


Downstream flow field structure in voice prosthesis and its effect on sound generation around the esophageal wall

Cite as: Phys. Fluids **35**, 025114 (2023); <https://doi.org/10.1063/5.0134627>

Submitted: 12 November 2022 • Accepted: 19 January 2023 • Accepted Manuscript Online: 20 January 2023 • Published Online: 07 February 2023

 Seiya Kosako (小迫誠弥), Mariko Hiramatsu (平松真理子), Yasushi Fujimoto (藤本保志), et al.



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[A POD-DMD augmented procedure to isolating dominant flow field features in a street canyon](#)

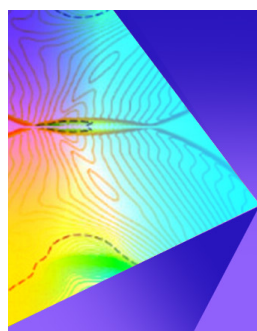
Physics of Fluids **35**, 025112 (2023); <https://doi.org/10.1063/5.0133375>

[Reactive-dynamic characteristics of a nanobubble collapse near a solid boundary using molecular dynamic simulation](#)

Physics of Fluids **35**, 022003 (2023); <https://doi.org/10.1063/5.0139169>

[Discovering explicit Reynolds-averaged turbulence closures for turbulent separated flows through deep learning-based symbolic regression with non-linear corrections](#)

Physics of Fluids **35**, 025118 (2023); <https://doi.org/10.1063/5.0135638>



Physics of Fluids

Special Topic: Shock Waves

Submit Today!

Downstream flow field structure in voice prosthesis and its effect on sound generation around the esophageal wall

Cite as: Phys. Fluids **35**, 025114 (2023); doi: [10.1063/5.0134627](https://doi.org/10.1063/5.0134627)
Submitted: 12 November 2022 · Accepted: 19 January 2023 ·
Published Online: 7 February 2023




View Online



Export Citation



CrossMark

Seiya Kosako (小迫誠弥),^{1,a)}  Mariko Hiramatsu (平松真理子),² Yasushi Fujimoto (藤本保志),³
and Yoshiyuki Tsuji (辻義之)

AFFILIATIONS

¹Department of Energy Engineering, Graduate School of Engineering Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8603, Japan

²Department of Otorhinolaryngology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya-shi, Aichi 466-8550, Japan

³Department of Otorhinolaryngology, Aichi Medical University, 1-1 Yazakokarimata, Nagakute-shi, Aichi 480-1195, Japan

^{a)} Author to whom correspondence should be addressed: s-kosako@energy.nagoya-u.ac.jp

ABSTRACT

This study investigates the effects of downstream flow in the voice prosthesis (Provox® Vega) on the esophageal wall during shunt vocalization to identify the dominant factors influencing the sound generated in the esophagus (the original sound in shunt vocalization). The trachea and esophagus were simulated to measure flow in the esophageal section through the voice prosthesis by using particle image velocimetry, and pressure distribution in the esophageal section was calculated by solving Poisson's equation. Downstream flow in the voice prosthesis fluctuated over time due to oscillations of the valve. This flow impinged on the esophageal wall and created a strong shear region that fluctuated over time to give rise to fluctuations in pressure at the same frequency as that of oscillations of the valve on the esophageal wall near the point of impingement. The sound pressure, estimated by using the measured velocity field based on the Ffowcs Williams–Hawkings equation, was compared with the sound pressure measured by a microphone. The results showed that temporal variations in pressure downstream of the voice prosthesis and its impinging on the esophageal wall were the main sources of the sound generated from the voice prosthesis within the esophageal section. The frequency of the generated sound depended on the frequency of oscillations of the valve.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0134627>

I. INTRODUCTION

Human speech originates in sounds generated in the larynx and the propagation of their resonance through the vocal tract. The exhalation of air from the lungs causes the larynx to oscillate and sound to be emitted from the mouth [Fig. 1(a)]. The larynx contains a fold-like tissue called the vocal cords. The vocal cords vibrate in response to expiratory flow from the lungs during vocalization, and the original speech sound is generated by the disturbance in expiratory flow due to these vibrations (Kaburagi *et al.*, 2010). In a previous paper, we examined the flow of air and sound near a vibrating body similar to the vocal cords and found that sound was emitted due to a disturbance in flow after it passed through the vibrating body. This indicates that the interference in expiratory flow in the body is important for the origination of sound (Kosako *et al.*, 2021).

A total laryngectomy is required in the case of severe laryngeal or hypopharyngeal cancer and results in the loss of voice. This significantly degrades the patient's postoperative quality of life (QoL). In case a laryngectomy is carried out, the trachea and esophagus are completely separated, and a tracheostoma is surgically created to enable breathing on the anterior surface of the throat while keeping the windpipe open. Shunt vocalization is one method that is used to help patients regain their voice, in which the tracheal airway and esophagus are connected by a one-way valve called a voice prosthesis, as shown in Figs. 1(b) and 1(c). During shunt vocalization, the tracheostoma is blocked by a finger or some other object, and expiratory air-flow from the lungs is pumped into the esophagus through the voice prosthesis, represented by the blue line in Fig. 1(b). An enlarged view around the Provox Vega is shown in Fig. 1(c). During this process, the

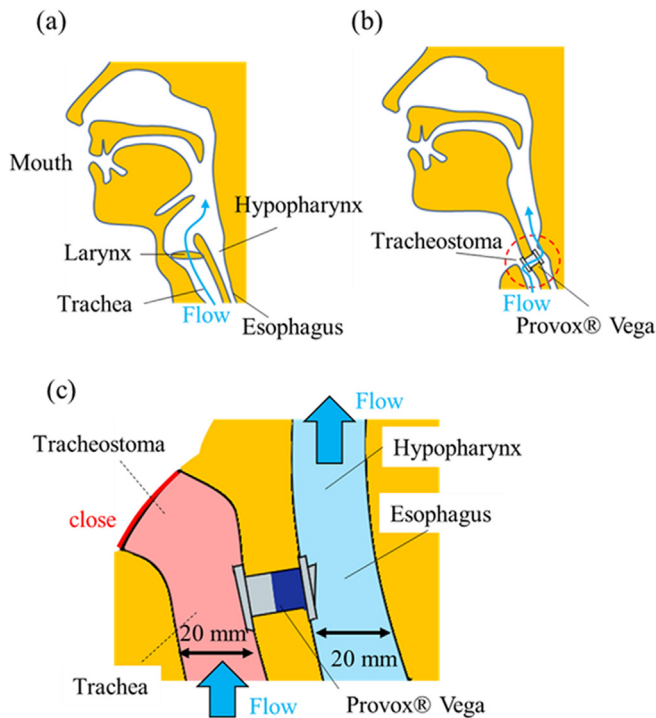


FIG. 1. Schematic of the sagittal plane of the head and neck region: (a) normal vocalization, (b) shunt speech with voice prosthesis, and (c) schematic view of the trachea and esophagus connected by a voice prosthesis.

