



OPEN

Numerical investigation of effects of tongue articulation and velopharyngeal closure on the production of sibilant [s]

HsuehJui Lu^{1,2}, Tsukasa Yoshinaga³, ChungGang Li^{1,2}✉, Kazunori Nozaki⁴, Akiyoshi Iida³ & Makoto Tsubokura^{1,2}

A numerical simulation of sibilant /s/ production with the realistically moving vocal tract was conducted to investigate the flow and acoustic characteristics during the articulation process of velopharyngeal closure and tongue movement. The articulation process was simulated from the end of /u/ to the middle of /s/ in the Japanese word /usui/, including the tongue elevation and the velopharyngeal valve closure. The time-dependent vocal tract geometry was reconstructed from the computed tomography scan. The moving immersed boundary method with the hierarchical structure grid was adopted to approach the complex geometry of the human speech organs. The acoustic characteristics during the co-articulation process were observed and consistent with the acoustic measurement for the subject of the scan. The further simulations with the different closing speeds of the velopharyngeal closure showed that the far-field sound during the co-articulation process was amplified with the slower closing case, and the velum closure speed was inverse proportional to the sound amplitude with the slope value of -35.3 dB s/m. This indicates possible phonation of indistinguishable aeroacoustics sound between /u/ and /s/ with slower velopharyngeal closure.

Speech disorders impact directly a person's social activities and career. In Japan, the number of children who have articulation disorder is estimated approximately 556,000; the number of adults who have articulation disorder is estimated approximately 293,000; the number of people who have stuttering is estimated approximately 710,000 in 2016¹. Although the articulation disorder can be caused by a variety of reasons, the current study put the emphasis on the morphological, and neurological articulation process and investigate the effect of tongue articulation and velopharyngeal closure on the speech production of sibilant /s/ with the numerical simulation.

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 formation². 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^{3,4}. The sibilant fricatives are characterized as broadband noise in the frequency range above the characteristic peaks, approximately 4–7 kHz for /s/⁵.

According to Kummer⁶, the velopharyngeal dysfunction is a speech disorder that the velopharyngeal valve does not close consistently and/or completely during the production of oral consonants. The causes for the velopharyngeal dysfunction are usually reported as abnormal anatomy due to cleft palate which leads to the velopharyngeal insufficiency, abnormal neurophysiology which leads to the velopharyngeal incompetence, and functional articulation errors as the velopharyngeal mislearning. The velopharyngeal valve, which consists of the velum (soft palate) as well as the lateral and posterior pharyngeal walls, controls the amount of airflow going to the nasal cavity. During the normal nasal breathing or the nasal consonant productions, the soft palate rests, and the velopharyngeal valve opens the flow channel. Then, the airflow goes through the nasal cavity, which

¹Computational Fluid Dynamics Laboratory, Department of Computational Science, Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan. ²Complex Phenomena Unified Simulation Research Team, RIKEN, Advanced Institute for Computational Science, Kobe 650-0047, Japan. ³Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan. ⁴Osaka University Dental Hospital, 1-1 Yamadaoka, Suita, Osaka 565-0871, Japan. ✉email: cgli@aquamarine.kobe-u.ac.jp

results in an acoustic coupling between the oral and nasal cavities during the nasal consonant productions. On the contrary, during the production of the oral consonants, such as fricatives and stops (e.g., /s/ and /p/), the complete closure of the velopharyngeal valve is required to increase the oral air pressure. Hence, the velum moves superior and posterior directions to close the velopharyngeal valve against the airflow going to the nasal cavity.

If the flow escapes through the nasal cavity with the velopharyngeal insufficiency during the sibilant sound production, the nasal sound emission occurs due to the jet flow generation at the valve. In this situation, the speech sound is distorted by an acoustic resonance in the nasal cavity. The effects of the velopharyngeal opening and the mechanism of the nasal emission have been investigated by the measurements^{7,8} and numerical simulations^{9–12}. Oren et al.¹³ provided a brief review of the types of nasal emission and the potential causes, and they reported that when there is a characteristic velopharyngeal insufficiency, the audibility of nasal emission can give clues for the relative size of the valve opening. Kummer et al.^{14,15} compared the velopharyngeal gap size in patients with the different level of the perceived characteristics of speech and indicated that the degree of the velopharyngeal insufficiency can be predicted to some extent based on the perceptual assessment. Bunton et al.⁷ 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 /a/. 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 cm², respectively.

Sundström et al.^{9–12} conducted a series of studies by utilizing the numerical flow simulations to investigate the effect of the velopharyngeal insufficiency. To determine the flow behaviour and the sound generating mechanism in the vocal tract, they conducted the simulations of /s/ with the different sizes of the velopharyngeal valve^{9,10}. They confirmed that the turbulence in the nasal cavity indeed affected the speech sound and explained that the additional sound was generated by interaction of the turbulence with the solid nasal wall. In addition, to clarify the relationship between the flow characteristics in the nasal cavity and the size of the velopharyngeal opening, they conducted another numerical simulation by changing the size of the velopharyngeal opening^{14,15} and found that the flow velocity and the turbulence intensity were inversely proportional to the size of the opening if the flow resistance across the velopharyngeal valve was smaller than that across the oral constriction. The series of results 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. However, almost all studies of the flow simulation used a static vocal tract geometry which only provides the stationary result instead of the time-varying sound production in the actual human speech process.

Therefore, in the current study, the numerical simulation of sibilant /s/ production with the realistically moving vocal tract was conducted to investigate the effect of velopharyngeal closure and tongue movement during the articulation process. The speech production process after the vowel /u/ to sibilant /s/ was simulated in the part of the Japanese word /usui/, which is an example of VCV sequence that induces natural movement of the tongue and velum closure. The flow and acoustic field during the phonation process were clearly revealed. Additionally, previous experimental observations indicated that the moving speed and timing of the speech organ is also an important factor in the sound production during the phonation process^{16–18}. Hence, simulations with faster and slower closing velopharyngeal valve were conducted to investigate the effect of the velum closure speed during the sibilant sound. For the first step, the velopharyngeal valve motion was accelerated and decelerated from the original speed within the range of normal articulation speed. By analyzing the spectrogram of the predicted sound, the current study clarifies the effects of velopharyngeal closure speed during the articulation process.

