INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Experimental validation of sound generated from flow in simplified vocal tract model of sibilant /s/

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 coupled numerical simulation of flow and sound generation in a simplified vocal tract model of sibilant /s/ were validated with experimental measurements. The simplified model consists of incisors and four rectangular channels representing a throat, constriction, space behind the incisors, and lips. Velocity distribution and far-field sound were measured by a hot-wire anemometer and an acoustic microphone, respectively. Simulated amplitude of velocity fluctuation at the flow separation region was stabilized by increasing the grid resolution, and agreed with those of the measurement. Amplitude of sound pressure simulated by the low-resolution grids was larger than that of the high-resolution grids, indicating that calculation accuracy of velocity fluctuation at the separation region is required to simulate sound generation of the sibilant /s/.

Index Terms: speech production, aeroacoustics, large-eddy simulation, sibilant fricative.

1. Introduction

Sibilant /s/ is one of unvoiced fricative sounds that are generated by turbulent jet flow in the vocal tract. The jet flow occurs from a constricted channel formed by a tongue and upper jaw palate. Impingement of the jet flow on incisors and lips is believe to generate broad-band noise over 4 kHz [1]. The production mechanisms of sibilant /s/ have been investigated by modeling the flow pattern in simplified vocal tract [2-4]. The modeling of turbulent shear layer on the surfaces of incisors [3] demonstrated that the generation of aeroacoustic sound source nearby the incisors forms the sound characteristics of sibilant /s/.

Large-eddy simulation (LES) in the simplified vocal tract models was conducted to uncover the relationship between flow and sound source characteristics in the vocal tract [5-7]. Results of the simulation [6] showed that increase of turbulent intensity enhances magnitude of the sound source at flow-separation region downstream of incisors. However, these studies presented only the distribution of sound source, and did not show the simulation of sound generation. The difficulty in the aeroacoustic simulation is found even in the general simulation for low Mach number flow [8]. The accuracy of simulated flow and sound is largely affected by the grid resolution. Consequently, the experimental validation is required to simulate the flow and sound generation in the simplified vocal tract model.

Therefore, in this study, flow and sound generated by the simplified vocal tract model of sibilant /s/ were simulated for three sets of computational grids. Then, two physical indicators, velocity distribution and sound spectrum of the simulations were validated with experiments using a hot-wire anemometer and an acoustic microphone. This validation will provide the potential requirement for the grid conditions of vocal tract of sibilant fricatives.

2. Material and Method

2.1. Simplified vocal tract model

The simplified vocal tract model consists of incisors and four rectangular flow channel representing a throat, constriction formed by tongue and upper jaw, space behind the incisors, and lip cavity. Overall geometry of the model was illustrated in Figure 1. Dimensions of the channel were determined based on the vocal tract geometry of 32 years old Japanese male subject, measured by Cone-Beam Computed Tomography (CBCT) [9]. The CT images were obtained while the subject sustained the sibilant /s/ for 9.6 s in a seated position. The cross-sectional areas and vertical heights at five positions (throat, constriction, cervical area, gap between incisors, lip cavity) were used to construct the model [7]. It has been confirmed that the model reproduces the subject's sibilant /s/ in the frequency range 3 to 15 kHz.

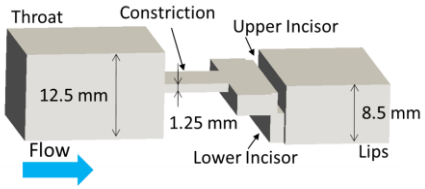

Figure 1. *Schematic illustration for simplified vocal tract model of sibilant /s/.*

2.2. Numerical Simulation

In preliminary calculation, the maximum mean velocity at the center of constriction was approximately U = 50 m/s with physiological flow rate 400 cm^3/s. Thus, Reynolds number ($Re = Ud/\nu$) and Mach number ($M = U/c$) of this simulation were estimated to be 4167 and 0.145, respectively, based on height of the constriction d = 1.25 mm, kinematic viscosity ν = 1.5x10^{-5} m^2/s, and speed of sound c = 344 m/s. Since Mach number is low ($M^2 << 1$), we assume that feedback effects of the generated sound on flow is negligible. Therefore, flow and