airflow due to exhalation causes the mucous membrane of the hypopharynx, located just above the esophagus, to vibrate, and this generates a sound in place of the larynx (Tateda *et al.*, 2010). In addition to its ease of learning and high rate of voice reacquisition, shunt vocalization has the advantage of being able to produce a voice that is closer to natural than other methods of vocalization because it uses the same expiratory flow as the larynx. Shunt vocalization has been widely used in Japan in recent years (Sano *et al.*, 2020). However, it does not yield an adequate quality of speech in terms of intelligibility and range of frequency. In a systematic review, van Sluis *et al.* (2018) showed that the fundamental frequency of shunt speech is around 70–150 Hz. The average fundamental frequency of speech in healthy adult males in Japan is around 140 Hz (Kujirai, 1983) and is relatively close to that of shunting speech. However, because the average fundamental frequency of speech in a normal female is about 220 Hz (Kujirai, 1983), the shunting voice has a comparatively low frequency, and this poses a major psychological barrier to female patients. In a follow-up study on the status of patients after the implantation of voice prosthesis, Sano (2021) reported that six of 53 patients had been unable to regain their voice (about 10%). To solve these problems, it is necessary to clarify the physical mechanism of sound generation in shunt speech.

Many studies on shunt vocalization have proposed a relationship between expiratory flow around the voice prosthesis and the sound of the voice. Weinberg and Moon (1984) calculated the airflow-induced resistance based on the ratio of pressure difference between the inlet and outlet of the voice prosthesis to the flow rate and proposed that lowering the resistance to the passing exhalation flow due to airflow

can improve the results of voice rehabilitation. In response to this requirement, Atos Medical Co., Ltd. developed Provox as a voice prosthesis designed to lower airflow resistance. Hilgers *et al.* (1993) experimentally confirmed that it greatly reduced airflow resistance. Provox Vega was introduced in 2010. It was designed to reduce the size of the device while maintaining low resistance due to airflow. Provox Vega is the only voice prosthesis approved for use in Japan at present and has been used in the aftermath of shunt vocalization surgery (Tomifuji, 2020). However, Hilgers *et al.* (2010) reported that high-frequency noise is noticeable in Provox Vega, and its quality of voice deteriorates despite the lowered resistance due to airflow. Therefore, the relationship between flow and voice quality cannot be considered based only on flow resistance. The authors measured the flow field that develops downstream of the voice prosthesis (Provox Vega) by using particle image velocimetry (PIV) (Kosako *et al.*, 2021). The results showed that the velocity of flow near the valve of the voice prosthesis fluctuated at a frequency close to that of the sound generated by the shunt under an expiratory flow rate. This suggests that the fluctuations in velocity of flow generate a sound in the range of frequency of the sound of shunt speech. To better understand the mechanism of sound generation in shunt speech, it is necessary to determine whether the sound is generated by interactions between flow downstream of the voice prosthesis and the esophagus, or by the flow itself. The turbulent structures in the flow field may be related to sound generation, but this has not yet been investigated.

This study seeks to clarify the influence of flow downstream of the voice prosthesis (Provox Vega) on the esophagus to identify the dominant factors influencing the sound generated in the esophagus during shunt vocalization. The expiratory flow in the esophagus during shunting was modeled by a rectangular duct that mimics the esophagus downstream of the voice prosthesis. The velocity field was measured by a PIV system, the sound pressure was calculated by the Ffowcs Williams–Hawkings theory (Williams and Hawkings, 1969), and the dominant factors influencing the sound generated in the esophagus were clarified through a comparison between the simulated and the measured sound pressures. The structure of the flow field and fluctuations in pressure in the esophagus were statistically analyzed together with the characteristic motion of the valve of the voice prosthesis. The relations among the structure of the flow field, vibrations of the valve, and the generation of sound are also discussed.

II. EXPERIMENTAL SETUP AND CONDITIONS

A. Tracheal-esophageal model

A schematic of the test section is shown in Figs. 2(a) and 2(b), and its photograph is provided in Fig. 2(c). It was made of acrylic so that the internal flow could be visualized. The red part imitated the trachea, and the blue part represented the esophagus and hypopharynx. These sections are referred to hereinafter as the tracheal section and the esophageal section, respectively. The blue and red parts in Fig. 2 correspond to the trachea (in red) and esophagus (in blue) in Fig. 1. The tracheal and esophageal sections were connected by the voice prosthesis (Provox Vega).

The esophageal section was designed as a rectangular channel to avoid distortion in the captured image and had dimensions of $W20 \times D20 \times D400 \text{ mm}^3$. The upper part, which is the side of the mouth in humans, was open to the outer environment. The voice prosthesis was placed perpendicular to the wall at a height of 200 mm from the bottom. The diameter of the flow channel was set to 20 mm, which is

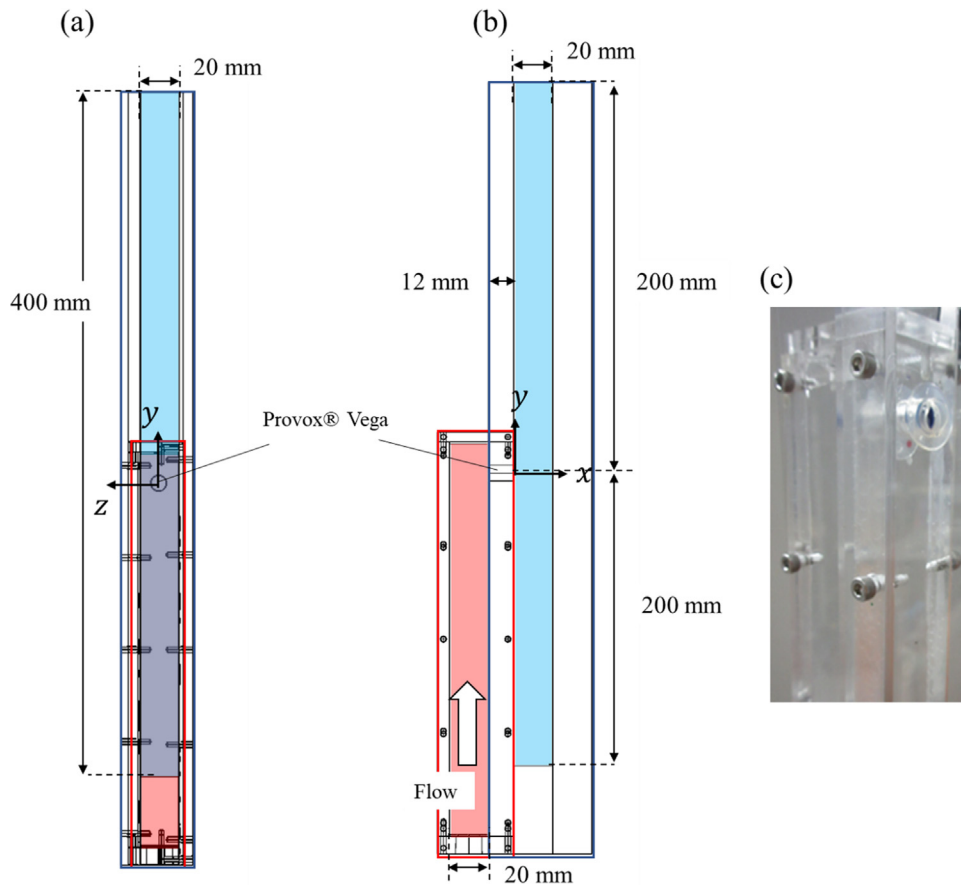


FIG. 2. Schematic view of the test section, with Provox Vega connecting the trachea and the esophageal section. The blue and red parts correspond to those in Fig. 1. (a) Front view. (b) Side view. (c) Photograph of the test section.