Methods

Vocal tract geometry. To simulate the articulatory process, the geometry of a vocal tract was extracted from four-dimensional computed tomography (4D-CT) images of a subject pronouncing a Japanese word /usui/ (“thin” in English) which using VCV sequence including natural movement of the tongue and velum closure. 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 × 512 pixels). The subject pronounced the Japanese word /usui misosiru/ while the subject was sitting down. The vocal tracts were collected from the scans of the articulatory process and the geometries were extracted from the CT images based on brightness values using the open source software “elasti”¹⁹ using the insight segmentation and registration toolkit (ITK), as shown in Fig. 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/ (Fig. 1a,d) to the sibilant sound /s/ (Fig. 1c,f), which includes the movement of tongue elevation and the closure of the velopharyngeal valve. During the simulation, the geometries at each time step were interpolated from these three geometries (Fig. 1d–f) to mimic the realistic motion of the vocal tract. All methods were performed in accordance with the relevant guidelines and regulations of the ethics committee. The informed consent was obtained from all subjects. The ethics committee of the graduate school of Osaka University certified this study (H26-E39).

The cross-sectional areas of the tongue constriction and the velopharyngeal valve during the movement are shown in Fig. 2a. The preliminary simulation with the stationary geometry was from $t = -0.005$ s to $t = 0$ s to avoid the effects of unstable state on the results of the far-field sound. After $t = 0.0$ s, the cross-sectional areas of the tongue and velum constrictions are reduced from 121 mm² and 55.6 mm² to 12.9 mm² and 0.01 mm², respectively.

To investigate the effect of the velopharyngeal closure speed, the speed of the velum was changed by modifying the cross-sectional area of the velopharyngeal valve at the middle between /u/ and /s/ ($t = 0.05$ s). According to the previous study¹¹, the ratio of area between the velopharyngeal valve and the oral constriction can be considered as a factor for the production of sibilant /s/. Furthermore, the most severe nasal emission was

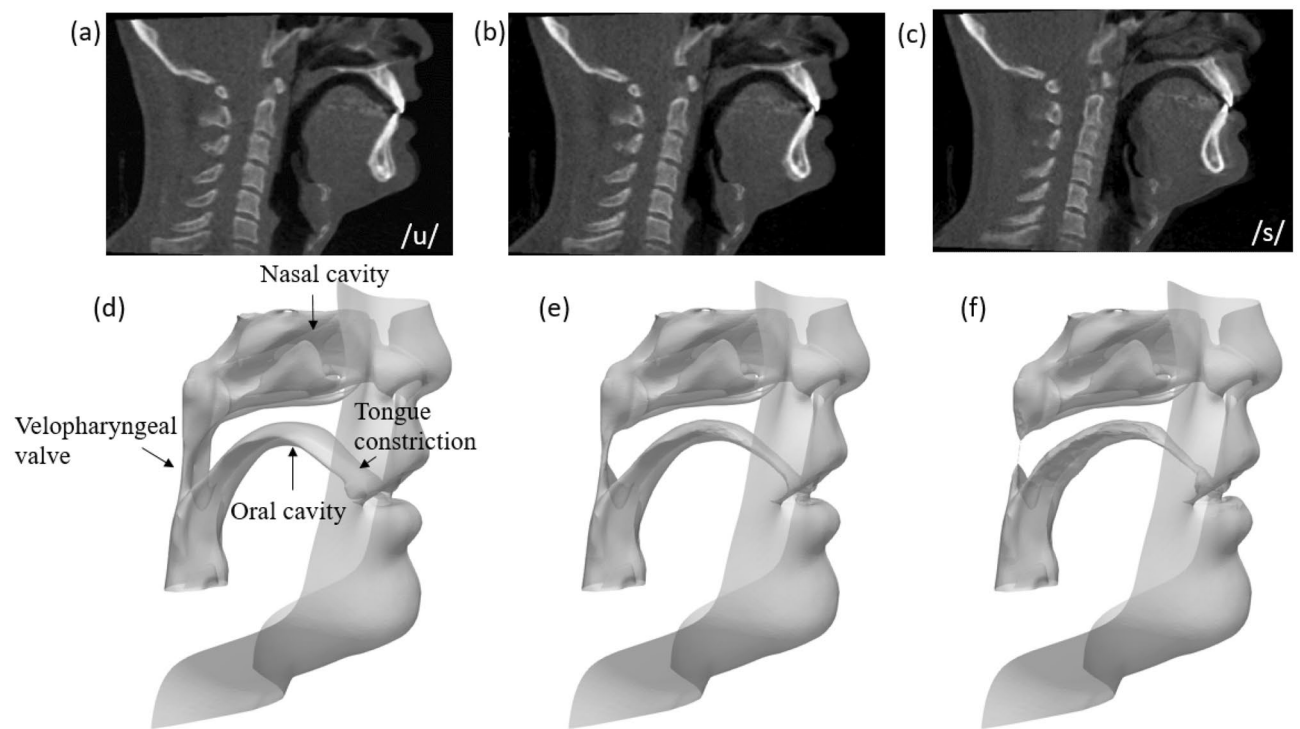


Figure 1. CT images and vocal tract geometry of (a) the end of vowel /u/ to (c) the sibilant sound /s/ in the phonation process of /usui/.

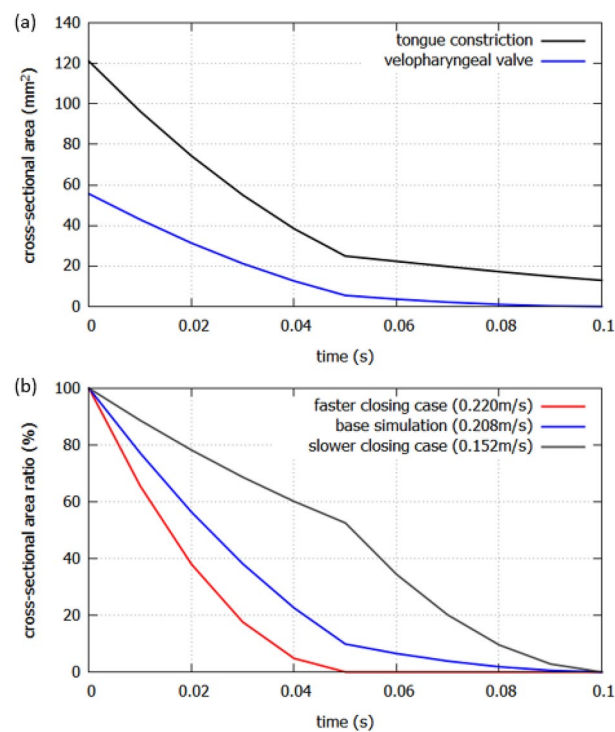


Figure 2. (a) The cross-sectional area of the tongue constriction and the velopharyngeal valve during the movement. (b) The cross-sectional area ratio of different closing speed.

