INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Individual Differences of Airflow and Sound Generation in the Vocal Tract of Sibilant /s/

Tsukasa Yoshinaga[1], *Kazunori Nozaki*[2], *Shigeo Wada*[3]
[1]Toyohashi University of Technology, Japan
[2]Osaka University Dental Hospital, Japan
[3]Graduate School of Engineering Science, Osaka University, Japan
yoshinaga@me.tut.ac.jp, knozaki@dent.osaka-u.ac.jp, shigeo@me.es.osaka-u.ac.jp

Abstract

To clarify the individual differences of flow and sound characteristics of sibilant /s/, the large eddy simulation of compressible flow was applied to vocal tract geometries of five subjects pronouncing /s/. The vocal tract geometry was extracted by separately collecting images of digital dental casts and the vocal tract of /s/. The computational grids were constructed for each geometry, and flow and acoustic fields were predicted by the simulation. Results of the simulation showed that jet flow in the vocal tract was disturbed and fluctuated, and the sound source of /s/ was generated in different place for each subject. With an increment of the jet velocity, not only the overall sound amplitude but also the spectral mean was increased, indicating that the increment of the jet velocity contributes to the increase of amplitudes in a higher frequency range among different vocal tract geometries.

Index Terms: speech production, large eddy simulation, sibilant fricatives

1. Introduction

Sibilant /s/ is pronounced by generating a turbulent jet flow in the front part of the vocal tract. The jet flow is generated by a constricted flow channel formed by a tongue tip and alveolar ridge. The generation mechanisms of sibilant /s/ have been investigated by several researchers by theoretically modeling the sound source in the simplified vocal tract [1-3]. In the modeling, the sound source was assumed to be generated when the jet flow impinged on the teeth obstacle [1-2] or when the jet passed through the gap between upper and lower teeth [3]. Then, by using cross-sectional areas of the vocal tract, the modeling captured the frequency characteristics of /s/.

In contrast to the theoretical modeling, the flow configuration in the vocal tract geometry of /s/ has been investigated by flow simulations. To simulate the jet flow and sound generation in the vocal tract, large eddy simulation (LES) of compressible flow was applied to the simplified geometry [4]. The LES of compressible flow enables to capture both the formation of small vortexes in the turbulent flow and propagation of the sound from the flow sources. Moreover, the LES was also applied to the realistic vocal tract geometry of /s/, and emergence of the sound source on a teeth surface was visualized [5]. However, these studies considered only a single geometry for each study, and effects of geometrical differences of individual vocal tracts on the flow and sound generation is still unclear.

Therefore, in this study, we applied the LES for the vocal tract of five subjects obtained by magnetic resonance imaging (MRI) to clarify the effects of the geometrical differences of five subjects on the flow and acoustic characteristics of sibilant /s/.

2. Method

2.1. Vocal tract acquisition

The vocal tract geometry of five subjects pronouncing /s/ without vowel context was obtained by MRI. The five subjects are male native Japanese speakers aged 22 to 26 years (mean: 23.6). The images of the vocal tract were collected using MAGNETOM Prisma fit 3T (Siemens, Germany). Since the MRI cannot capture the tooth geometry while the subject is sustaining /s/, the digital tooth cast was separately collected by the MRI, and the cast was superimposed on the images of /s/ as described in [6]. The resolutions of the MRI for /s/ and the tooth cast are 0.5×0.5×2 mm and 0.25×0.25×1 mm, respectively. The obtained image of /s/ and vocal tract geometry were shown in Fig. 1. The geometry from glottis to pharynx was omitted to reduce computational costs of the simulation.