the average esophageal diameter in adult males (Kuo and Urma, 2006). The tracheal section was the same as the test section used in a previous report by the authors. The interested reader can refer to the details provided in Kosako *et al.* (2021). The test section simulated expiratory flow from the lungs during shunt speech. The flow applied from the lower part of the tracheal section [indicated by the white arrow in Fig. 2(b)] passed through the Provox Vega into the esophageal section and spurted out from its upper part into free space. Flow in the esophageal section downstream of the voice prosthesis was measured in this study.

B. Experimental apparatus and coordinate axes

A brief description of the entire experimental setup is given below. The flow applied by an air compressor (EARTH MAN, ACP-10A) entered the test section through a calibrated flow meter and a seeding generator (LaVision, 1108926). The flow rate in the channel was controlled by the calibrated flowmeter. The flow rate q_v was set to six values— $q_v = 0, 5, 10, 15, 20,$ and 25 l/min—for each measurement. Preliminary experiments confirmed that the valve of the voice prosthesis and sound generation were not affected by tracer particles (DEHS) applied by the seeding generator (see Appendix A).

The coordinate axes were set such that x represented the direction of flow, y represented the orthogonal direction, and z represented

the span direction. The origin was set at the center of the valve of the voice prosthesis, as shown in Fig. 2.

C. PIV measurements

Flow from the voice prosthesis into the esophageal section was measured by using PIV. The system consisted of a high-speed camera (Photron, FASTCAM-Mini-AX50), a CW-YAG laser (Finess diode pumped solid-state laser, 4 W), and a seeding generator. The high-speed camera was placed perpendicular to the test section to avoid image distortion due to the acrylic. The laser beam was shaped into a 1-mm-thick sheet by an optical system to visualize the tracer particles (Di-Ethyl-Hexyl-Sebacate; DEHS) generated by the seeding generator in the xy section.

The range of measurement was ($0 \text{ mm} < x < 20 \text{ mm}$, $-30 \text{ mm} < y < 60 \text{ mm}$), and the spatial resolution of the high-speed camera was 0.11 mm/pixel. The sampling frequency of the camera was set to 7.2 kHz. A total of 58 231 images were captured by using the Photron FASTCAM Viewer 4. The flow rate was mainly examined at $q_v = 15$ l/min, which is within the range of the exhalation flow rate during speech. Davis 10.2 of LaVision was used for PIV analysis, the multi-pass (decreasing size) method was used for PIV calculations, and the mesh after PIV analysis consisted of 32×129 cells with a spatial resolution of 0.66 mm.

D. Measuring vibrations of the valve of voice prosthesis

Images over the time series were recorded to investigate the relationship between fluctuations in the velocity of flow in the esophageal section and oscillations of the valve inside the voice prosthesis. A high-speed camera (Photron, FASTCAM-Mini-AX50) was placed parallel to the yz plane ($x = 500$ mm, $y = 0$ mm, $z = 0$ mm) to capture the internal voice prosthesis from the side of the esophageal test section. At this measurement position, the size of images captured by the high-speed camera was 256×256 pixels at a spatial resolution of 0.0402 mm/pixel. A total of 32 768 images were taken under each flow rate at a sampling frequency of 5 kHz and an exposure time of 200 μ s. The flow was measured in the range from $q_v = 0$ to 25 l/min.

E. Measuring the sound generated by the test section

The sound pressure generated by the test section was also measured. An omni-directional microphone (Aco, TYPE 7016) was set at $x = 1.5$ m, $y = 0$ m, and $z = 2$ m. The ratio ξ of the sound pressure of the acoustic wave to that of the pseudo-acoustic wave is defined as $\xi = 2\pi fr/c_0$, where f is the frequency of sound, r is the distance between the source and the microphone, and c_0 is the speed of sound. Because the ratio was in the range of $4.7 \leq \xi \leq 5.7$ at the position where measurements of the microphone were recorded, the sound wave was about 4.7–5.7 times larger than the pseudo-sound wave. Converting the present ratio ξ to units in decibels yielded 13 to 15 dB. The effect of the pseudo-sound wave can be considered to be negligibly small if ξ is 10 dB or higher (Iida, 2003). Therefore, the frequency of the sound wave was higher than that of the pseudo-sound wave at this position. Signals of the sound pressure were recorded using LabVIEW 2017 from National Instruments. The sampling frequency of the microphone was 20 kHz, and the measurement time was set to 30 s. The measurements were made under flow conditions ranging from $q_v = 0$ to 25 l/min to investigate the dependence of the flow rate on the generated sound. Signals of the sound pressure were recorded using LabVIEW 2017 from National Instruments. The sampling frequency of the microphone was 20 kHz, and the measurement time was set to 30 s. The measurements were made under flow conditions ranging from $q_v = 0$ to 25 l/min to investigate the dependence of the flow rate on the generated sound.

F. Sound pressure estimation

The Lighthill equation, which is the equation governing aerodynamic sound, can be expressed as follows:

$$\left(\frac{1}{c_0^2} \frac{\partial^2}{\partial t^2} - \nabla^2\right) [c_0^2(\rho - \rho_0)] = \frac{\partial^2}{\partial x_i \partial x_j} T_{ij}, \tag{1}$$

$$T_{ij} = \rho u_i u_j + [(p - p_0) - c_0^2(\rho - \rho_0)] \delta_{ij} - \sigma_{ij}, \tag{2}$$

where \mathbf{u} is the velocity of flow, p is the pressure, ρ is the density, p_0 and ρ_0 are the standard states of each physical quantity, δ_{ij} is Kronecker delta, and σ_{ij} is the viscous stress tensor. T_{ij} is called the Lighthill tensor. The left-hand side of Eq. (1) refers to the propagation of sound waves in a uniform stationary medium, and the right-hand side represents the source propagated by the left-hand side. When the Reynolds number is high, the viscous stress term is small by

comparison and can be neglected. When flow occurs at a low Mach number, the condition $\rho = \rho_0$ is satisfied. This can also be assumed to be isentropic flow. Considering these factors, the Lighthill tensor can be approximated as

$$T_{ij} \approx \rho_0 u_i u_j. \tag{3}$$

The integral solution of the Lighthill equation in the presence of a moving object is given by the following equation, which is called the Ffowcs Williams–Hawkings (FW–H) equation:

