12th International Symposium on Turbulence and Shear Flow Phenomena (TSFP12)
Osaka, Japan (Online), July 19-22, 2022

Numerical simulation of the sibilant /s/ sound articulation process

HsuehJui Lu
Department of Computational Science
Kobe University
1-1 Rokkodaicho, Nada Ward, Kobe, Hyogo
Lsray@stu.kobe-u.ac.jp

Sakuya Sugimoto
Department of Computational Science
Kobe University
1-1 Rokkodaicho, Nada Ward, Kobe, Hyogo
sugimotodot@stu.kobe-u.ac.jp

Tsukasa Yoshinaga
Department of Mechanical Engineering
Toyohashi University of Technology
1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi
yoshinaga@me.tut.ac.jp

ChungGang Li
Department of Computational Science
Kobe University
1-1 Rokkodaicho, Nada Ward, Kobe, Hyogo
cgli@aquamarine.kobe-u.ac.jp

Kazunori Nozaki
Dental Hospital
Osaka University
1-8 Yamadaoka, Suita, Osaka
knozaki@dent.osaka-u.ac.jp

Akiyoshi Iida
Department of Mechanical Engineering
Toyohashi University of Technology
1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi
iida@me.tut.ac.jp

Makoto Tsubokura
Department of Computational Science
Kobe University
1-1 Rokkodaicho, Nada Ward, Kobe, Hyogo
tsubo@tiger.kobe-u.ac.jp

ABSTRACT

The numerical simulation of the articulation process of the sibilant /s/ sound was conducted with the moving realistic vocal tract geometry of the Japanese word /usui/ using an implicit compressible flow solver. The moving immersed boundary method with the hierarchical structure grid was adopted to approach the complex geometry of the human speech organs. The sound emitted from the vocal tract was analyzed by the spectrogram to clarify the effect of velopharyngeal closure. The sibilant /s/ sound was gradually generated by reducing the constriction size. However, while the velopharyngeal valve was not completely closed, the sound was affected by the nasal emission. After the closure of the valve, a typical /s/ sound was produced by the turbulent airflow in the vocal tract. These results provide the underlying insights necessary for clinical treatment of speech difficulty for sibilant sound.

INTRODUCTION

The sibilant sound is one of the most difficult articulations in the consonant production, and it is known to acquire in the latest leaning process. Therefore, it is often reported in articulation disorders due to the insufficient tongue formation. The sibilant /s/ is generated by the airflow passing through a constriction formed between the alveolar ridge and the tongue tip. When the airflow is forced to go through the narrow constriction (sibilant groove) of the tongue tip, the turbulent jet flow is generated, and the velocity fluctuation downstream from the constriction is considered as the source of the sound. The sibilant fricatives are characterized as broadband noise in the frequency range above the characteristic peaks, approximately 4–7 kHz for /s/.

The velopharyngeal valve, which consists of the velum (soft palate), the lateral and posterior pharyngeal walls, controls the amount of airflow going to the nasal cavity. If the flow escapes through the nasal cavity during the sibilant sound production, the nasal emission occurs. The nasal emission is inaudible with the large velopharyngeal opening. However, when the velopharyngeal opening becomes small, the undesired sound might be generated.

The effect of the velopharyngeal opening and the mechanism of the nasal emission have been investigated by the measurements and numerical simulations. Bunton and Story (2012) examined the relationship among perceptual ratings of the hypernasality, and noted the differences in nasalance, size of the nasal valve opening and perceptual ratings of hypernasality among the three English vowels, /i/, /u/, and /ɑ/. The results indicated that the listeners could detect the hypernasality for the high and low vowels with nasal port areas from 0.01 to 0.15 cm^2, respectively. Moreover, the further work (Bunton, 2015) indicated that the perceptual ratings of the hypernasality and the nasalance increased until nasal port areas reached 0.16 cm^2.

A series of studies (Sundström et al., 2020) was conducted by utilizing the numerical flow simulations to investigate the effect of the velopharyngeal insufficiency and indicated that the ratio of channel areas at the velopharyngeal valve and the oral constriction can be considered as a factor to determine the airflow configuration during the production of sibilant /s/ sound. In addition, the results showed that the acoustic energy in the nasal cavity and the far-field sound were directly related to the size of the velopharyngeal opening. However, the almost all studies of the flow simulation used a fixed vocal tract geometry which only provides the stationary result instead of the time-varying sound production in the actual human speech process.