produced when the size of the velopharyngeal opening was equal to the size of the oral constriction. Therefore, the velum opening area was enlarged from 5.48 mm² (original model) to 29.2 mm² which is close to the area of oral constriction for the slower closing case, and reduced to 0.01 mm² for the faster closing case. As shown in Fig. 2b, a faster velopharyngeal closure (red line) and a slower velopharyngeal closure (gray line) were simulated with the original closure speed (blue line). The initial and the end of the velopharyngeal valve size of all cases are set to be the same, however, the velopharyngeal channel was totally closed at $t=0.05$ s in the faster case, while the channel was still opening 50% at $t=0.05$ s in the slower case. The closure speeds of the velum are 0.220 m/s, 0.208 m/s and 0.152 m/s for three cases.

Governing equation. The governing equations are the compressible 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

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

ρ is the density, u_i is the velocity components ($i = 1, 2, 3$), and δ_{ij} indicates the Kronecker delta.

The total energy e is as

$$e = \frac{P}{\rho(\gamma - 1)} + \frac{1}{2}(u_1^2 + u_2^2 + u_3^2), \tag{4}$$

and μA_{ij} is the stress term, while

$$A_{ij} = \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3}(\nabla \cdot u)\delta_{ij}. \tag{5}$$

The pressure P is given from the ideal gas equation:

$$P = \rho RT. \tag{6}$$

The working fluid is air, and the computational condition is standard temperature and pressure²⁰. The dynamic viscosity μ and thermal conductivity k follows Sutherland’s law:

$$\mu(T) = \mu_0 \left(\frac{T}{T_0} \right)^{\frac{3}{2}} \frac{T_0 + 110}{T + 110}, \tag{7}$$

$$k(T) = \frac{\mu(T)\gamma R}{(\gamma - 1) Pr}, \tag{8}$$

where $\rho_0 = 1.1842$ kg/m³, $\mu_0 = 1.85 \cdot 10^{-5}$ N·s/m², $T_0 = 298.06$ K, $\gamma = 1.4$, $R = 287$ J/kg, and the Prandtl number (Pr) is 0.71.

The following numerical framework was adopted to solve the three-dimensional compressible flow governed by Eq. (1). We applied the second-order-accurate implicit lower–upper symmetric Gauss–Seidel scheme (LUSGS) for time integration. The Roe scheme²¹ with a preconditioning method and dual time stepping is adopted, and the discretized form of Eq. (1) with artificial time step $\Delta \tau$ is

$$\begin{aligned} & \Gamma \frac{\bar{U}_p^{k+1} - \bar{U}_p^k}{\Delta \tau} + \frac{3\bar{U}^{k+1} - 4\bar{U}^n + \bar{U}^{n-1}}{2\Delta t} + \frac{1}{\Delta x_1} \left(\bar{F}_1^{k+1(i+\frac{1}{2},j,k)} - \bar{F}_1^{k+1(i-\frac{1}{2},j,k)} \right) \\ & + \frac{1}{\Delta x_2} \left(\bar{F}_2^{k+1(i,j+\frac{1}{2},k)} - \bar{F}_2^{k+1(i,j-\frac{1}{2},k)} \right) + \frac{1}{\Delta x_3} \left(\bar{F}_3^{k+1(i,j,k+\frac{1}{2})} - \bar{F}_3^{k+1(i,j,k-\frac{1}{2})} \right) = 0, \end{aligned} \tag{9}$$

where Γ is the preconditioning matrix proposed by Weiss and Smith²², U_p is the primitive form $[P, u_1, u_2, u_3, T]$, τ and t indicated artificial time and physical times, the superscripts k is the iteration numbers in artificial time step and n is the preceding step of real time. The quantities associated with the artificial time term of the $(k + 1)$ th iteration are transferred approximately to quantities of the $(n + 1)$ th time step in real time when the term $\partial U_p / \partial \tau$ converges to zero. Then, Eq. (9) will be turned back to the original Navier–Stokes equation including the transient term.

Finally, Eq. (1) can be wrote as

$$\left[\frac{I}{\Delta\tau} + \Gamma^{-1}M\frac{3}{2\Delta t} + \Gamma^{-1}(\delta_{x_1}A_p^k + \delta_{x_2}B_p^k + \delta_{x_3}C_p^k) \right] \Delta U_p = \Gamma^{-1}R^k, \tag{10}$$

where $M = \partial U / \partial U_p$, $A_p^k = \partial F_1^k / \partial U_p$, $B_p^k = \partial F_2^k / \partial U_p$, $C_p^k = \partial F_3^k / \partial U_p$ are the flux Jacobian, $R^k = -(3U^k - 4U^n + U^{n-1}) / (2\Delta t) - (\delta_{x_1}\bar{F}_1^k + \delta_{x_2}\bar{F}_2^k + \delta_{x_3}\bar{F}_3^k)$, and δ_{x_i} is the central-difference operator.