$$p - p_0 = \frac{1}{4\pi} \frac{\partial^2}{\partial x_i \partial x_j} \int_V \left[\frac{T_{ij}}{r}\right] dV - \frac{1}{4\pi} \frac{\partial}{\partial x_i} \oint_S \left[\frac{(p - p_0)\delta_{ij} - \sigma_{ij}}{r} n_j\right] dS - \frac{1}{4\pi} \frac{\partial}{\partial x_i} \oint_{V_s} \left[\frac{\rho_0 \dot{V}_i}{r}\right] dV + \frac{1}{4\pi} \frac{\partial^2}{\partial x_i \partial x_j} \oint_{V_s} \left[\frac{\rho_0 V_i V_j}{r}\right] dV, \tag{4}$$

where $r = |\mathbf{x} - \mathbf{x}_{so}|$ is the distance between the observation point and the source, \mathbf{n} is the vertical vector from the boundary of the object to the outside, \mathbf{V} and $\dot{\mathbf{V}}$ are the moving velocity and the acceleration of the object, respectively, and V_s is the internal volume of the object. The brackets $[\]$ mean that the calculation takes into account the delay due to propagation at the speed of sound from each source \mathbf{x}_{so} to the observation point \mathbf{x} (Williams and Hawkings, 1969).

III. RESULTS AND DISCUSSION

A. Sound generated by the test section

Figure 3 shows the frequency spectrum of sound pressure 3 measured by the microphone. The background data were subtracted from the data obtained at each flow rate. For $q_v = 5, 10,$ and 15 l/min, the fundamental peaks of frequency were identified in the range from 103 to 120 Hz (117 Hz at $q_v = 15$ l/min), as shown by the black dashed line in the figure. The peaks near 200 Hz were their harmonics. As the flow rate was increased from $q_v = 5$ to $q_v = 15$ l/min, the sound pressure at the fundamental peak frequency increased and the frequency increased as well, but when the flow rate was increased further to

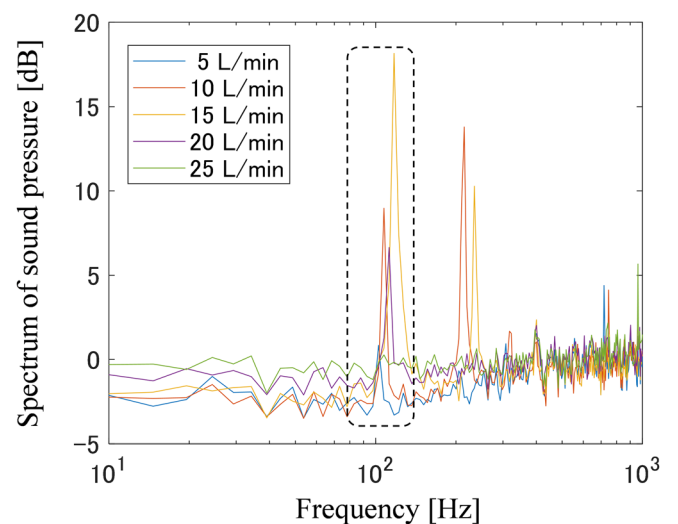


FIG. 3. Frequency spectrum of the sound generated by the test section.

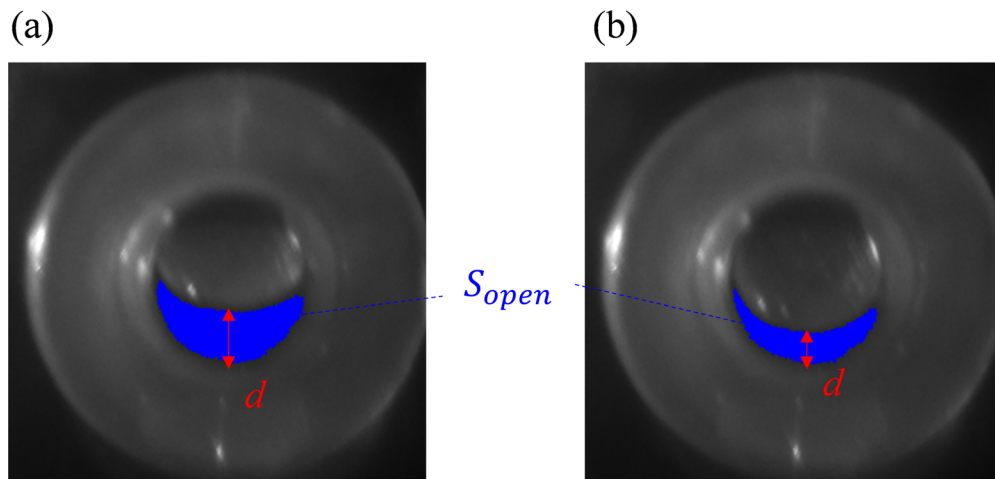


FIG. 4. Instantaneous image of voice prosthesis. d is the size of the aperture, and S_{open} is the area of the opening of the valve, shown in blue. (a) Aperture phase. (b) Closed phase.

$q_v = 20$ l/min, the sound pressure decreased and the peak frequency become smaller than that at $q_v = 15$ l/min. Then, at $q_v = 25$ l/min, we could not observe the sound generation around 100 Hz. In other words, there was a threshold of sound generation around 100 Hz in the range from $q_v = 20$ to $q_v = 25$ l/min. The frequency of shunt vocalizations has been reported to be in the range of 70–150 Hz (van Sluis *et al.*, 2018), indicating that the sounds generated from this test section were close in frequency to those of shunt speech.

B. Vibration of valve of voice prosthesis

The valve inside the voice prosthesis oscillated at a specific flow rate. The instantaneous images of the opening and closing phases of the valve are shown in Fig. 4. The distance indicated by the red line is defined as the opening distance d (mm) of the valve, and the temporal variations in d were studied using image processing. Figure 5 shows the frequency spectrum of variations in d under flow conditions ranging from $q_v = 5$ to 25 l/min. It confirms that at flow rates from $q_v = 5$ to 20 l/min, the valve vibrated and had a peak fundamental frequency between 103 and 120 Hz, as indicated by the black dashed line in the figure. The frequency at each flow rate in Fig. 5 is consistent with that of the generated sound in Fig. 3, indicating that the generated sound was affected by the frequency of vibrations of the valve. In the range from $q_v = 5$ to 15 l/min, the valve repeatedly opened and closed, and its opening distance increased with the flow rate. At $q_v = 20$ l/min, the valve oscillated around a fixed opening distance, but its amplitude of oscillations decreased. At $q_v = 25$ l/min, the valve remained open and did not oscillate.

The blue area indicated in Fig. 4 is defined as the area of opening of the valve S_{open} , and was calculated by image processing according to the durations of different flow rates. The relationship between the flow rate and the time-averaged area of the opening S_{open} is shown in Fig. 6. The red line shows the approximate line obtained by the least squares method, and the value of its quality (R-squared) was 0.98. The flow rate was proportional to S_{open} regardless of vibrations in the valve.