In the present study, therefore, the numerical simulation of sibilant /s/ production with the realistically moving vocal tract was conducted to investigate the articulation process of velopharyngeal closure and tongue movement. We simulate the sibilant /s/ sound production process in the part of the Japanese

word /usui/ which is extracted after the vowel production of /u/ to the sibilant sound /s/, including the tongue elevation and the velopharyngeal valve closure. The acoustic field was directly predicted with the flow field by the compressible Navier–Stokes equations (Li and Tsubokura, 2017). To simulate the airflow in the complex geometry of the human vocal tract, the moving immersed boundary method with the hierarchical structure grid system was applied. By analyzing the spectrogram of the generated sound, current study clarifies the effect of tongue movement and velopharyngeal closure during the articulation process of the sibilant /s/ production.

GOVERNING EQUATION

The governing equations are the Navier-Stokes equations,

$$\frac{\partial U}{\partial t}+\frac{\partial F_1}{\partial x_1}+\frac{\partial F_2}{\partial x_2}+\frac{\partial F_3}{\partial x_3}=0 \tag{1}$$

where t is the time, x_i indicates three directions in Cartesian coordinate system (i = 1, 2, 3), and the conservative vector U is

$$U=(\rho\ \ \rho u_1\ \ \rho u_2\ \ \rho u_3\ \ \rho e)^T \tag{2}$$

where the flux vectors F_i are

$$F_i=\begin{pmatrix}\rho u_i\\ \rho u_i u_1+P\delta_{i1}-\mu A_{i1}\\ \rho u_i u_2+P\delta_{i2}-\mu A_{i2}\\ \rho u_i u_3+P\delta_{i3}-\mu A_{i3}\\ (\rho e+P)u_i-\mu A_{ij}u_j-k\dfrac{\partial T}{\partial x_i}\end{pmatrix}\quad \forall i=1,2,3 \tag{3}$$

where ρ is the density, u_i indicates the velocity components (i = 1, 2, 3), P is the pressure, δ_{ij} is the Kronecker delta, μA_{ij} is the stress term, e is the total energy, k is the thermal conductivity, and the T is the temperature.

The current framework mainly based on the previous study (Lu et al., 2021). The LUSGS implicit algorithm and the adaptively switched time stepping scheme were applied for the time advancement. The Roe scheme with 5th order MUSCL was used to approximates the convective terms. And the second order central difference method was adopted to calculate the magnitudes of the viscous terms. The immersed boundary method was adopted for the geometry of the realistic human vocal tract and the tongue and velum movement during the phonation process. To reduce the grid generation time for the geometry and to provide better load balancing and higher performance on parallelization, the hierarchical structure grid was applied in the simulations. Using the hierarchical structure grid system can make the working time required to build the computational grids shorter and simultaneously provide better load balancing and higher performance for parallel computations.

VOCAL TRACT MODEL

To simulate the pronouncing process, the geometry of a vocal tract was extracted from 4D-CT images of a subject pronouncing the Japanese word /usui/. The subject was a 42-year-old Japanese male who self-reported no speech disorders with a normal dentition of angle class I. The CT scan was conducted by a 320-row Area Detector CT (20 fps; 320 slices of 512 x 512 pixels). The subject pronounced the Japanese word /usui/ for 1.6 s while the subject was sitting down. The vocal tract geometry was collected from the scans of the phonation process and extracted from the images based on the brightness values of scan, as shown in Figure 1, including the oral cavity, nasal cavity, tongue constriction and velopharyngeal valve. The geometries used in the current study were the duration from the end of vowel /u/ (Figure 1(a) and 1(b)) to the sibilant sound /s/ (Figure 1(c) and 1(d)), which includes the movement of the tongue elevation and the closure of the velopharyngeal valve. The cross-sectional areas of the tongue and velum constrictions are reduced from 121 mm^2 and 55.6 mm^2 to 12.9 mm^2 and 0.01 mm^2, respectively.

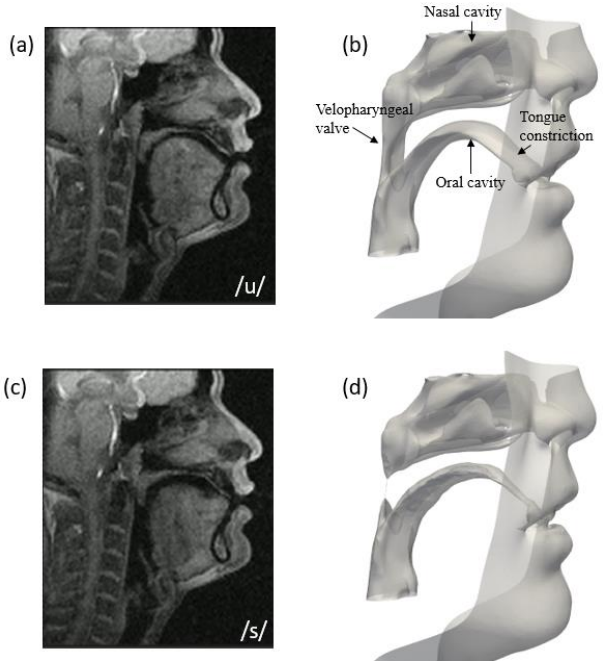

Figure 1: MRI images and vocal tract geometry of (a)(b) the end of vowel /u/ to (c)(d) the sibilant sound /s/ in the phonation process of /usui/.