The solution-limited time stepping (SLTS) method adaptively adjusting the CFL number and determine $\Delta\tau$ in the governing equation was proposed by Lian et al.²³ to accelerate the convergence speed. When adopting LUSGS method, the estimation value ΔQ_{est} defined as.

$$\Delta Q_{est} = -\Delta\tau M^{-1}[-(3U^k - 4U^n + U^{n+1}) / (2\Delta t) - (\delta_{x_1}\bar{F}_1^k + \delta_{x_2}\bar{F}_2^k)], \tag{11}$$

and ΔQ_{ref} is defined as.

$$\Delta Q_{ref} = \begin{pmatrix} \alpha_1 \times \max[0.5 \times \rho(u_1^2 + u_2^2), \Delta P_{sur}, P_{global} \times 10^{-9}] \\ \alpha_2 \times \max \left[(u_1^2 + u_2^2), \frac{\Delta P_{sur} \times c}{\gamma P}, V_{global} \times 10^{-9} \right] \\ \alpha_3 \times \max \left[(u_1^2 + u_2^2), \frac{\Delta P_{sur} \times c}{\gamma P}, V_{global} \times 10^{-9} \right] \\ \alpha_4 \times T \end{pmatrix} \tag{12}$$

where ΔP_{sur} indicates the maximum difference between the pressure at surrounding points and P_{global} , V_{global} denotes the global value that ensures the reference values are always greater than 0, c is the speed of sound, and γ is the heat capacity ratio. Equation (12) provide a criterion for determining the stability of the calculation. According to Lian²³, the factor $[\alpha_1 \alpha_2 \alpha_3 \alpha_4]$ is the coefficient of the maximum allowable fractional change. Under the SLTS method, the larger physical time step and the faster speed of convergence can be achieved for the fluid simulations. However, SLTS method is not suitable for aeroacoustic simulations even when the convergence criteria are satisfied because of the Newton linearization error of the term $\partial U_p / \partial \tau$. Hence, adaptively switched time stepping scheme (ASTS)^{24,25} was applied to maintain the accuracy in the aeroacoustic simulation at the same time reducing the computational cost.

In the calculation of R^k on the right-hand side of Eq. (10), the terms involving F_i in Eq. (3) can be divided into an inviscid term $F_{inviscid}$, and a viscous term $F_{viscous}$ as shown below:

$$F_{inviscid} = \begin{pmatrix} \rho u_i \\ \rho u_i u_1 + p \delta_{i1} \\ \rho u_i u_2 + p \delta_{i2} \\ \rho u_i u_3 + p \delta_{i3} \\ (\rho e + p) u_i \end{pmatrix}, \tag{13}$$

$$F_{viscous} = - \begin{pmatrix} 0 \\ \mu A_{i1} \\ \mu A_{i2} \\ \mu A_{i3} \\ \mu A_{ij} u_j + \lambda \frac{\partial T}{\partial x_i} \end{pmatrix}, \tag{14}$$

When employed the Roe scheme in Eq. (14), $F_{inviscid}$ term will be discretized into

$$F_{inviscid,i+1/2} = \frac{1}{2} [F_R(U) + F_L(U)] + F_d, \tag{15}$$

where F_d is the Roe dissipation term, which is composed of jumps of properties of work fluids. For the reconstruction of F_R and F_L , we adopted the fifth-order monotone upstream-centered scheme for conservation laws (MUSCL)²⁶ without a limiter function to prevent turbulent fluctuations from attenuating. In addition to the inviscid term, the derivative terms in A_{ij} in the viscous term of Eq. (3) are calculated using the second-order central difference. The detail of the current framework can be found in previous study^{27,28}.

Computational conditions. The three-view diagram of the overall computational domain is shown in Fig. 3a–c. 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. The time step was set as 2×10^{-6} s and the minimum grid size was chosen to 0.05 mm. The grid convergence was tested and the computational accuracy is enough to capture the acoustic characteristics of /s/, as described in the supplementary information. The total grid number was approximately 1.82×10^8 . The grid distribution on the mid-sagittal plane is shown in Fig. 3d. The immersed boundary method was adopted for the geometry of the realistic human vocal tract and the tongue and velum movement during the articulatory process. To reduce the grid generation time for the geometry and to provide better load balancing and higher performance on parallelization, the hierarchical structure grid²⁹ 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.

Although it is different from the physiologically changing flow rate in real phonation process, the constant flow rate was set to focus on the effect of the movement of the speech organ. The uniform inlet velocity was set