Figure 7 shows the relationship between the flow rate and the frequency of vibrations of the valve. The solid orange line represents the fitting line obtained by the least squares method from $q_v = 5$ to 15 l/min. The value of its quality (R-squared) was 0.96. The flow rate and the frequency of vibrations of the valve had a linear relationship from $q_v = 5$ to 15 l/min, but this relation did not obtain when $q_v = 20$ l/min. The frequency spectrum in Fig. 5 shows that the peaks of frequency were composed of the fundamental frequency and its harmonic for $q_v = 5, 10,$ and 15 l/min, while the only noticeable peak of frequency at $q_v = 20$ l/min occurred at 109 Hz. In other words, there might have been a threshold where vibrations occurred between $q_v = 20$ l/min and $q_v = 25$ l/min. However, the flow-dependent frequency changed in the range of 5 l/min $\leq q_v \leq 15$ l/min. Therefore,

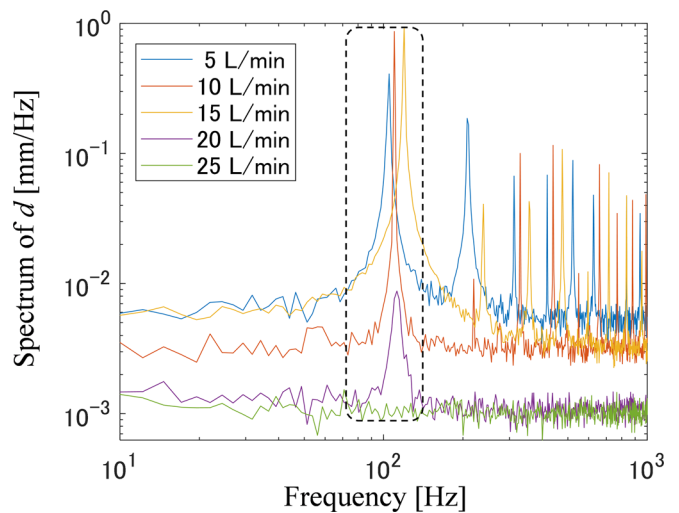


FIG. 5. Spectrum of variations in the aperture d of the valve inside the voice prosthesis.

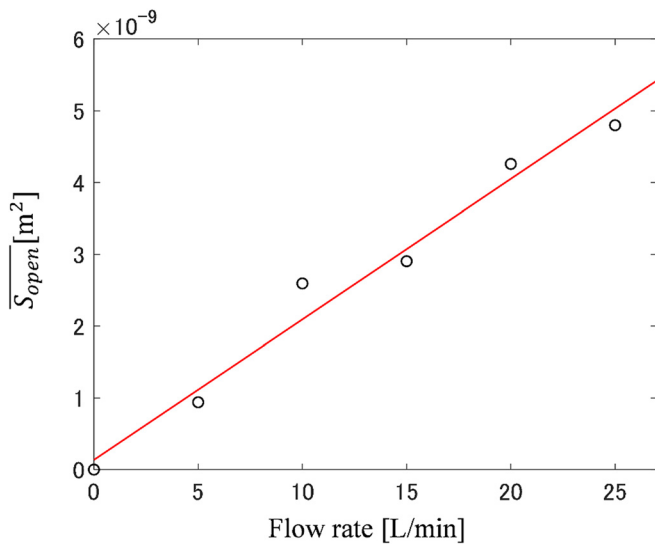


FIG. 6. Relationship between flow rate and area of the aperture. \circ : Area of valve opening; red solid line: linearly fitted line.

the mode of vibration changed when the flow rate was varied from 151/min to $q_v = 201$ /min.

C. Flow field in the esophagus

The instantaneous velocity vector and contours of the magnitude of flow in the esophageal section at $q_v = 15$ l/min are shown in Fig. 8. The voice prosthesis is indicated by the white area. The time $t = 0$ was defined as the time when the velocity of flow was the highest in a cycle of fluctuations in velocity downstream of the voice prosthesis. It was determined using the visualized images. T represents the period of fluctuation of the oscillation in flow, and was $T = 0.0093$ s for $q_v = 15$ l/min based on Fig. 10. Figure 8(a) shows that a fast flow at

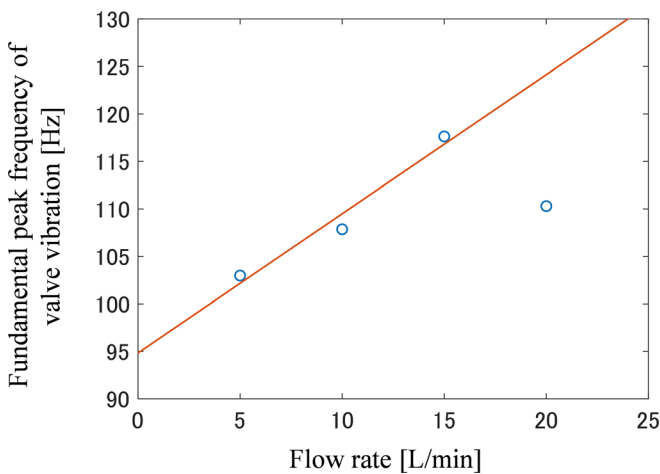


FIG. 7. Relationship between the flow rate and the frequency of vibrations of the valve. \circ : Fundamental frequency of vibrations of the valve; red solid line: linearly fitted line.

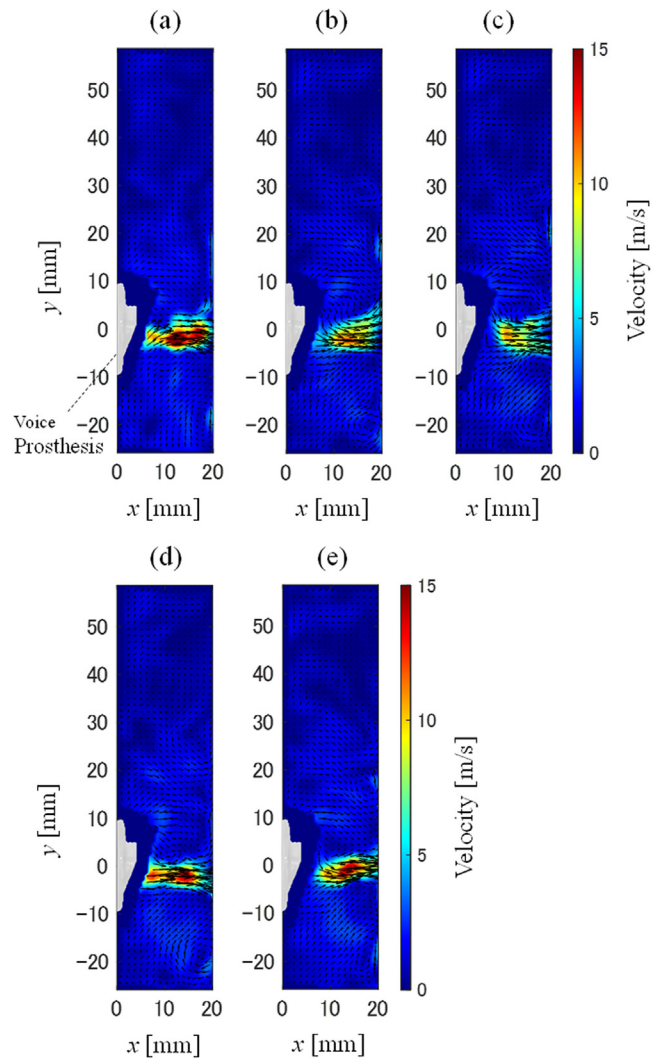


FIG. 8. Instantaneous velocity vectors and contours of the magnitude of flow in the esophageal section: (a) $t/T = 0$, (b) $t/T = 1/4$, (c) $t/T = 1/2$, (d) $t/T = 3/4$, and (e) $t/T = 1$.