COMPUTATIONAL CONDITION

After testing the time and space resolutions in the preliminary simulation, the time step was set as 2×10^{-6} s and the minimum grid size was chosen to 0.05 mm near the constriction to keep the accuracy of the immersed boundary around the turbulent region. The total grid number was approximately 1.82×10^8. The grid distribution on the mid-sagittal plane is shown in Figure 2(a). Although the physiologically flow rate varies during the real phonation process, the constant flow rate was set in this simulation to focus on examining the effect of the movement of the speech organs. The uniform inlet velocity was set to 1.5 m/s on the pharynx inlet (the red line shown in Figure 2(a)), which resulted in a physiological flowrate of 450 cm^3/s. Since the sibilant /s/ sound will be generated by the geometry of the vocal tract (Yoshinaga et al., 2019), therefore, there is no voice source set in this simulation.

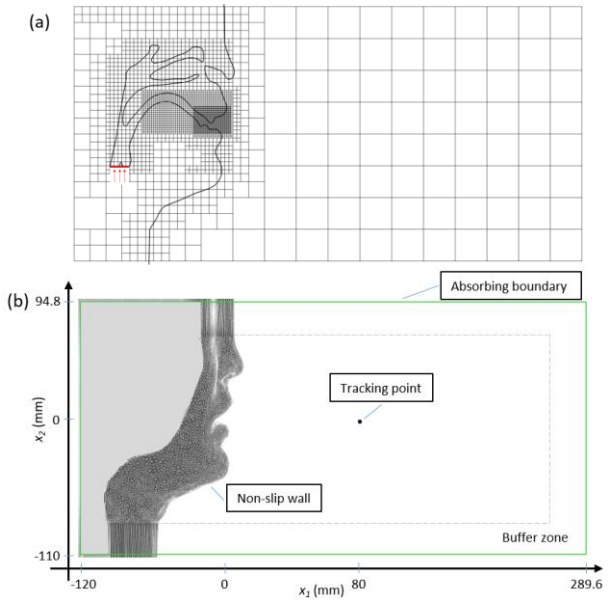

Figure 2: (a) The grid distribution (every 16th gridline is shown for clarity) (b) The computational domain.

The computational domain is shown in Figure 2(b). The x_1 is defined as the anterior-posterior direction; the x_2 is defined as the inferior–superior direction; the x_3 is defined as the transverse direction. To keep the flow in the computational domain from being polluted by reflecting pressure waves at the outer computational boundary, an absorbing boundary condition was used as the outlet condition. Fast Fourier transform (FFT) using the Hann window was applied to the pressure waveforms sampled at 80 mm from the lip outlet to analyze the far-field sound spectrogram. The FFT sampling frequency was 50 kHz with 256 points. The overlap ratio was 0.9 for the spectrogram calculation.

RESULT AND DISCUSSION

The flow and the pressure fluctuation field after the /u/ to the /s/ during the phonation process at the sagittal plane are shown in Figure 3. At t = 0.01 s (Figure 3(a) and 3(d)), the velopharyngeal valve was opened with the cross-sectional area of 55.6 mm^2 which let the airflow go to both nasal and oral cavities. Specifically, there was 35% airflow go through the velopharyngeal valve and 65% airflow goes to the oral cavity. Therefore, the velocity of the mainstream at the oral cavity was only around 10 m/s. In addition, because the tongue was at the lower position, the constriction of oral cavity was not narrow enough to generate the /s/ sound. At t = 0.05 s (Figure 3(b) and 3(e)), due to the closing velopharyngeal valve, the higher flow resistance to the upper chamber decreased the flow rate from 35% to 13% in the nasal cavity and increased the flow rate from 65% to 87% in the oral cavity. Furthermore, the reduction of the cross-sectional area of the oral cavity caused by the elevated tongue position leaded the velocity of the mainstream in the oral cavity was accelerated to around 20 m/s and the periodic pressure wave was observed near the lips in Figure 3(e). Besides, because of the narrow velopharyngeal valve, the velocity became turbulent at the nasal cavity which resulted in the pressure wave at the exit of the nose in Figure 3(e). At t = 0.10 s (Figure 3(c) and 3(f)), since the velopharyngeal valve was completely closed, there was no nasal sound emission observed in pressure fluctuation field. At this time, since the tip of the tongue moved posterior direction and elevated towards the hard palate, the velocity of the mainstream in the oral cavity reached around 25 m/s and the amplitude of the sound was increased at the far-field (Figure 3(f)).