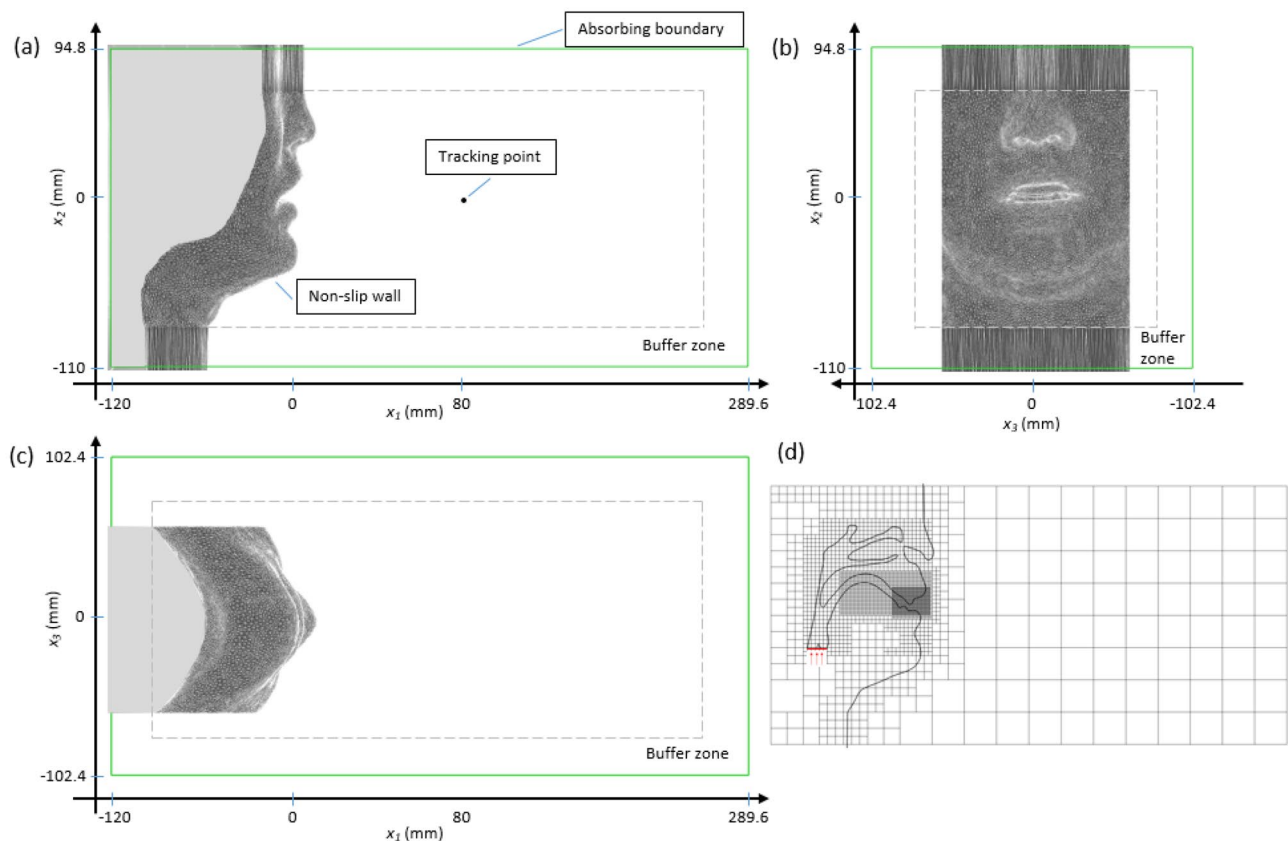


Figure 3. (a)–(c) The three-view diagram of overall computational domain. (d) grid distribution (every 16th gridline is shown for clarity).

to 1.5 m/s as the red line shown in Fig. 3d, which resulted in a physiological flowrate of $450 \text{ cm}^3/\text{s}$. The sibilant /s/ sound will be generated by the geometry of the vocal tract automatically, therefore, there is no voice source set in this simulation. 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. The utilized absorbing boundary condition is based on Freund³⁰ and extended by Li²⁸ which is adjusted for the low Mach number simulation. 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. The sound pressure level (SPL) was calculated based on the reference pressure $P_{ref} = 20 \times 10^{-6} \text{ Pa}$.

Results and discussion

The instantaneous flow field and the pressure fluctuation along the sagittal plane from /u/ to /s/ are shown in Fig. 4 for the original motion of the CT scans with the velum speed of 0.208 m/s. At $t = 0.01 \text{ s}$ (Fig. 4a,d), the tongue was at the lower position so that the size of the tongue constriction was not small enough to produce the sound of /s/. In addition, the velopharyngeal valve was opened 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 and there was no sound propagation from both nasal and oral cavities. At $t = 0.05 \text{ s}$ (Fig. 4b,e), since the velopharyngeal valve kept closing the channel to the nasal cavity, the higher flow resistance 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. Moreover, the elevated tongue position caused the reduction of the cross-sectional area of the oral cavity so that 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. 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 Fig. 4e. At $t = 0.1 \text{ s}$ (Fig. 4c,f), 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. At this time, since the velopharyngeal valve was completely closed, there was no nasal sound emission.

The spectrogram of the propagating sound collected at 80 mm from the lip outlet, is shown in Fig. 5a. For $t = 0.01 \text{ s}$ to $t = 0.03 \text{ s}$, since the cross-sectional area of the oral tract was decreased with the tongue elevation and velopharynx closure, the amplitude peak at 6 kHz gradually appeared after $t = 0.02 \text{ s}$. The amplitude of the broadband noise around 5–9 kHz was observed with relatively small amplitudes since the velopharyngeal valve was still open and some of the airflows kept going through the nasal cavity. For $t = 0.03 \text{ s}$ to $t = 0.07 \text{ s}$, because of the

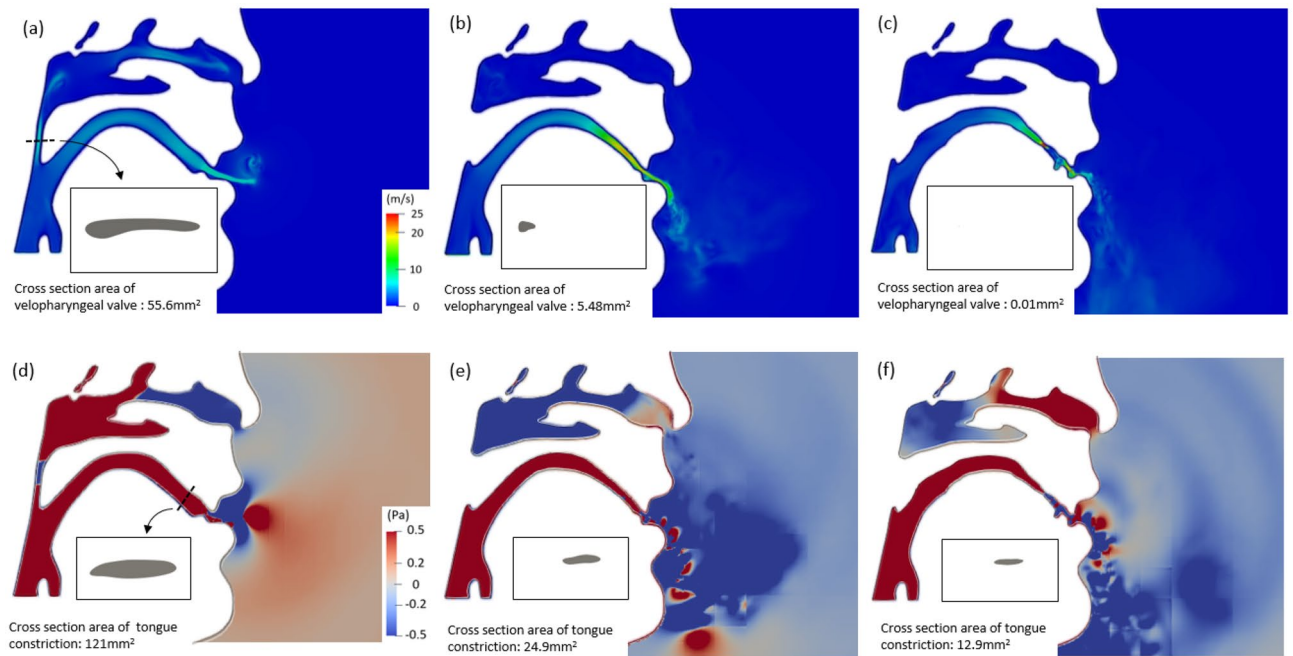


Figure 4. (a)–(c) Instantaneous flow field and (d)–(f) pressure fluctuation along the sagittal plane at (a, d) $t=0.01$ s, (b, e) $t=0.05$ s and (c, f) $t=0.10$ s.