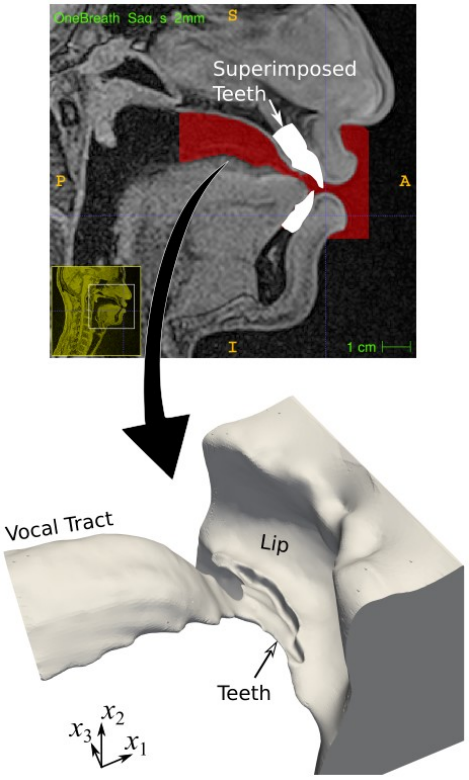

Figure 1: *MRI and extracted vocal tract geometry of /s/ (Subject A).*

http://dx.doi.org/10.21437/Interspeech.2019-1376

2.2. Large eddy simulation

To simulate the turbulent flow and sound generation in the vocal tract, the LES of compressible flow was applied to the geometries of five subjects. Governing equations of the LES are three-dimensional compressible Navier-Stokes equations and equation of state:

$$\frac{\partial \bar{\rho}}{\partial t} = -\frac{\partial \bar{\varphi}_j}{\partial x_j}, \quad (1)$$

$$\frac{\partial \bar{\varphi}_i}{\partial t} + \frac{\partial \bar{\varphi}_i \tilde{u}_j}{\partial x_j} = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial \sigma_{ij}}{\partial x_j}, \quad (2)$$

$$\frac{\partial \bar{\rho}\bar{E}}{\partial t} + \frac{\partial \bar{\varphi}_j \bar{E}}{\partial x_j} = -\frac{\partial \bar{p}\tilde{u}_j}{\partial x_j} + \frac{\partial \sigma_{ij}\tilde{u}_j}{\partial x_j} - \frac{\partial q_j}{\partial x_j}, \quad (3)$$

$$\bar{p} = (\gamma - 1)\bar{\rho}\bar{e}, \quad \bar{e} = C_v \bar{T} \quad (4)$$

where, $\bar{\rho}$ is density, $\tilde{u}_i$ is velocity components (i = 1, 2, 3), $\bar{\varphi}_i = \overline{\rho u_i}$ is momentum flux, $\bar{p}$ is pressure, $\bar{E} = 1/2\,|\tilde{u}_i|^2 + \bar{e}$ is the total energy, $\bar{e}$ is the internal energy, and $\bar{T}$ is temperature. The viscous stress tensor σ_{ij} and the heat flux q_j are calculated as

$$\sigma_{ij} = 2\bar{\rho}(\nu + \nu_{SGS})\left(s_{ij} - \frac{1}{3}\delta_{ij}s_{ll}\right), \quad (5)$$

$$q_j = -\bar{\rho}\gamma C_v(\alpha + \alpha_{SGS})\frac{\partial \bar{T}}{\partial x_j} \quad (6)$$

with grid-scale (GS) kinematic viscosity, subgrid-scale (SGS) viscosity ν_{SGS}, GS thermal diffusivity α, and SGS thermal diffusivity α_{SGS} [7]. The strain rate tensor s_{ij} is calculated as $s_{ij} = 1/2\left(\partial\tilde{u}_i/\partial x_j + \partial\tilde{u}_j/\partial x_i\right)$. The gas constants C_v and γ are determined from the values of air at atmospheric pressure and temperature 20 ℃. The equations were implemented and solved in the finite volume software OpenFOAM 2.3.1 (OpenCFD Ltd).

To examine the sound source configuration in the flow field, the source term in Lighthill's equation with low Mach number assumption [8]

$$\psi = \frac{\partial^2 \rho u_i u_j}{\partial x_i \partial x_j}, \quad (7)$$

was calculated with the air density ρ and the velocity components u_i.

2.3. Computational grids and boundary conditions

In the extracted geometry of five subjects, the computational grids are constructed, and the density, the velocity components, the total energy, and the pressure are calculated at each grid point. The computational grids for the vocal tract of subject A are shown in Fig. 2. A far-field region with a length 110 mm in x_1 direction was added in front of the lip to enable the sound propagation. The grid sizes in the region from the constriction to the lips were reduced to capture small vortexes of the jet flow. The total number of grid elements for each vocal tract is summarized in Table 1.

