# A relationship between simplified and realistic vocal tract geometries for Japanese sibilant fricatives

Tsukasa Yoshinaga[1], Kazunori Nozaki[2], Shigeo Wada[1]

[1]*Graduate School of Engineering Science, Osaka University, Osaka, Japan*
[2]*Osaka University Dental Hospital, Osaka, Japan*

t.yoshinaga@biomech.me.es.osaka-u.ac.jp, knozaki@dent.osaka-u.ac.jp
shigeo@me.es.osaka-u.ac.jp

## Abstract

*The production mechanisms of sibilant fricatives have been investigated by using simplified vocal tract models or flow simulation in the realistic vocal tract geometry. In this study, in order to investigate the relationship between the simplified and realistic models, compressible flow simulation was applied to both simplified and realistic vocal tract geometries of Japanese sibilant fricatives. The realistic geometry was constructed from medical images of a subject pronouncing /s/. Cross-sectional areas and heights at five positions in the realistic geometry were used to construct the simplified geometry. Simulated flow fields inside the vocal tracts showed that the flow configuration in the simplified geometry was consistent with that in the realistic geometry. Predicted sound spectra at a far-field point showed that the simplified geometry reproduced the characteristic peaks of the subject's /s/. These results indicate that the flow and acoustic fields in the simplified geometry represent those in the subject's vocal tract geometry pronouncing /s/.*

**Keywords**: speech production, turbulence, vocal tract geometry, sibilant fricatives

## 1. Introduction

Sibilant fricatives are sounds generated by turbulent flow formed by a constriction at the anterior part of the vocal tract (Stevens, 1971). To study the production mechanisms of the sibilant fricatives, several simplified vocal tract models have been proposed. Shadle (1985) proposed a simplified vocal tract model consisted of a constricted channel and obstacle in a cylinder, and experimentally investigated the sound generated by the model. Howe and McGowan (2005) proposed a one-dimensional simplified model, and analytically predicted the far-field sound spectrum of sibilant /s/.

Meanwhile, to examine flow configuration in the realistic vocal tract geometry, flow simulation has been applied to the vocal tract geometry of /s/. Nozaki (2010) conducted large eddy simulation (LES) on the realistic vocal tract geometry of /s/ constructed from computed tomography (CT) images. With the simulation, sound sources occurring on the surface of the lower teeth were visualized.

Based on the flow and acoustic fields obtained with these geometries, the production mechanisms of /s/ have been widely discussed. However, the relationship between the simplified models and sound generation mechanisms in the realistic geometry is still unclear. Therefore, in this study, we conducted the LES on both the simplified and realistic vocal tract geometries of /s/, and investigated the relationship of the flow and acoustic fields in the vocal tract geometries.

## 2. Methodology

### 2.1. Vocal tract geometries

The realistic vocal tract geometry was constructed based on CT images of a male Japanese pronouncing /s/. The subject of the CT images is 32-year-old native Japanese speaker. With the CT scan, 512 sagittal slices of 512 pixel × 512 pixel (isotropic 0.1 × 0.1 × 0.1 mm voxels) were obtained in 9.6 s while the subject sustained /s/ in a seated position. The vocal tract surface was extracted based on CT values, and computational grids were constructed. The computational grids of the realistic geometry are depicted in Figure 1 (a).

The simplified vocal tract geometry was constructed based on the cross-sectional areas and vertical heights at six positions: the pharynx; the constriction; the cervical area of upper teeth; space between upper and lower teeth; and the lip cavity, in the realistic geometry. The computational grids of the simplified geometry are depicted in Figure 1 (b).

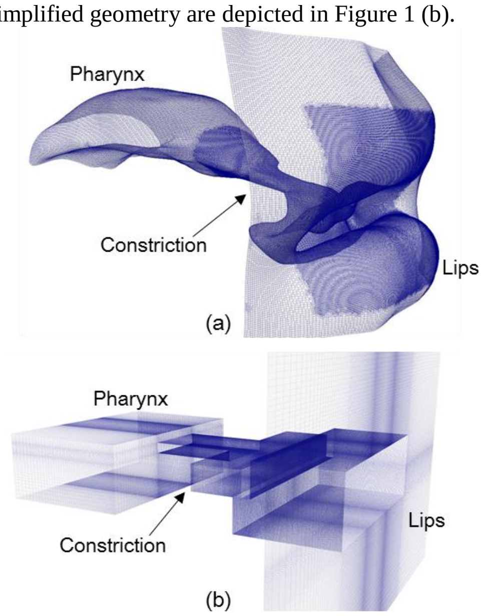


Figure 1: *Computational grids of realistic geometry (a) and simplified geometry (b) for the vocal tact of /s/.*

### 2.2. Computational methods

To simulate both flow and acoustic fields in the vocal tract geometries, we conducted the LES of compressible flow. The one-equation type subgrid-scale model was used to model the turbulent eddy viscosity in a subgrid scale. The spatial derivatives were discretized by the second-order accurate central differencing scheme, and the time integration was performed by the second-order accurate Crank-Nicolson method. The equations were implemented and solved in the finite volume method software OpenFOAM 2.3.1 (OpenCFD Ltd). At the inlet of the vocal tract, the uniform velocity was set to produce the subject's physiological flow rate 300 $cm^3/s$. At the outlet boundary, no reflection boundary condition was imposed. From the simulated pressure fields, the pressure amplitude at a far-filed point 68 mm from the lips was collected, and the sound spectrum was calculated by discretized Fourier transform with six ensemble average of 256-point frames sampled at 100 kHz.