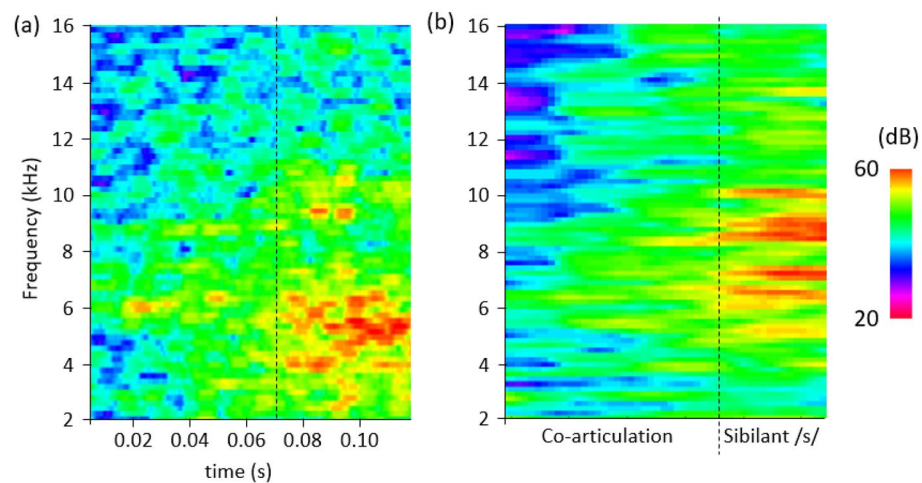


Figure 5. The spectrogram during co-articulation and sibilant /s/ sound (a) by the current study (b) measurement for the subject of the CT scan²⁵.

closing velopharyngeal valve, the broadband noise was amplified in the extended frequency range of 3–11 kHz. After $t=0.07$ s, when 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. Since the tongue constriction at this moment was narrow enough for the production of the sibilant sound, the typical sibilant sound was generated with the characteristics peak around 4–6 kHz, and the broadband noise was observed around 4–12 kHz. Although the frequency range of the /s/ sound is different, the spectrogram during the co-articulation is similar to that shown in Fig. 5b which was pronounced by the subject of CT scan³¹. 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^{11,14,15}.

To investigate the effect of the closing speed of the velopharyngeal valve, simulations with closing speeds of 0.220 m/s, 0.208 m/s and 0.152 m/s were conducted with the cross-sectional area ratio shown in Fig. 2b. To clarify the relationship between the sound amplitude and velum closing speed, the spectrograms of the different closing speeds are shown in Fig. 6 based on the velopharyngeal valve opening. The horizontal axis is plotted with the velum opening ratio, and the complete closure (0% opening) was around $t=0.05$ s for the case with 0.220 m/s, $t=0.09$ s for the original case, and $t=0.10$ s for the case with 0.152 m/s. As seen in the figure, at 80% valve opening

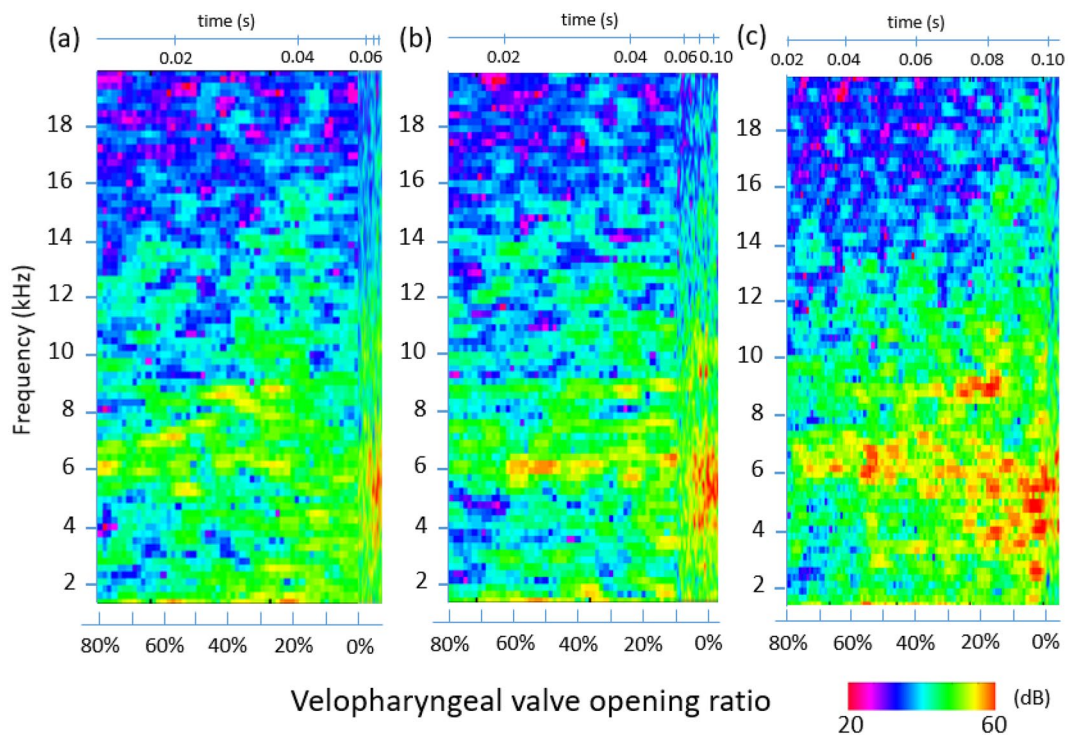


Figure 6. The far-field spectrogram with the closing speed of (a) 0.220 m/s (faster closing case), (b) 0.208 m/s (base simulation) and (c) 0.152 m/s (slower closing case) base on the velopharyngeal valve opening ratio.

ratio, peaks were mainly observed around 6 kHz in all cases. While the valve opening ratio reached 20%, the frequency range of the broadband noise was increased to 3–10 kHz in all cases. Although the amplitudes of the three cases were different, the tendency of the broadband noise indicated that the sound is strongly related to the opening of the velopharyngeal valve. Besides, the amplitudes of 0.152 m/s in the frequency range 3–7 kHz were larger than those of 0.208 m/s and 0.220 m/s.