Boundary conditions are depicted in Fig. 3. At the inlet of the vocal tract geometry, a uniform velocity was imposed to set the constant flow rate for each vocal tract. Since we could not measure the flow rate while the subjects were sustaining /s/ in the MRI machine, we used a general flow rate of /s/ 250 cm^3/s [9]. Meanwhile, the flow rate of subject D was increased up to 300 cm^3/s since the sound amplitude was too small to compare with other subjects. The flow rates for five subjects are summarized in Table 1. For the vocal tract wall, no-slip and adiabatic boundaries were used. At the outlet, the no-reflection boundary condition was imposed to prevent artificial sound reflection from the outlet.

The time step for the time integration was set to 1×10^{-7} s, and 13.2×10^4 iterations were performed. To evaluate acoustic characteristics of the predicted sounds, overall sound pressure level (OASPL) and the first spectral moment (spectral mean, M_1) of the pressure at x_1 = 100 mm from the lips were calculated by the discretized Fourier transform (DFT) with 256-point frames sampled at 100 kHz.

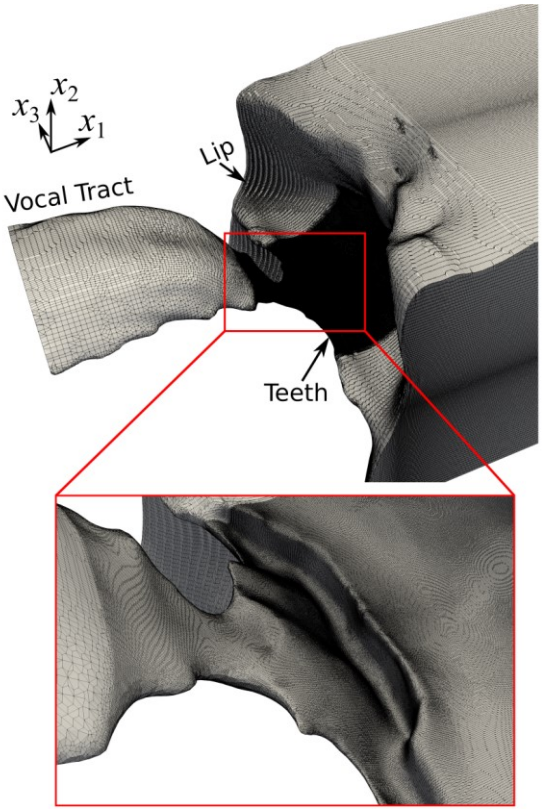

Figure 2: *Computational grids of subject A.*

Table 1: *Total number of elements for computational grids and volume flow rate Q at the inlet.*

Subject	Total number of elements (×10^6)	Q (cm^3/s)
A	41.2	250
B	35.2	250
C	42.0	250
D	39.4	300
E	37.4	250

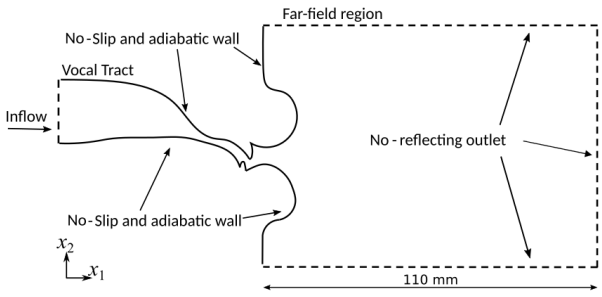

Figure 3: *Boundary conditions.*

3. Results and Discussion

Predicted instantaneous velocity fields on the sagittal plane of the vocal tract for subject A, B, C, D, and E are shown in Fig. 4 (a), (b), (c), (d), and (e), respectively. The position of the sagittal plane is set at the centre of the maximum constriction for each subject. The velocity magnitude is normalized by the maximum mean velocity at the constriction.

