with a simplified vocal tract model and numerical airflow simulations indicated that the differences in a width of the constriction in the coronal plane and periodic vortex generations are the causes of the acoustic difference in the vocal tracts with the same place of articulation. However, it is still unclear whether these changes occur in the actual vocal tract and how it leads to the variability in [ç] and [ç].

Therefore, in this study, we reconstructed the three-dimensional realistic vocal tract geometry of [ç] in /hi/ for the three repetitive MRI measurements with the same subject and investigated the airflow and sound characteristics in the realistic vocal tracts. The repetitive measurements with the identical subject enable to analyze the variability in both vocal tract geometry and its sound source characteristics of [ç].

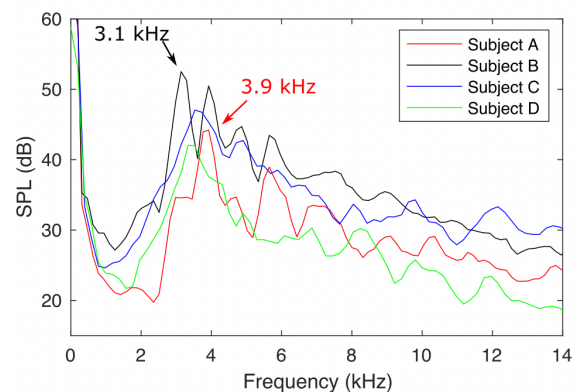


Figure 1: Spectra of [ç] in /hi/ uttered by four subjects. The sound is recorded at 30 cm from lips.

2. Materials and Methods

2.1. Vocal tract acquisition

The vocal tract geometry of a subject pronouncing [ç] in /hi/ was obtained by the MRI. The subject is a male native Japanese speaker aged 51 years, and he pronounces [ç] and [ε] with the same place of articulation [3]. The vocal tracts were scanned by a 3T MRI system (MAGNETOM Prisma fit 3T, Siemens, Germany), at ATR Brain Activity Imaging Center, Kyoto. The MRI measurements were conducted as a part of constructing a real-time MRI articulatory movement database [11].

To measure the variability of the vocal tract geometry of [ç], three MRI scans were conducted. The first scan was acquired on slices of the coronal plane with 128×128 pixels and the spatial resolution of $2 \times 2 \times 2 \text{ mm}^3$ voxels. The second and third scans were obtained on slices of the sagittal plane with 256×256 pixels and the spatial resolution of $1 \times 1 \times 3 \text{ mm}^3$ voxels. Fig. 2 presents the comparison of the vocal tract images obtained with two setups. The subject sustained [ç] in the context of /hi/ for 20 s in a supine position. A total of three vocal tracts was reconstructed from the three MRI scans. The vocal tract region in the MR images was segmented by an automatic segmentation algorithm in the software program itk-SNAP. Additionally, frontal teeth (i.e. incisors) images were measured by pressing tongue and inner lip surfaces on the upper and lower incisors, and the segmented geometry was superimposed on the vocal tract.

The obtained vocal tracts are shown in Fig. 3 from the pharynx to the lips. Since the first scan resolution was 2 mm, the height of the palate constriction was larger than those of the other vocal tracts. The configuration and position of the constriction are presented in Table 1. The constriction heights of vocal tracts 2 and 3 were almost the same, while the width of vocal tract 2 was larger than the other vocal tracts.

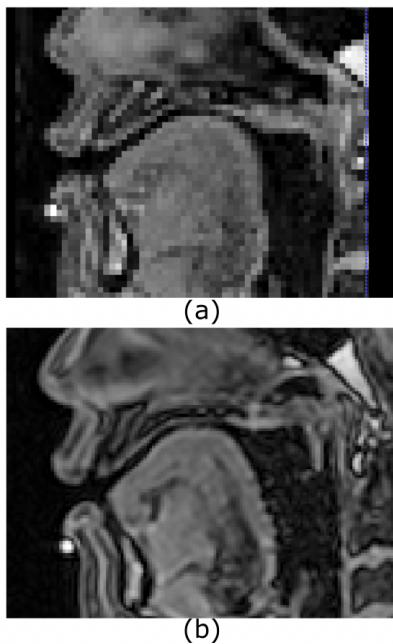


Figure 2: Mid-sagittal planes of the subject sustained [ç]. The slice resolution is $2 \times 2 \text{ mm}$ pixels in (a), whereas the resolution is $1 \times 1 \text{ mm}$ pixels in (b).