The overall sound pressure level (OASPL) values from 1.5 to 20 kHz during the articulatory process are shown in Fig. 7a. The OASPL of the 0.220 m/s was larger than the other cases around $t = 0.01$ s, because of the higher flow rate through the oral cavity caused by the higher flow resistance in the nasal cavity. However, the OASPL of 0.152 m/s became larger than that of 0.220 m/s after $t = 0.04$ s. The amplitude of 0.220 m/s at this moment became the lowest since the velopharyngeal valve was totally closed. On the contrary, because the channel to the nasal cavity was still 50% opened for 0.152 m/s, the audible nasal emission made the OASPL higher at $t = 0.05$ s. To identify the effect of the velum closure speed, the OASPL based on the velopharyngeal valve opening ratio was shown in Fig. 7b. As shown in the figure, the amplitude was amplified for 0.152 m/s in the opening ratio from 75 to 5%. The mean OASPL of the 0.220 m/s, 0.208 m/s, and 0.152 m/s are 61.8 dB, 62.1 dB, and 64.2 dB, respectively, during the articulatory process ($t = 0-0.1$ s). At the velum opening of 55%, the OASPL of 0.152 m/s was 5 dB louder than that of 0.220 m/s because the area ratio of velopharyngeal to oral channel was larger in slower closing case.

Moreover, the relation between the mean OASPLs and the velum closure speed was plotted in Fig. 8. The relation was in the inverse proportion with the slope equaled to -35.3 dB·s/m. This indicated that the sound generated before the /s/ will be enlarged while the slower closing speed of the velopharyngeal valve. This amplified sound might be related to area ratio of velopharyngeal to oral constriction mentioned by Sundström¹¹. Consequently, the slower velum closure speed will result in the larger sound amplitude during the co-articulation and might disturb the acoustic characteristics of pronounce sounds. Although we performed the simulation only for this subject, further simulation with additional patients with hypernasality may help the diagnosis of the velopharyngeal insufficiency.

Conclusion

To investigate the effects of the velopharyngeal closure speed and the tongue constriction formation on the sibilant sound production, a numerical simulation of the articulatory process after the vowel /u/ to the sibilant sound /s/ was conducted. The broadband noise generation during the co-articulation process was observed and consistent with the measurement result. To clarify the effect of velopharyngeal closure speed, the simulations with different velum closing speed were conducted. The spectrograms based on the velopharyngeal valve opening showed that the frequency range of the broadband noise of /s/ was extended from 6 to 3–7 kHz during the co-articulation process in all cases, and this indicated that the sound amplitude is strongly related to the velopharyngeal opening. In addition, the sound amplitude increased in the case with the slower velum closure

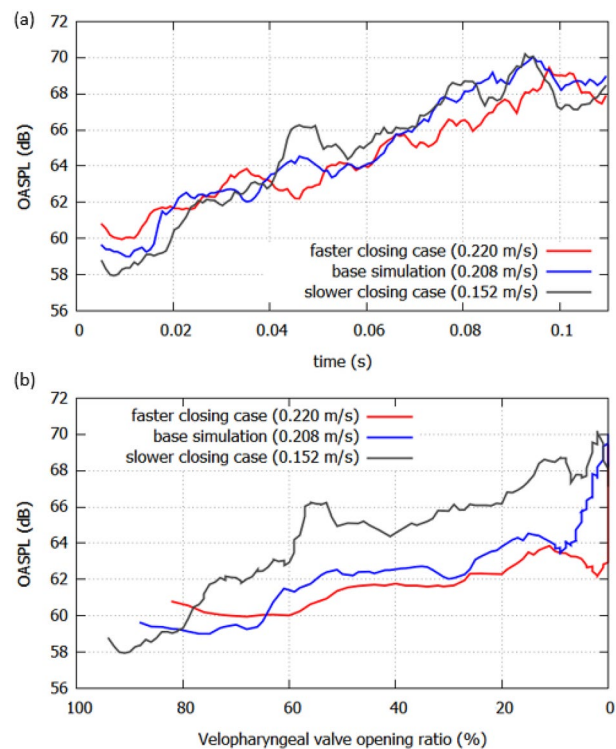


Figure 7. (a) The OASPLs with the different closing speed. (b) The OASPLs based on the velopharyngeal valve opening ratio.

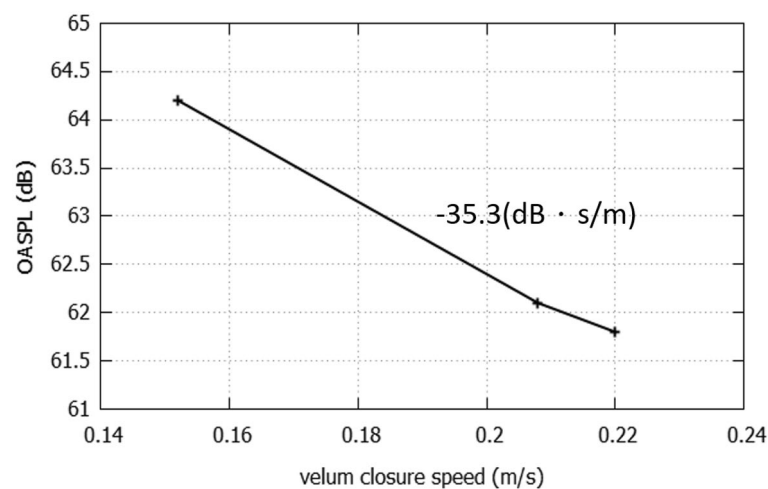


Figure 8. The relation between OASPLs during the phonation process and the velum closure speed.

during the co-articulation process, indicating the possible emission of indistinguishable sound. These findings can provide the underlying insights necessary to the oral surgical treatment for generating the sibilant sound. Moreover, the present framework will enable us to predict the effects of the vocal tract shapes for more subjects on the production of sibilant fricatives with shorter time to build the database which benefits patients prior to prosthetic surgery.

Data availability

The datasets generated or analysed during the current study are available from the corresponding author on reasonable request.