Copyright © 2016 ISCA

http://dx.doi.org/10.21437/Interspeech.2016-1597

Table 1. *Parameters for numerical simulations*

Total number of elements	Minimum element size (mm)	Δt	Simulated time (s)	Size of far-field region (cm)
10 million	2.9×10^{-2}	5×10^{-7}	0.005 - 0.018	10 x 6.8 x 6.3
30 million	4.9×10^{-2}	5×10^{-7}	0.028 - 0.045	10 x 6.8 x 6.3
45 million	1.7×10^{-2}	1×10^{-7}	0.016 - 0.020	20 x 20 x 20

sound were simulated separately. First, flow fluctuations were simulated by LES of incompressible fluid as flow source. Then, acoustic fields were simulated by using Lighthill's acoustic analogy in frequency domain [10]:

$$-\nabla^2\rho' - k^2\rho' = M^2\rho_0 \frac{\partial^2 u_i u_j}{\partial x_i \partial x_j}, \quad (1)$$

where ρ' is density fluctuation, k is wave number, ρ_0 is constant density of air, and u_i $(i = 1,2,3)$ is the velocity component at each grid. Sound source in frequency domain (right-hand side in Eq. (1)) was calculated by discretized Fourier transform (DFT) for flow field at each grid. Both flow and acoustic simulations were conducted by second-order scheme finite element method (FEM) softoware FrontFlow/blue ver. 8.1 [11].

The boundary conditions for flow and acoustic simulations were presented in Figure 2. For flow simulation, uniform velocity and constant pressure were set on inlet and outlet boundaries, respectively, to yield steady flow rate 400 cm^3/s at the inlet. No-slip boundary condition was used on surfaces of the model. For acoustic simulation, non-reflecting boundary condition (NRBC) was set on inlet and outlet boundaries, and surfaces of the model were set as ridged wall. Computational parameters for three grid sets were summarized in Table 1. Compare to 10 million grids (10M), minimum element size of 30 million grids (30M) is larger since grid stretching nearby the surface of the model is moderate. In 45 million grids (45M), Δt was smaller than those of other grid sets because of the minimum element size. In addition, 45M has larger far-field region to evaluate the influence of size of the region on sound propagation. At the inlet of the constriction in 30M and 45M, round edge with radius 0.1 mm was formed to smooth the contraction flow between throat and constriction. For every grid condition, DFT of flow source was calculated with average of 4 sets of 32 values extracted with sampling frequency 20 kHz.

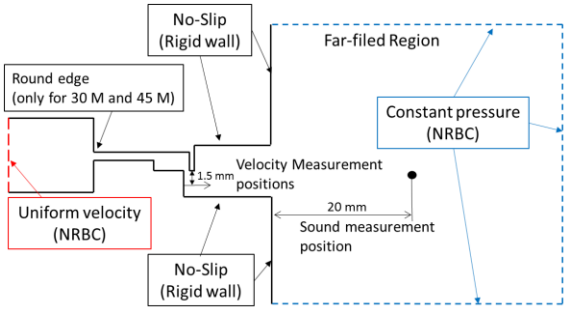

Figure 2: *Boundary conditions for numerical simulations.*

2.3. Experimental setup

The simplified model (illustrated in Figure 1) was constructed by 3D printer (Z printer, 3D systems) with plaster. Steady airflow was inserted to the model by an air compressor (YC-4RS, Yamazaki). Mass flow controller (MQV0050, Azbil) was used to fix the flow rate as 400 cm^3/s. Velocity distribution at 1.5 mm downstream of upper incisor was measured by using a hot-wire anemometer (0251R-T5, Kanomax) which was calibrated in a small wind tunnel (Model 1065, Kanomax) from 2 m/s to 50 m/s. Far-field sound was measured by a microphone (Type 4939, Bruel & Kjaer) at 20 mm from outlet of the model. Signals of velocity and sound were recorded for 1 s with sampling frequency 100 kHz and 44.1 kHz, respectively, by data acquisition system (PXIe-4492, National Instruments). Power spectral density (PSD) of the sound was calculated using fast Fourier transform (FFT) with time average of 4 sets of 30% overlapped 128 point signals convoluted by Hanning window.

3. Results and Discussion

Instantaneous velocity fields between the constriction and the lip cavity in midsagittal plane were shown in Figure 3. In 10M, flow at the constriction became fully turbulent and passed nearby the surface of upper and lower incisors. In 30M and 45M, flow at the constriction became laminar and formed recirculatory flow nearby the surface of upper incisor. This recirculatory flow widened angle between the jet flow and the surface of lower incisor. The difference of flow state at the constriction was caused by the difference of edge shape (*i.e.* angular edge in 10M and round edge in 30M and 45M) at the inlet of the constriction. By making the round edge, flow passed smoothly through the edge of constriction and was not disturbed.