nearly 15 m/s was ejected from the voice prosthesis in the upper-right direction $(+x, +y)$. Figures 8(b) and 8(c) show that the flow from the voice prosthesis slowed and changed to the lower-right direction. Then, Figs. 8(d) and 8(e) show that the velocity increased, the direction of flow changed to the upper-right, and the stage from (a) to (e) was repeated. When the valve opened from the bottom to the top, jet flow was concentrated in a localized area at the bottom of the valve such that the flow spurted toward the $(+x, +y)$ direction from the bottom, as shown in Figs. 8(a) and 8(e). When the valve spacing d increased, the area of the jet increased and the velocity of flow near the valve decreased [Fig. 8(c)]. Flow was generated along the valve to the lower-right direction $(+x, -y)$ (the top of the valve was the fixed end, because of which its position changed diagonally to the lower right). The direction of flow close to downstream of the voice prosthesis ($0 \text{ mm} < x < 10 \text{ mm}$) fluctuated over time, but the position along the

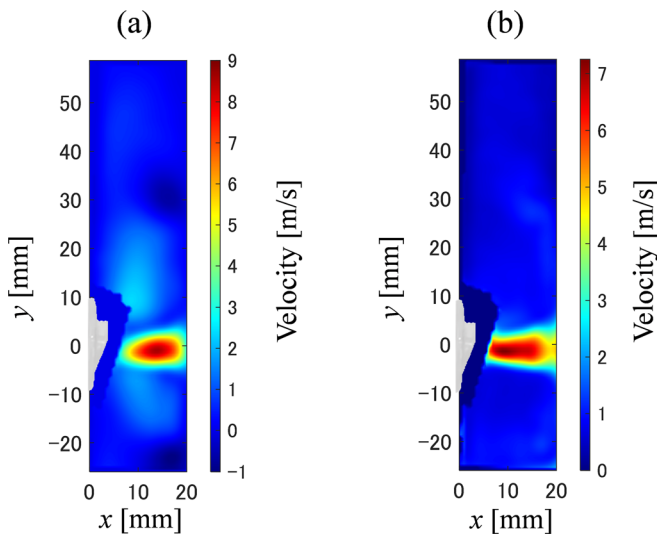


FIG. 9. Contours of the statistical value of the streamwise velocity: (a) mean velocity distribution and (b) root mean-squared (RMS) distribution.

y direction where the flow came into contact with the wall at $x = 20$ mm (-5 mm $< y < 5$ mm) indicated little change over time. The very fast flow from the valve dispersed upward and downward after impinging on the wall, but lost momentum and almost decayed in the range of -5 mm $< y < 5$ mm.

Figure 9(a) shows the time-averaged velocity in the direction of flow and (b) shows the distribution of its RMS in the esophageal section. It is clear from Fig. 9(a) that the mean flow emanating from the voice prosthesis was distributed in a direction almost parallel to the x -axis. Figure 9(b) shows that the distribution of the RMS was slightly tilted to the lower right, and the intensity of the RMS is the highest near the voice prosthesis.

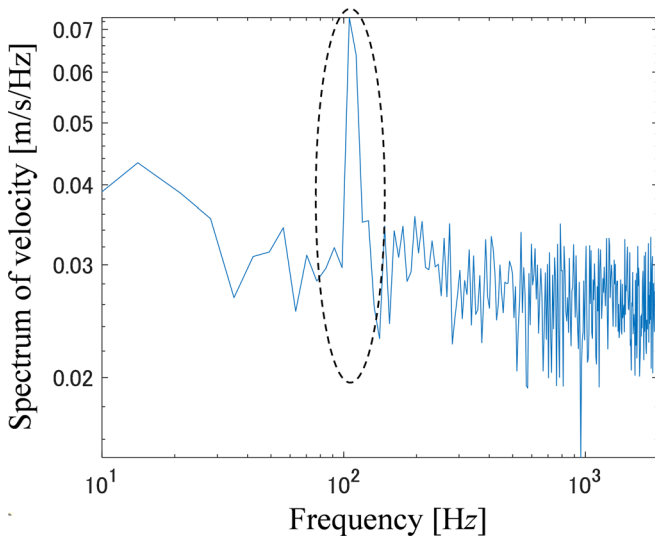


FIG. 10. Spectra of streamwise fluctuations in velocity at the position where its RMS was the maximum ($x = 8$ mm, $y = -1.4$ mm).

The spectrum of the fluctuations in velocity along the direction of flow at the point of maximum fluctuation ($x = 8$ mm, $y = -1.4$ mm) is shown in Fig. 10. A fundamental peak in frequency of 117 Hz was observed, as shown by the black dashed line, and was the same as the frequency of the generated sound and the valve inside the voice prosthesis. It can be concluded that the fluctuation in flow downstream of the voice prosthesis was determined by vibrations of the valve.

Assuming that the flow was two-dimensional, the second invariant Q ($Q = -1/2(\partial u_i \partial u_j / \partial x_i \partial y_j)$) in the xy section was obtained from the PIV-measured velocity field by using the spatial derivative of velocity through the central difference. Regions with positive values of Q indicated the presence of a vortex center, while regions with negative Q indicated a strong velocity shear. Figure 11 shows the instantaneous distribution of Q in the esophageal section, where the time in