Received: 16 February 2022; Accepted: 18 August 2022

Published online: 13 September 2022

References

- Kariyasu, M., Toyama, M. & Matsuhira, T. Epidemiology of communication disorders—Prevalence and estimates of the number of speech-language-hearing disabled. *KUAS Dep. Bull. Paper* **1**, 1–12 (2016) **(in Japanese)**.
- Wood, S., Wishart, J., Hardcastle, W., Cleland, J. & Timmins, C. The use of electropalatography (EPG) in the assessment and treatment of motor speech disorders in children with Down's syndrome: evidence from two case studies. *Dev. Neurorehabil.* **12**(2), 66–75 (2009).
- Shadle, C. H. Articulatory-acoustic relationships in fricative consonants. In *Speech Production and Speech Modelling* 187–209, Springer, Dordrecht (1990).
- Jongman, A., Wayland, R. & Wong, S. Acoustic characteristics of English fricatives. *J. Acous. Soc. Am.* **108**(3), 1252–1263 (2000).
- Badin, P. Fricative consonants: Acoustic and X-ray measurements. *J. Phon.* **19**(3–4), 397–408 (1991).
- Kummer, A. W. Types and causes of velopharyngeal dysfunction. In *Seminars in Speech and Language*, Vol. 32, No. 02, 150–158 (© Thieme Medical Publishers, 2011).
- Bunton, K. & Story, B. H. The relation of nasality and nasalance to nasal port area based on a computational model. *Cleft Palate Craniofac. J.* **49**(6), 741–749 (2012).
- Bunton, K. Effects of nasal port area on perception of nasality and measures of nasalance based on computational modeling(2015).
- Sundström, E. & Oren, L. Pharyngeal flow simulations during sibilant sound in a patient-specific model with velopharyngeal insufficiency. *J. Acoust. Soc. Am.* **145**(5), 3137–3145 (2019).
- Sundström, E. & Oren, L. Sound production mechanisms of audible nasal emission during the sibilant/s. *J. Acoust. Soc. Am.* **146**(6), 4199–4210 (2019).
- Sundström, E., Boyce, S. & Oren, L. Effects of velopharyngeal openings on flow characteristics of nasal emission. *Biomech. Model. Mechanobiol.* **19**, 1–13 (2020).
- Sundström, E. & Oren, L. Change in aeroacoustic sound mechanism during sibilant sound with different velopharyngeal opening sizes. *Med. Biol. Eng. Comput.* **59**(4), 937–945 (2021).
- Oren, L., Kummer, A. & Boyce, S. Understanding nasal emission during speech production: A review of types, terminology, and causality. *Cleft Palate Craniofac. J.* **57**(1), 123–126 (2020).
- Kummer, A. W., Curtis, C., Wiggs, M., Lee, L. & Strife, J. L. Comparison of velopharyngeal gap size in patients with hypernasality, hypernasality and nasal emission, or nasal turbulence (rustle) as the primary speech characteristic. *Cleft Palate Craniofac. J.* **29**(2), 152–156 (1992).
- Kummer, A. W., Briggs, M. & Lee, L. The relationship between the characteristics of speech and velopharyngeal gap size. *Cleft Palate Craniofac. J.* **40**(6), 590–596 (2003).
- Moll, K. L. & Daniloff, R. G. Investigation of the timing of velar movements during speech. *J. Acoust. Soc. Am.* **50**(2B), 678–684 (1971).
- Dalston, R. M. & Seaver, E. J. Nasometric and phototransductive measurements of reaction times among normal adult speakers. *Cleft Palate J.* **27**(1), 61–67 (1990).
- Warren, D. W., Dalston, R. M. & Mayo, R. Hypernasality in the presence of “adequate” velopharyngeal closure. *Cleft Palate Craniofac. J.* **30**(2), 150–154 (1993).
- Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. Elastix: A toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2009).
- Bejan, A. *Convection Heat Transfer* (Wiley, 2013).
- Li, C. G. & Tsubokura, M. An implicit turbulence model for low-Mach Roe scheme using truncated Navier–Stokes equations. *J. Comput. Phys.* **345**, 462–474 (2017).
- Weiss, J. M. & Smith, W. A. Preconditioning applied to variable and constant density flows. *AIAA J.* **33**(11), 2050–2057 (1995).
- Lian, C., Xia, G. & Merkle, C. L. Solution-limited time stepping to enhance reliability in CFD applications. *J. Comput. Phys.* **228**(13), 4836–4857 (2009).
- Lu, H. *et al.* Numerical investigation of effects of incisor angle on production of sibilant /s/. *Sci. Rep.* **11**(1), 1–11 (2021).
- Lu, H., Li, C. & Tsubokura, M. Adaptively switched time stepping scheme for direct aeroacoustic computations. *AIP Adv.* **12**(3), 035340 (2022).
- Kim, K. H. & Kim, C. Accurate, efficient and monotonic numerical methods for multi-dimensional compressible flows: Part II: Multi-dimensional limiting process. *J. Comput. Phys.* **208**(2), 570–615 (2005).
- Li, C. G. A compressible solver for the laminar–turbulent transition in natural convection with high temperature differences using implicit large eddy simulation. *Int. Commun. Heat Mass Transf.* **117**, 104721 (2020).
- Li, C. G., Tsubokura, M., Fu, W. S., Jansson, N. & Wang, W. H. Compressible direct numerical simulation with a hybrid boundary condition of transitional phenomena in natural convection. *Int. J. Heat Mass Transf.* **90**, 654–664 (2015).
- Nakahashi, K. Building-cube method for flow problems with broadband characteristic length. In *Computational Fluid Dynamics 2002*, pp. 77–81 (Springer, Berlin, 2003).
- Freund, J. B. Proposed inflow/outflow boundary condition for direct computation of aerodynamic sound. *AIAA J.* **35**(4), 740–742 (1997).
- Yoshinaga, T., Nozaki, K. & Wada, S. A simplified vocal tract model for articulation of [s]: The effect of tongue tip elevation on [s]. *PLoS ONE* **14**(10), e0223382 (2019).

Acknowledgements

This work was supported by the RIKEN Junior Research Associate Program, JSPS KAKENHI (Grant numbers: JP19H04124, JP19H03976 and JP20H01265) and JST CREST (Grant Number: JPMJCR20H7).

Author contributions

H.J.L., T.Y., C.G.L., K.N., A.I., and M.T. designed the project. M.T. and A.I. managed the project. T.Y., and K.N. prepared the experimental setups. T.Y. performed the experiments. K.N. organized the subject measurements. H.J.L. and C.G.L. conducted the numerical simulations and analyzed the data. The paper was mainly written by H.J.L., incorporating all authors' comments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18784-7>.

Correspondence and requests for materials should be addressed to C.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022