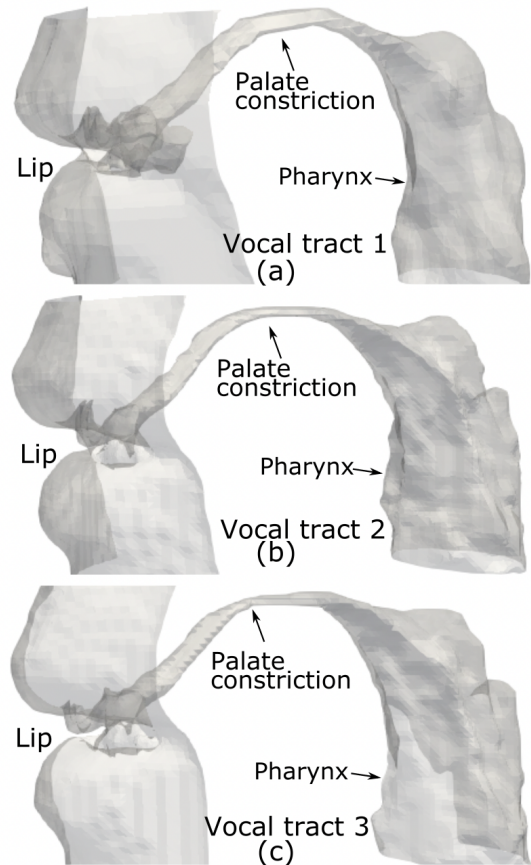


Figure 3: Reconstructed vocal tracts for three repetitive MRI measurements. (a) Vocal tract 1, (b) vocal tract 2, and (c) vocal tract 3.

Table 1: Configuration and position of the palate constriction for each vocal tract.

Vocal Tract	Height (mm)	Width (mm)	Area (mm ²)	Distance from lip (mm)
1	3.15	4.06	8.93	50.9
2	1.64	6.91	8.83	49.4
3	1.31	4.05	4.87	38.7

2.2. Numerical airflow simulation

To simulate the airflow and sound generation in the pronunciation of [ç], a large eddy simulation (LES) of the compressible flow was conducted on the obtained vocal tracts. 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 with the immersed-boundary method [12] that is applicable for the realistic vocal tract geometry. The spatial derivatives were solved by a sixth-order accuracy compact finite-difference scheme, while the time integration was performed using the third-order accuracy Runge-Kutta method. The LES was applied by filtering the turbulent energy using a 10th-order spatial filter as an implicit turbulence model. Details of the method are described in [3].

The computational grids on the mid-sagittal plane for vocal tract 1 are presented in Fig. 4. Near the palate constriction, the grid spacing was set to resolve the turbulent vortices, and the minimum grid size was $\Delta x = 0.1$ mm. To reduce the computational costs, the grid pacing was gradually increased towards the lip outlet, and the grid size was set to capture the sound waves. The sound pressure at 100 mm from the lower lip of each vocal tract was sampled to calculate the sound spectrum. To prevent the acoustic reflection at the outlet of the computational domain, the buffer region, which smoothly reduces the pressure fluctuations, was set at the outer computational domain. The total number of grid points was approximately 9.5×10^7 for each vocal tract.

As a boundary condition, the non-reflecting boundary was set on the outer computational domain. To reduce the computational cost, the vocal tract geometry upstream from the epiglottis was removed, and a pressure chamber of the back cavity ($37.7 \times 37.7 \times 56$ mm³) was set just below the pharynx. A constant pressure of 300 Pa was set at the inlet of the pressure chamber of vocal tracts 1 and 2. Meanwhile, the pressure was increased to 540 Pa for vocal tract 3 to match the sound amplitude, because vocal tract 3 had the smaller constriction area, and the resulting sound amplitude decreased with the same inlet pressure. The axes x_1 , x_2 , and x_3 are defined in the anterior-posterior, superior-inferior, left-right directions, respectively.

A time step for the time integration was set to 1.0×10^{-7} s, and 3.3×10^5 iterations were performed after 5×10^4 preliminary iterations. The sound and velocity were sampled at 100 kHz, and spectra were calculated using the DFT with a 790-point window, multiplied by a Hann window. The average of spectra was calculated with seven windows and an overlap ratio of 40%.