In the vocal tract of subject A, the jet flow directly impinged on the upper teeth. In subject B, the jet flow passed along a surface of the upper incisor and was disturbed at the gap between teeth. In subject C, the jet flow passed along the surface of the upper teeth and started fluctuating in a cavity between the upper teeth and the lower lip. In subject D, the jet flow directly impinged on the lower teeth and passed along the lower lip surface. In subject E, although the velocity magnitude started increasing at the constriction, the velocity reached the maximum at the gap between the tongue tip and the upper teeth. Although the flow configurations for five subjects were different, the formation of the maximum velocity by the tongue tip was consistent with the previous simulation [5] and experiment with a vocal tract replica [10].

Root mean square (RMS) values for time variation of the sound source ψ on the sagittal plane are shown in Fig. 5. Although large amplitudes of the sound source appeared in a region from the front part of the constriction to the lip outlet for every subject, positions of the maximum amplitude of the source varied depending on the subject. The maximum source for subject A appeared on the lower lip surface, whereas the maximum for subject B and C appeared in the space behind the upper teeth. The maximum source for subject D appeared at a gap between the lower teeth and the upper palate, whereas the maximum for subject E appeared at the tip of the upper teeth. The main source generation with the impingement on teeth for subject C and D agreed with the modeling of [1-2], whereas the source generation at the gap between teeth for subject A and B agreed with the modeling of [3]. In contrast, the main source generation with the jet impingement on the lower lip surface was firstly observed in subject E.

The maximum amplitude of the sound source was plotted with the maximum mean velocity at the constriction for each subject in Fig. 6 (a). With an increment of the velocity at the constriction, the amplitude of the sound source increased except for subject A. This result suggests that the vocal tract of subject A has a more winding flow channel than the other subjects, and the flow was largely disturbed by the vocal tract surfaces. The OASPL and the spectral mean M_1 of predicted sound spectrum were plotted with the maximum mean velocity in Fig. 6 (b) and (c), respectively. The bar shows the mean and standard deviation for seven frames of the DFT. With the increment of the velocity at the constriction, not only the OASPL but also M_1 was increased. These results suggest that with the increment of the velocity, the amplitude of flow source in higher frequency range was increased as predicted in [11], and both the OASPL and M_1 were increased among different vocal tract geometries. Note that the first characteristic peak of /s/ appeared at 6.0, 4.8, 5.6, 3.9, and 6.9 kHz for subject A, B, C, D, and E, respectively, and these values agreed well with those of sustained /s/ recorded with subjects in seated position (without vowel context).

With the simulation for five subjects, we could observe the relationship between the flow and acoustic characteristics for each subject. For further clarification of the cause of the