## 3. Results and Discussion

Instantaneous velocity magnitudes in the mid-sagittal plane of the realistic and simplified vocal tract geometries were shown in Figure 2. In the realistic geometry, the maximum velocity appeared at the center of the constriction, and the jet left from the constriction impinged on the upper teeth. Then, the impinged flow separated from the upper teeth surface, and formed large velocity fluctuation nearby the lower lip surface. The flow nearby the lower lip surface traveled towards the upper lips and left the lip cavity. In the simplified geometry, similar flow configuration was observed from the pharynx to the lower lip surface. This result indicates that the simplified vocal tract geometry reproduces most of the flow configuration in the subject's vocal tract pronouncing /s/. Note that the flow configuration in the realistic geometry of this subject is also consistent with that assumed in the previous simplified model (Howe and McGowan, 2005).

The spectra of sound pressure simulated with the realistic and simplified vocal tract geometries are plotted in Figure 3. For the case with the realistic geometry, the first characteristic peak appeared at 5 kHz, and the overall peak appeared at 9 kHz. For the case with the simplified geometry, the first and overall peaks appeared at 5.5 and 8 kHz, respectively. Although the amplitude of the first peak for the simplified geometry was approximately 8 dB larger than that for the realistic geometry, the amplitude above 8 kHz for the simplified geometry was similar in level with that for the realistic geometry (the maximum discrepancy was less than 7 dB). These results suggest that the flow and sound generated in the vocal tract pronouncing /s/ were represented by those in the simplified geometry. Further analysis and evaluation for the two geometries will be presented in the next full paper.

## 4. Acknowledgements

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

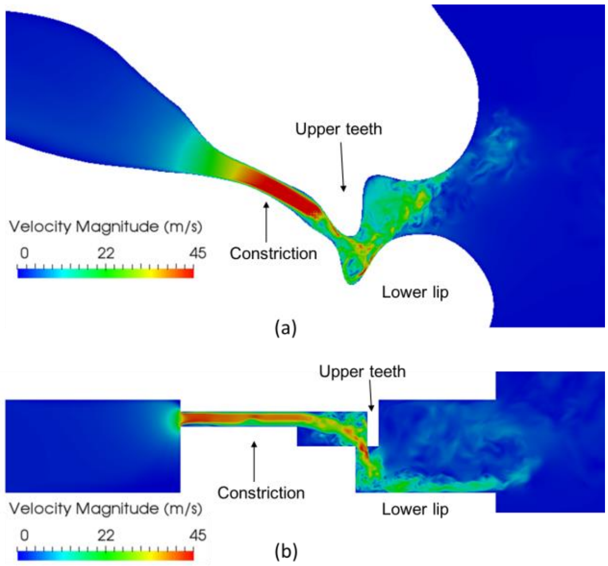


Figure 2: *Mid-sagittal plane of instantaneous flow velocity field in the realistic geometry (a) and the simplified geometry (b).*

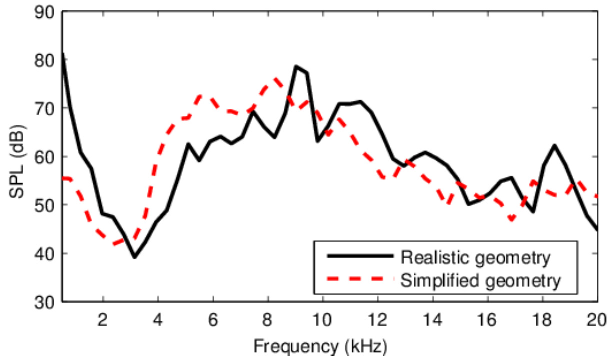


Figure 3: *Spectra of sound pressure at 68 mm from lips simulated with the realistic and simplified vocal tract geometries.*

## 5. References

Howe, M. S. and McGowan, R. S. (2005). "Aeroacoustics of [s]," Proc. R. Soc. A 461, 1005-1028.

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

Shadle, C. H. (1985). "The acoustics of fricative consonants," Ph. D thesis, Massachusetts Institute of Technology, Cambridge, MA, 194 pp.

Stevens, K. N. (1971). "Airflow and turbulence noise for fricative and stop consonants: static considerations," J. Acoust. Soc. Am. 50(4), 1180-1192.