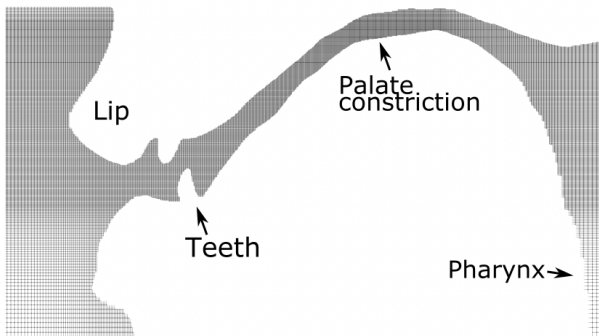


Figure 4: Computational grids on the mid-sagittal plane.

3. Results and Discussion

The spectra of sounds predicted by the flow simulation are plotted in Fig. 5. The sound generated from the three vocal tracts became broadband noise with the characteristic peak at 3 to 3.5 kHz. Vocal tracts 1 and 3 had the main peak at 3.0 kHz, while vocal tract 2 had the main peak at 3.5 kHz. The harmonic peak was also observed at 6.0 kHz for vocal tract 1 and 7.4 kHz for vocal tract 3.

Although the subject pronounced [ç] in the same vowel context, the spectral shape was slightly changed due to the different vocal tract geometry. In particular, a sharp peak appeared only for vocal tract 2, and sounds of the other vocal tracts were dominated by the broad aeroacoustic noise which has no significant peak. Nevertheless, the sound produced by

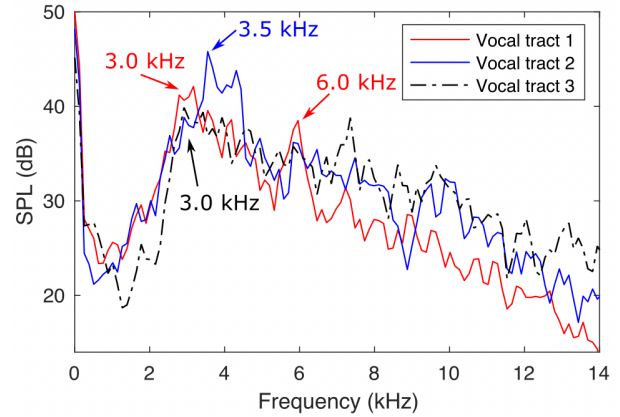


Figure 5: Spectra of sound predicted by the flow simulation. The sound pressure was sampled at 100 mm from the lower lip.

three vocal tracts had the main peak at around 3 kHz, and spectral shapes were similar to those measured with the other Japanese subjects (Fig. 1). These results indicate that the vocal tract acquisition via MRI and the numerical prediction with the flow simulation had enough accuracy to reproduce the phenomena in the pronunciation of [ç].

It has been argued that the peak frequency of fricative sounds is determined by the distance between the constriction and lip outlet [13-14]. Although the main peak frequency was almost the same for three vocal tracts, the distance between the minimum constriction area and lip outlet 38.7 mm of vocal tract 3 was approximately 10 mm shorter than those of vocal tracts 1 and 2 as shown in Table 1. If we consider the acoustic resonance of a one-closed tube at 3 kHz, 1/4 wavelength is 28.6 mm and much smaller than the distance between the constriction and lip outlet. This suggests that the acoustic resonance length was affected and became shorter by the jet flow generated from the constriction.

The airflow velocity magnitudes at the mid-sagittal plane of the three vocal tracts are shown in Fig. 6. The maximum velocity was 33.2, 28.7, and 42.3 m/s for vocal tracts 1, 2, and 3, respectively, and the color bar was normalized by each maximum velocity. As shown in the figure, the jet flow was generated by the constriction for all vocal tracts, while the beginning of the jet in vocal tract 3 was closer to the incisors and lips compared to the other vocal tracts due to the position of the minimum constriction.

The jet flow generated from the constriction became turbulent near the front part of the hard palate and produced the aeroacoustic sound. Although the maximum velocity of vocal tract 2 was smaller than the other vocal tracts, the large magnitude of the velocity remained from downstream of the constriction to the lip outlet. As a result, the sound amplitude of vocal tract 2 was the maximum among the three vocal tracts. In contrast, vocal tracts 1 and 3 had larger areas from the constriction to the lip outlet, and velocity fluctuations of turbulent flows distributed only near the palate and tongue surfaces.