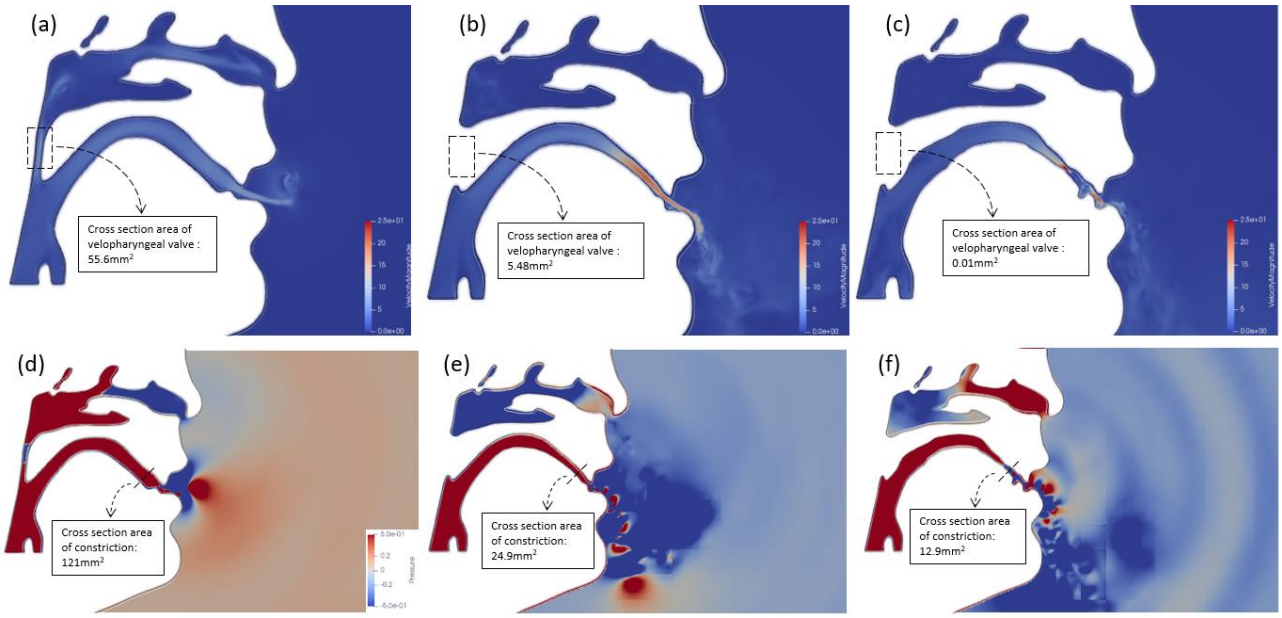

Figure 3: (a-c) Instantaneous flow field and (d-f) pressure fluctuation along the sagittal plane at (a)(d) t = 0.01 s, (b)(e) t = 0.05 s and (c)(f) t = 0.10 s.

The spectrogram of the propagating sound collected at 80 mm from the lips is shown in Figure 4(a). For t = 0.01 s to t = 0.03 s, the amplitude of the broadband noise around 5–9 kHz was observed with relatively small amplitudes due to the opening velopharyngeal valve which leaded to part of the airflow going through the nasal cavity. Besides, since the cross-sectional area of the oral tract was deceased with the tongue elevation, the amplitude peak at 6 kHz gradually appeared after t = 0.02 s. For t = 0.03 s to t = 0.07 s, because of the closing velopharyngeal valve, the broadband noise was amplified in the extended frequency range of 3–11 kHz. After t = 0.08 s, because the size of the velopharyngeal valve became smaller than that of the tongue constriction, the airflow tended to go to the oral cavity rather than the nasal cavity. In addition, since the tongue constriction at this moment was narrow enough for the production of the sibilant sound, the typical /s/ sound was generated with the characteristics peak after 4k Hz, and the broadband noise was observed around 4–12k Hz.

Figure 4(b) shows the measurement result of the phonation process by the real subject (Yoshinaga et al., 2019). Although the frequency range of the /s/ sound is different, the spectrogram during the co-articulation by the current study (t = 0.01 s to t = 0.08 s) is similar to that shown in Figure 4(b). Furthermore, the relation between the size of the velopharyngeal valve and the sound amplitude observed at the far-field was consistent with the previous studies (Kummer, 1992; Sundström, 2020).