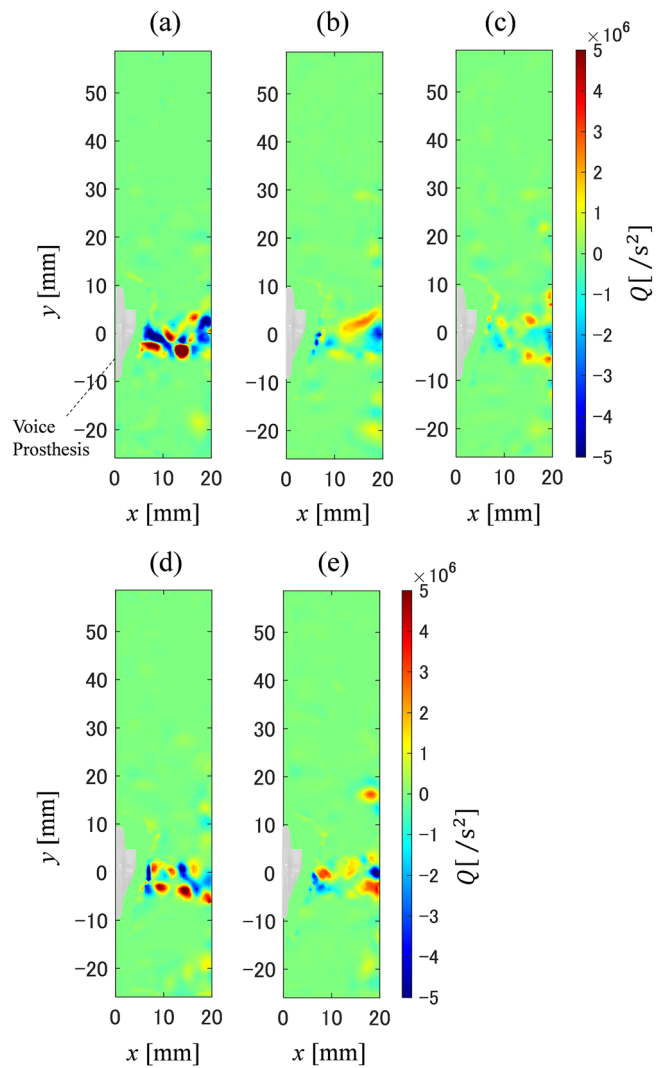


FIG. 11. Instantaneous distribution of Q in the esophageal section: (a) $t/T = 0$, (b) $t/T = 1/4$, (c) $t/T = 1/2$, (d) $t/T = 3/4$, and (e) $t/T = 1$.

Fig. 11 is the same as in Fig. 8. Figures 11(a)–11(e) show that discrete vortex structures were generated from the voice prosthesis and regions of strong shear were present between them.

In comparison with the distribution of the instantaneous velocity (Fig. 8), the vortex structures were alternately generated up and down along the region with a high velocity and spread out in the $\pm y$ direction when approaching the wall. As the distribution of velocity at the outlet of the voice prosthesis fluctuated with time, the direction of release of the vortex structure changed. The magnitude of Q decreased from (a) $t/T = 0$ to (c) $t/T = 1/2$ and then increased again from (c) $t/T = 1/2$ to (e) $t/T = 1$ as the velocity of flow fluctuated, thus repeating the cycle.

Figure 12(a) shows the time-averaged contour of Q in the esophageal section, and Fig. 12(b) shows the contour of its RMS. The frequency spectrum of temporal variations in Q at each position was calculated. The distribution of the frequency spectrum at 117 Hz, which was equal to the frequencies of the generated sound and the vibrations of the valve, is shown in Fig. 12(c). Figure 12(a) shows that the vortex structures were symmetrically distributed on average, with the center at $y = -1$ mm (indicated by the black line in the figure), and a very strong shear was observed due to the flow impacting the wall in the region between the vortex structures and the wall ($16 \text{ mm} < x < 20 \text{ mm}$, $-5 \text{ mm} < y < 3 \text{ mm}$). A strong shear was also observed in the vicinity of the valve inside the voice prosthesis. The velocity vectors in Figs. 8(a)–8(e) suggest that this shear was caused by the motion of valve and changed the direction of the main flow. Figure 12(b) shows that the fluctuation in Q was the largest near the valve, decreased with the distance from it along the x direction, and increased again near the wall. However, the results at 117 Hz,

shown in Fig. 12(c), show that the fluctuations in Q did not peak near the voice prosthesis but in the region where the flow impinged on the wall ($16 \text{ mm} < x < 20 \text{ mm}$, $-5 \text{ mm} < y < 3 \text{ mm}$). This region corresponded to the area where the shear increased.

D. Pressure distribution in the esophageal section

Poisson’s equation [Eq. (5)] was solved numerically to estimate the pressure distribution in the measured cross section

$$\nabla^2 p = -\rho_0 \nabla \cdot (u \cdot \nabla) u. \tag{5}$$

Poisson’s equation was discretized by using central differencing and iteratively calculated using the Gauss–Seidel method. The right-hand side of Poisson’s equation was calculated from the velocity field measured by PIV. In the measured region, $x = 0 \text{ mm}$, $-30 \text{ mm} < y < 60 \text{ mm}$ and $x = 20 \text{ mm}$, $-30 \text{ mm} < y < 60 \text{ mm}$ were the boundaries of the wall, and the pressure gradient was set to zero at the wall. Figure 13 shows the pressure distribution in the esophageal section. The interval in Fig. 13 is the same as those of the distribution of instantaneous velocity in Fig. 8. Figure 14 shows (a) the time-averaged distribution, (b) the RMS distribution, and (c) the distribution of variations in the frequency spectrum of pressure at 117 Hz. Figures 13(a)–13(e) show that the distribution of instantaneous pressure exhibited a trend of high pressure, low pressure, and high pressure in the direction of flow (x direction). Furthermore, the magnitude of the overall pressure fluctuated with time as the velocity of flow fluctuated. Figure 14(a) shows that pressure in the vicinity of the valve was always highly independent of the position of the valve. This is because strong shear ($Q < 0$) occurred near the valve due to the deformation

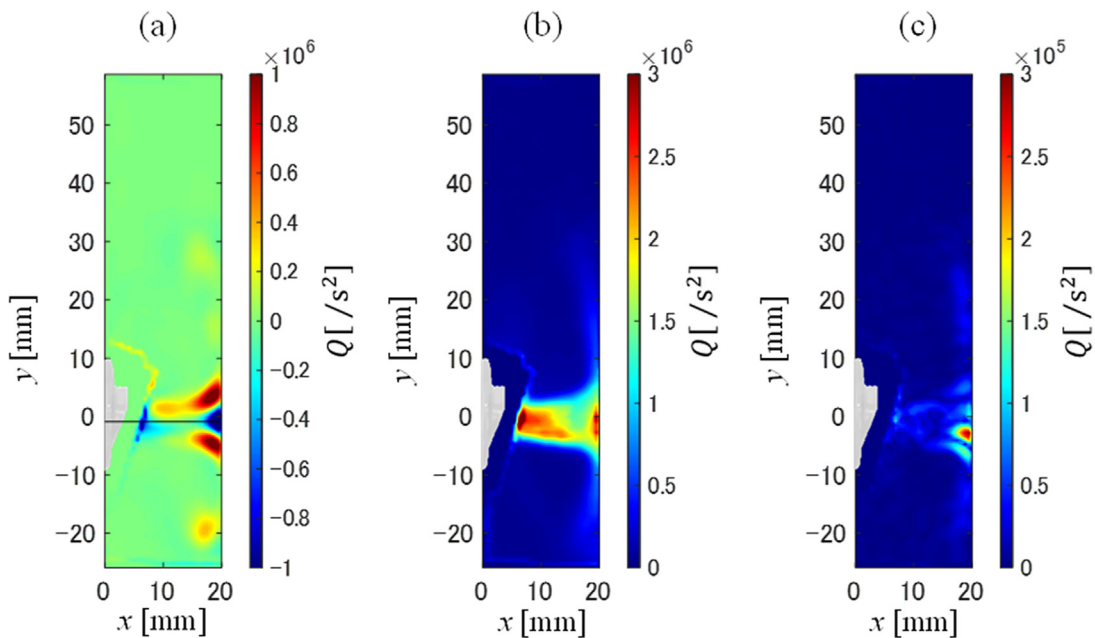


FIG. 12. Statistical distribution of Q in the esophageal section: (a) time-averaged distribution, (b) RMS distribution, and (c) distribution of variations in the frequency spectrum of Q at 117 Hz.