The non-sibilant fricatives are known to be distinguished by an absence of the strong turbulent noise due to the jet striking on the teeth obstacles [1]. However, we found that the turbulent flow can also remain from the palate to teeth in some of vocal tracts in [ç], and the flow striking on the teeth

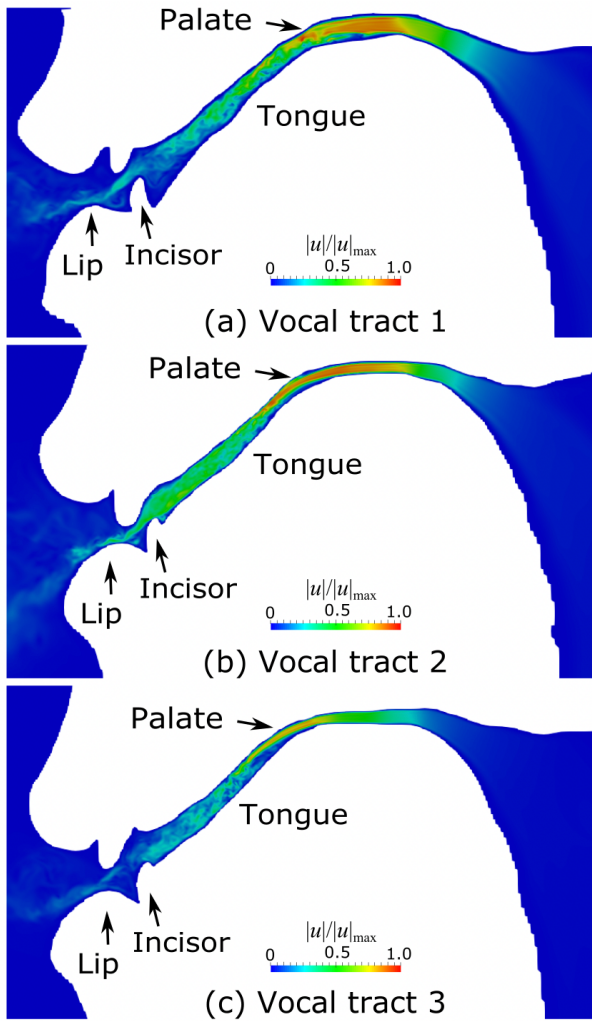


Figure 6: Velocity magnitudes on the mid-sagittal plane of each vocal tract. (a) Vocal tract 1, (b) vocal tract 2, and (c) vocal tract 3.

amplified the characteristic peak.

To examine the properties of the sound source in the turbulent flow, the velocity fluctuation u_1 in x_1 -direction (the anterior-posterior direction) was sampled, and the power spectrum density was calculated and plotted in Fig. 7. The velocity fluctuations in the broadband range are the turbulent sound source, and the characteristic peak is formed by the acoustic resonance in the oral tract. Since the jet flow configuration was different in each vocal tract, the velocity at 40 mm from the lower lip was sampled for vocal tract 1, whereas the velocity at 32 mm from the lower lip was sampled for vocal tracts 2 and 3. As shown in Fig. 7, the velocity fluctuation of vocal tracts 1 and 2 had a peak at 5.8 kHz, while there was no characteristic peak in the spectrum of the velocity fluctuation of vocal tract 3. The peaks in vocal tracts 1 and 2 indicate that the turbulent vortices were generated with a frequency of 5.8 kHz and these vortices became the sound source of the periodic sounds at 3 and 6 kHz. In addition, this result suggests that the periodic vortices, which were observed in the simplified vocal tract model in the previous study [3], also appeared in the realistic geometry. In contrast, there was no specific peak in the source of vocal tract

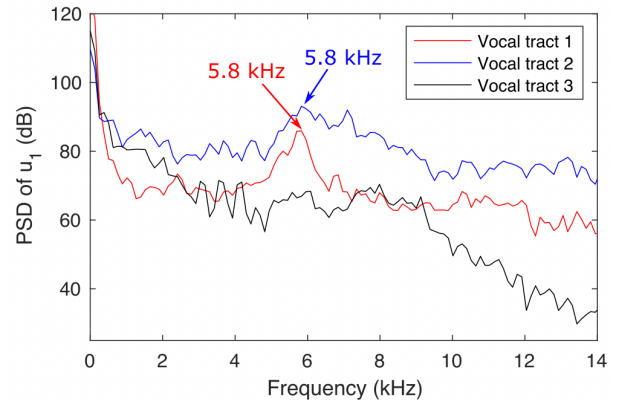


Figure 7: Power spectrum density of velocity fluctuation u_1 in x_1 -direction.