CONCLUSION

In the current study, the sibilant /s/ phonation process was simulated with the realistical vocal tract movement of the Japanese word /usui/. By analyzing the spectrogram, the interaction between the sound generated by the tongue movement and the nasal emission caused by the velopharyngeal opening were observed. The sibilant /s/ sound was generated gradually by the reduction of the constriction size with the tongue elevation. However, while the velopharyngeal valve was not completely closed, the sound was influenced by the nasal emission. At the end of the simulation, the typical sibilant /s/ sound was produced by the tongue constriction when the velopharyngeal was completely closed. These results provide clear description of the /s/ sound articulation process which contribute to the clinical treatment for the sibilant sound production.

REFERENCES

Bunton, K., and Story, B. H., 2012, The relation of nasality and nasalance to nasal port area based on a computational model. The Cleft palate-craniofacial journal, 49(6), 741-749.

Bunton, K., 2015, Effects of nasal port area on perception of nasality and measures of nasalance based on computational modeling.

Kummer, A. W., Curtis, C., Wiggs, M., Lee, L., and Strife, J. L., 1992, Comparison of velopharyngeal gap size in patients with hypernasality, hypernasality and nasal emission, or nasal turbulence (rustle) as the primary speech characteristic. The Cleft palate-craniofacial journal, 29(2), 152-156.

Li, C. G., and Tsubokura, M., 2017, An implicit turbulence model for low-Mach Roe scheme using truncated Navier–Stokes equations. Journal of Computational Physics, 345, 462-474.

Lu, H., Yoshinaga, T., Li, C., Nozaki, K., Iida, A., and Tsubokura, M., 2021, Numerical investigation of effects of incisor angle on production of sibilant/s. Scientific Reports, 11(1), 1-11.

Sundström, E., Boyce, S., and Oren, L., 2020, Effects of velopharyngeal openings on flow characteristics of nasal emission. Biomechanics and modeling in mechanobiology, 1-13.

Yoshinaga, T., Nozaki, K., and Wada, S., 2019, A simplified vocal tract model for articulation of [s]: The effect of tongue tip elevation on [s]. PloS one, 14(10), e0223382.

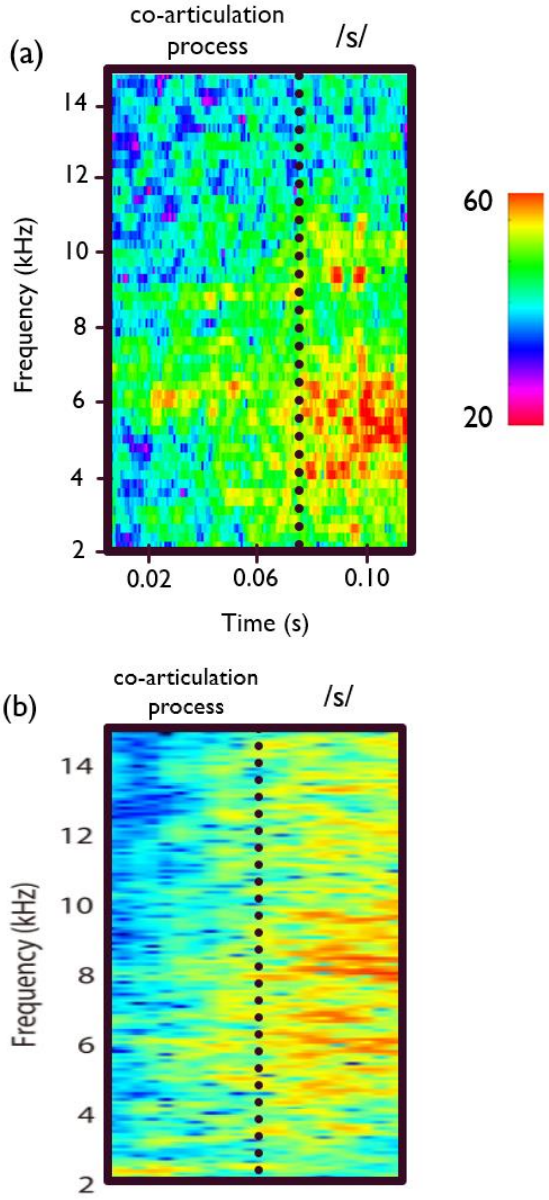

Figure 4. The spectrogram during co-articulation and sibilant /s/ sound (a) by the current study (b) measurement for the subject (Yoshinaga et al., 2019).