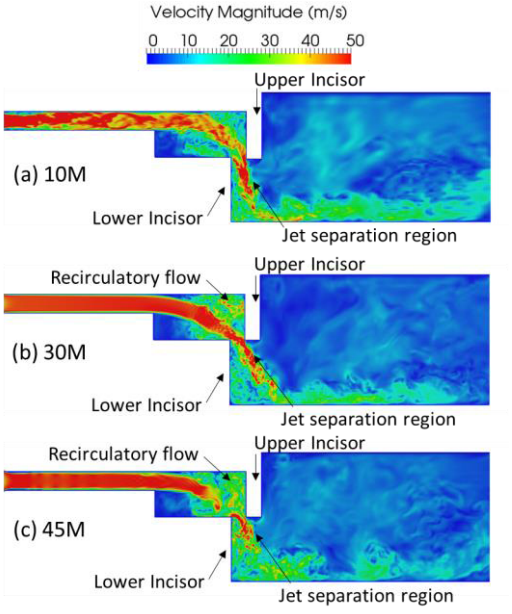

Figure 3. *Instantaneous velocity magnitude in midsagittal plane of (a) 10 million grids, (b) 30 million grids, and (c) 45 million grids.*

Mean and RMS values of velocity distribution at 1.5 mm downstream of upper incisor were plotted in Figure 4 (a) and (b). In order to consider the short simulated time, Mean and RMS values of experiment were calculated for shortened time (*i.e.* 0.004 s) and standard deviations of mean and RMS values

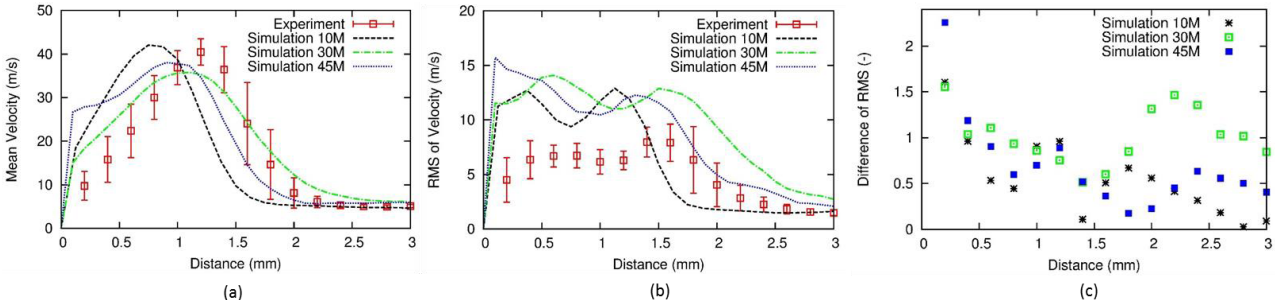

Figure 4. *(a) Mean and (b) RMS of velocity distribution at 1.5 mm downstream of incisors. (c) Difference of RMS values between experiment and simulation. The difference is normalized by experimental values at each measurement positions.*

were calculated for the recorded time (*i.e.* 1 s). Peak of mean velocity measured by the hot-wire anemometer appeared between 1 and 1.4 mm, and position of the peaks in simulation of 30M and 45M agreed with that of experiment. Meanwhile, the peak in 10M appeared at 0.8 mm and this position was different from the experiments. This shift of the peaks in simulation was caused by the flow state upstream of the jet flow as shown in Figure 3. RMS values at separation region of the jet flow (distance 1.5-2.5 mm) in 30M were decreased by increasing the grid resolution to 45M. This tendency agreed with the other case of simulations for flow-separation [8]. The difference of RMS values between experiment and simulation was plotted in Figure 4 (c). The difference in the case of simulation 30M was larger than those of 10M and 45M at the separation region.