3, indicating that the main peak at 3 kHz was formed only by the acoustic resonance in the vocal tract. In addition, the amplitudes of the velocity fluctuations downstream from the constriction as well as the far-field sound were smaller than those of the other vocal tracts, although the velocity at the constriction in vocal tract 3 was larger than those of the other vocal tracts. These results demonstrate that the small changes of vocal tract configurations influence the flow characteristics, and it causes the variability in the acoustic properties of [ç].

4. Conclusions

In this study, the vocal tracts of non-sibilant fricative [ç] in /hi/ were obtained from one subject, and the variability of flow and acoustic properties was examined by the numerical flow simulation. The obtained vocal tracts showed that the position and area of the constriction were almost the same for different MRI resolutions (vocal tracts 1 and 2), while those varied in the same resolution (vocal tracts 2 and 3). These results indicate that the vocal tract configuration easily varies in the production of [ç]. The predicted sound (Fig. 5) showed good agreement with the sounds measured for the other Japanese subjects (Fig. 1), suggesting that the numerical prediction with the flow simulation had enough accuracy to reproduce the phenomena in the pronunciation. The comparison of the flow and acoustic characteristics among three vocal tracts revealed that the slight changes of the constriction area and the flow channel downstream from the constriction influenced the sound source generation and the peak amplitudes at around 3 kHz. Hence, these variabilities in the acoustic properties may result in the phonetic/phonological confusion between [ç] /si/ and [ç] /hi/ in Japanese. Further analysis of the sound generation mechanisms of the characteristic peak is expected, while further simulations with the vocal tracts of different subjects and different languages are anticipated.

5. Acknowledgements

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI, Grant Number: JP19H03976, JP19K21641, and JP20H01265.

6. References

[1] P. Ladefoged, and I. Maddieson, *The Sounds of the World's Languages*. London: Blackwell, 1996.

- [2] K. N. Stevens, "On the quantal nature of speech," *Journal of Phonetics*, vol. 17, pp. 3–45, 1989.
- [3] T. Yoshinaga, K. Maekawa, and A. Iida, "Aeroacoustic differences between the Japanese fricatives [ɕ] and [ç]," *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2426–2436, 2021.
- [4] National Language Research Institute, *Linguistic Atlas of Japan I-V*. Tokyo: National Printing Bureau, 1966–1974.
- [5] S. Jannedy, and M. Weirich, "Sound change in an urban setting: Category instability of the palatal fricative in Berlin," *Laboratory Phonology*, vol. 5, no 1, pp. 91–122, 2014.
- [6] F. Conrad, "Regional differences in the evolution of the merger of /ʃ/ and /ç/ in Luxembourgish," *Journal of the International Phonetic Association*, pp. 1–18, 2021.
- [7] P. Badin, "Fricative consonants: Acoustic and x-ray measurements," *Journal of Phonetics*, vol. 19, no. 3, pp. 397–408, 1991.
- [8] S. S. Narayanan, A. A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 98, no. 3, pp. 1325–1347, 1995.
- [9] M. Toda, and K. Honda, "An MRI-based cross-linguistic study of sibilant fricatives," in *Proceedings of 6th international seminar on Speech Production*, pp. 1–6, 2003.
- [10] M. Toda, S. Maeda, and K. Honda, "Formant-cavity affiliation in sibilant fricatives," in *Interface Explorations*, edited by S. Fuchs, M. Toda, and M. Zygis, pp. 343–374, 2010.
- [11] K. Maekawa, "Production of the utterance-final moraic nasal in Japanese: A real-time MRI study," *Journal of the International Phonetic Association*, pp. 1–24, 2021.
- [12] Q. Liu, and O. V. Vasilyev, "A Brinkman penalization method for compressible flows in complex geometries," *Journal of Computational Physics*, vol. 227, no 2, pp. 946–966, 2007.
- [13] C. H. Shadle, "The acoustics of fricative consonants," *Ph.D. dissertation*, Massachusetts Institute of Technology. Cambridge, MA, 1985.
- [14] T. Yoshinaga, K. Nozaki, and S. Wada, "Aeroacoustic analysis on individual characteristics in sibilant fricative production," *The Journal of the Acoustical Society of America*, vol. 146, no. 2, pp. 1239–1251, 2019.