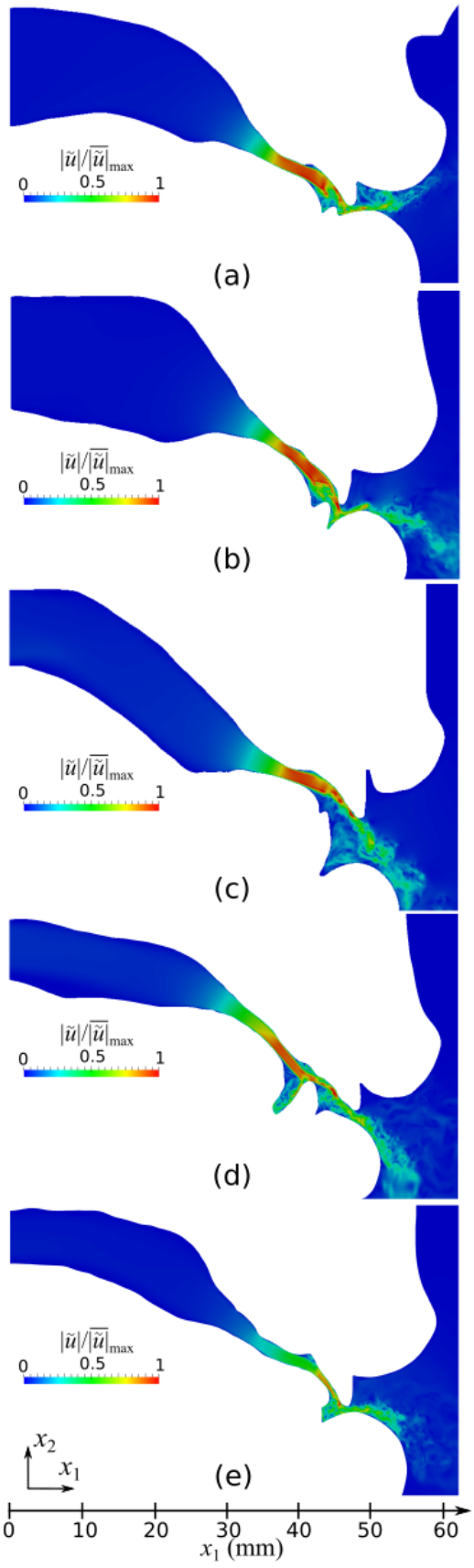

Figure 4: *Velocity magnitude in the vocal tract of subject A (a), B (b), C (c), D (d), and E (e).*

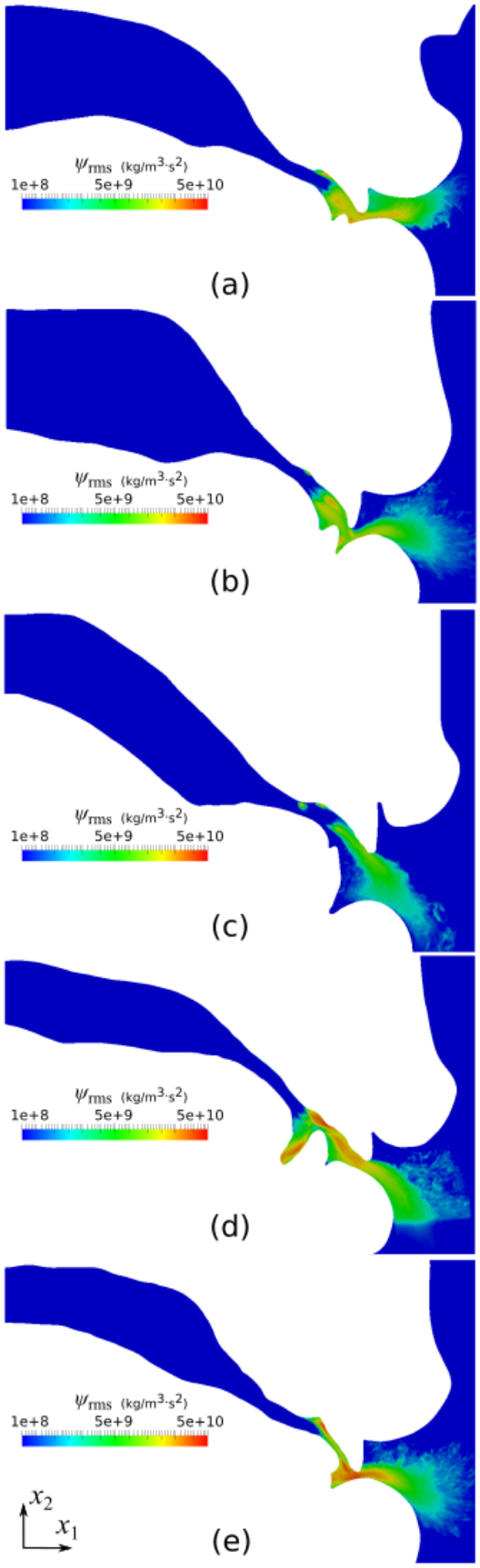

Figure 5: *RMS values of sound source ψ in the vocal tract of subject A (a), B (b), C (c), D (d), and E (e).*