Frequency spectra of sound at 20 mm from outlet of the model were shown in Figure 5 (a). The spectral shape of 10M and 45M roughly agreed with that of the experiment in the frequency range 1 to 5 kHz. The difference of PSDs between experiment and simulation was plotted in Figure 5 (b). The differences at the frequency under 3.5 kHz were larger than

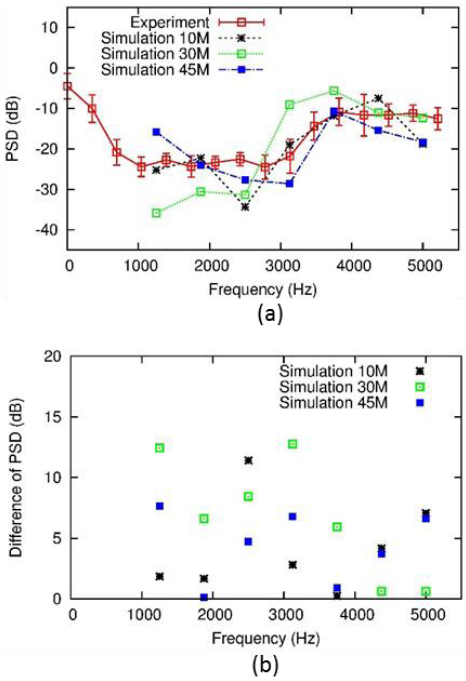

Figure 5. *(a) Frequency spectra of sound at 20 mm from outlet of the model. (b) Difference of PSDs between experiment and simulation.*

those over 3.5 kHz. Those larger values might be decreased by increasing the number of averaging values in DFT of flow source since lower frequency sound consists of longer wave length. Meanwhile, agreement of sound level in 10M indicated that size of far-field region in 10M is enough to simulate the sound spectrum at 20 mm from the outlet of the model.

PSDs at 1.2 kHz and 3.1 kHz simulated in 30M were 12 dB larger than those of experiment. The sound source in the vocal tract is mainly generated by velocity fluctuation in the jet flow downstream of incisors [5-7]. Therefore, this larger PSD was caused by the over prediction of velocity fluctuation (*i.e.* RMS values) at the flow separation region which was caused by the low grid resolution. Noted that PSD at 3.1 kHz simulated in 10M was smaller than that of 30M since minimum element size nearby the separation region of 10M was smaller than that of 30M (see Table 1). Hence, this result indicates that it is significant to increase the grid resolution at the separation region downstream of incisors in order to accurately simulate the sound spectrum of sibilant /s/.

4. Conclusions

The flow and sound simulated in the simplified vocal tract model of sibilant /s/ were validated with experimental measurements. Mean velocity distribution showed that position of the jet flow downstream of incisors was accurately simulated by forming laminar flow with round edge at the constriction. The over prediction of RMS values at the flow separation region downstream of incisors was improved by increasing the grid resolution from 30M to 45M. The result of sound spectrum indicated that over prediction of RMS values in 30M caused the larger PSD at 3.1 kHz. Consequently, high grid resolution to predict the amplitude of velocity fluctuation is needed to simulate the sound spectrum of sibilant /s/. By increasing the number of grids and simulated time, further agreement in flow and sound can be expected. In addition, the same computational methodology could be applied for further analysis on the other fricative consonants, *e.g.* /ʃ/ and /*f*/.

5. Acknowledgements

This research used computational resources of the K computer and other computers provided by the AICS and Cyber Media Center, Osaka University through the HPCI System Research Project (Project ID: hp150274). This research was also supported by JSPS Grant-in-Aid for JSPS Research Fellows (A15J004130), and the Program for Leading Graduate Schools of MEXT (Humanware Innovation Program).We acknowledge Prof. Chisachi Kato for helpful comments and advices.

6. References

[1] P. Badin, "Fricative consonants: acoustic and X-ray measurements," *Journal of Phonetics* vol. 19, 397-408, 1991.

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

[5] J. Cissoni, K. Nozaki, A. Van Hirtum, and S. Wada, "A parameterized geometric model of the oral tract for aero acoustic simulation of fricatives," *International Journal of Information and Electronics Engineering*, vol. 1, 1-3, 2011.

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

[7] T. Yoshinaga, N. Koike, K. Nozaki, and S. Wada, "Study on production mechanisms of sibilant /s/ using simplified vocal tract model," *Proceedings of Inter-noise 2015*, San Francisco, CA, 2015.

[8] C. Kato, Y. Ymade, H. Wang, Y. Guo, M. Miyazawa, T. Takaishi, S. Yoshimura, and Y. Takano, *Computers & Fluids*, vol. 36, 53-68, 2007.

[9] K. Nozaki, T. Yoshinaga, and S. Wada, "Sibilant /s/ simulator based on computed tomography images and dental casts," *Journal of Dental Research*, vol. 93, 207-211, 2014.

[10] A. A. Overai, F. Roknaldin, and T. Hughes, "Computational procedures for determining structuaral-acoustic response due to hydrodynamic sources," *Computer methods in Applied Mechanics and Engeneering*, vol. 190, 345-361, 2000.

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