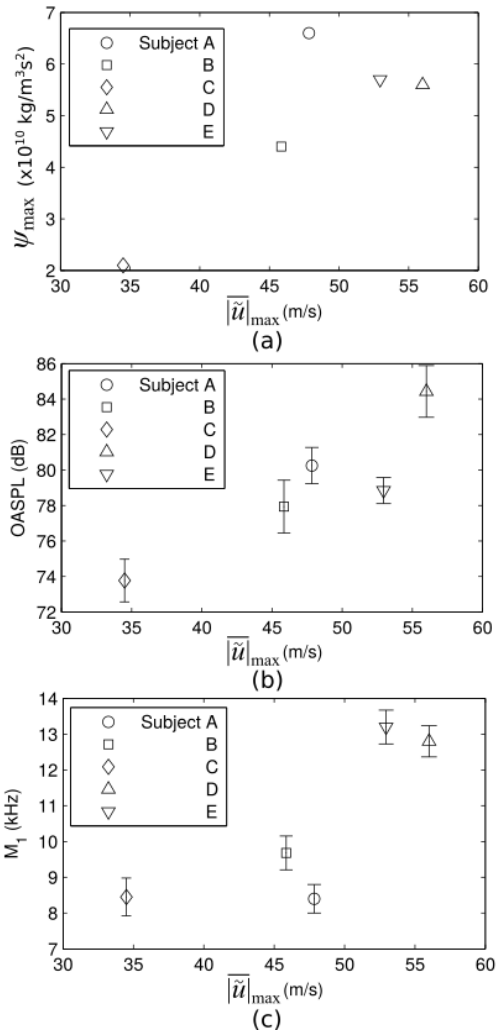

Figure 6: *The maximum amplitude of sound source, OASPL, and M1 vs. the maximum mean velocity at the constriction.*

individual differences, it is desirable to analyse the vocal tract geometry with the simplified model considering three-dimensional vocal tract geometry [12].

4. Conclusion

In this study, the LES of compressible flow was applied to the vocal tract geometry of five subjects extracted from MRI, in order to investigate the individual differences of flow and sound generation of sibilant /s/. We observed the differences of flow and its source configuration in each subject. For the five subjects, the both OASPL and M_1 were increased with the increment of the flow velocity at the constriction and the source amplitude downstream from the constriction.

5. Acknowledgements

This work was supported by MEXT as “Priority Issue on Post-K computer” (hp180202) and JSPS Grant-in-Aid for Scientific Research on Innovative Areas (A18H050770). We would like to thank Dr. Yasuhiro Shimada of ATR-Promotions Inc. for his contribution on the vocal tract imaging.

6. References

[1] K. N. Stevens, *Acoustic Phonetics*, The MIT Press (Cambridge, MA), 1998.

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

[3] M. Howe and R. McGowan, "Aeroacoustics of [s]," *Proceedings of the Royal Society A*, vol. 461 1005-1028, 2005.

[4] T. Yoshinaga, K. Nozaki, and S. Wada, "Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract," *Physics of Fluids* 30, 035104, 2018.

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

[6] H. Takemoto, T. Kitamura, H. Nishimoto, and K. Honda, "A method of tooth superimposition on mri data for accurate measurement of vocal tract shape and dimensions," *Acoustical science and technology* 25, 468–474, 2004.

[7] C. Fureby, "On subgrid scale modeling in large eddy simulations of compressible fluid flow," *Physics of Fluids* 8, 1301–1311, 1996.

[8] M. J. Lighthill, "On sound generated aerodynamically I. General Theory," *Proceedings of the Royal Society A* 211, 564–587, 1952.

[9] Y. Fujiso, K. Nozaki, and A. Van Hirtum, "Estimation of minimum oral tract constriction area in sibilant fricatives from aerodynamic data," *Journal of Acoustical Society of America* 138, EL20–EL25, 2015.

[10] K. Nozaki, T. Yoshinaga, and S. Wada, "Sibilant /s/ simulator based on computed tomography images and dental casts," *Journal of Dental Research* 93, 207–11, 2014.

[11] L. M. Jesus and C. H. Shadle, "A parametric study of the spectral characteristics of European Portuguese fricatives," *Journal of Phonetics* 30, 437-464, 2002.

[12] T. Yoshinaga, K. Nozaki, and S. Wada, "Effects of tongue position in the simplified vocal tract model of Japanese sibilant fricatives /s/ and /ʃ/," *Journal of the Acoustical Society of America* 141, EL314-EL318, 2017.