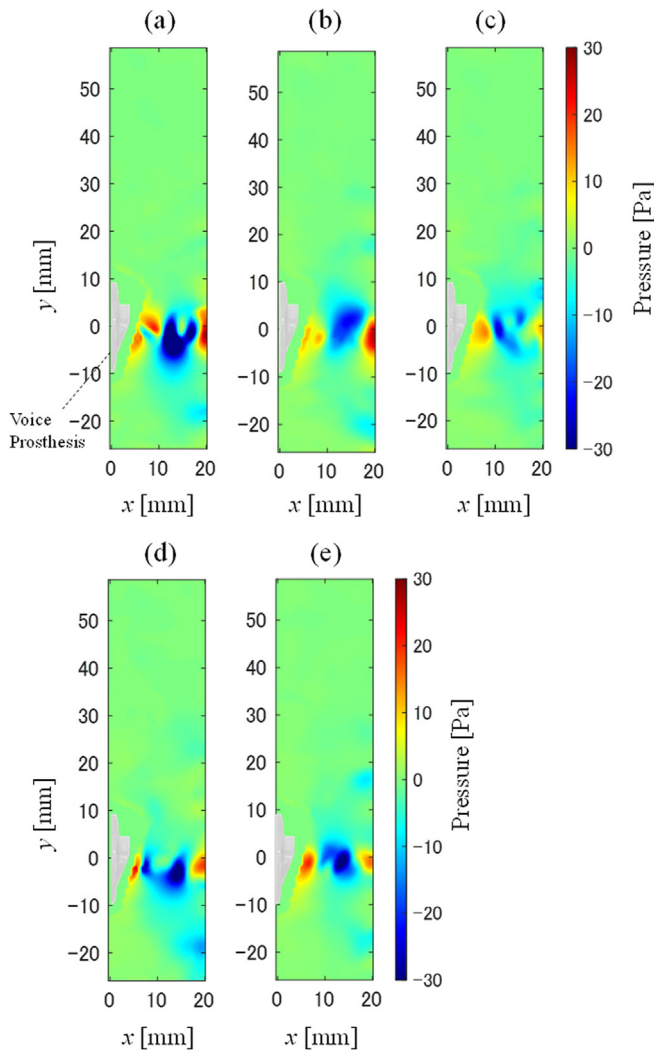


FIG. 13. Instantaneous pressure distribution in the esophageal section: (a) $t/T = 0$, (b) $t/T = 1/4$, (c) $t/T = 1/2$, (d) $t/T = 3/4$, and (e) $t/T = 1$.

of the jetting flow structure caused by the vibration of the valve. The flow pushed the valve away and caused it to open, while the high-pressure region near the valve acted as a force to close it.

In Fig. 14(a), the low-pressure region is widely distributed in the range of $10 \text{ mm} < x < 18 \text{ mm}$ and a strong high-pressure region occurs in the range of $18 \text{ mm} < x < 20 \text{ mm}$. A comparison between Figs. 12(a) and 14(a) shows that the low-pressure region extended from area with positive values of Q , and the vortex structure released along the flow from the voice prosthesis generated a large low-pressure region. In addition, a high-pressure region was formed near the wall owing to the strong shear generated by the collision of flow with the wall. Figure 14(b) shows that the region of large fluctuations in pressure spread in the $\pm y$ direction from the exit of the voice prosthesis to the wall at $x = 20 \text{ mm}$, and the largest fluctuations in pressure were observed at the position where the flow impinged on the wall. Figure 14(c) shows that the highest pressure at 117 Hz

corresponded to the time when the vortical structures impinged on the wall.

We now summarize the above discussion. The flow through the voice prosthesis generated fluctuations in velocity in the range of frequency of speech (117 Hz). This flow impinged on the wall and created a strong, time-varying shear region that caused fluctuations in pressure in the range of frequency of speech on the esophageal wall near the impinging point.

The test section was made of acrylic resin to visualize the flow field inside it. However, this did not simulate the vibrations of the esophageal wall when the flow impinged on it. We verify the effects of vibrations in the esophageal wall on the results in Appendix B.

E. Estimated sound pressure and the dominant factors influencing it

The sound generated by flow in the esophageal section was estimated based on the FW–H equation [Eq. (4)], which is the analytical solution of the Lighthill equation. The third and fourth terms in Eq. (4) refer to the effects of volume exclusion due to vibrations of the valve inside the voice prosthesis, but were ignored because they were negligibly small. The first term, which represents direct sound due to the turbulence in flow (vortex structure) in the esophageal section, and the second term, which represents the sound pressure estimated from fluctuations in pressure on the esophageal wall, were calculated.

Flow was assumed to be two-dimensional in the first term, and the Lighthill source term ($\partial^2 \rho_0 u_i u_j / \partial x_i \partial x_j$) was then calculated. The sound pressure was estimated by adding the contributions from all sources in the space measured by PIV at each frequency.

For the second term, we defined $x = 0 \text{ mm}$, $-30 \text{ mm} < y < 60 \text{ mm}$ and $x = 20 \text{ mm}$, $-30 \text{ mm} < y < 60 \text{ mm}$ as the esophageal wall and used the distribution of instantaneous pressure calculated in Sec. III D to obtain temporal variations in the pressure gradient in the x direction on the wall.

The frequency spectrum of sound pressure calculated from the first and second terms of the FW–H equation at ($x = 1.5 \text{ mm}$, $y = 0 \text{ mm}$, $z = 2 \text{ mm}$) is shown in Fig. 15. It was calculated on the assumption that the flow was 2D. The level of sound pressure was obtained by multiplying the estimated sound pressure by 4.5 mm, which is the angle of aperture of the voice prosthesis along the z direction. This procedure was carried out for both the first and the second terms. As shown by the black dashed line in Fig. 15, the sound pressure estimated from the second term had a fundamental frequency peak at 117 Hz, which was identical to the frequency of the sound generated by the esophageal section and the vibrational frequency of the valve of the voice prosthesis. The sound pressure estimated from the first term had a small fundamental frequency peak at 117 Hz, but was not as clear as that from the second term, and had a peak value that was about 15 dB smaller than that of the second term. Therefore, the sound source owing to temporal variations in pressure in the esophageal wall was the dominant factor in the sound generated by flow in the esophageal section at $q_v = 15 \text{ l/min}$. Figure 14(c) shows that the fluctuations in pressure in the esophageal wall at the point of collision between the flow and the wall ($x = 20 \text{ mm}$, $-5 \text{ mm} < y < 3 \text{ mm}$) accounted for most of the sound sources.

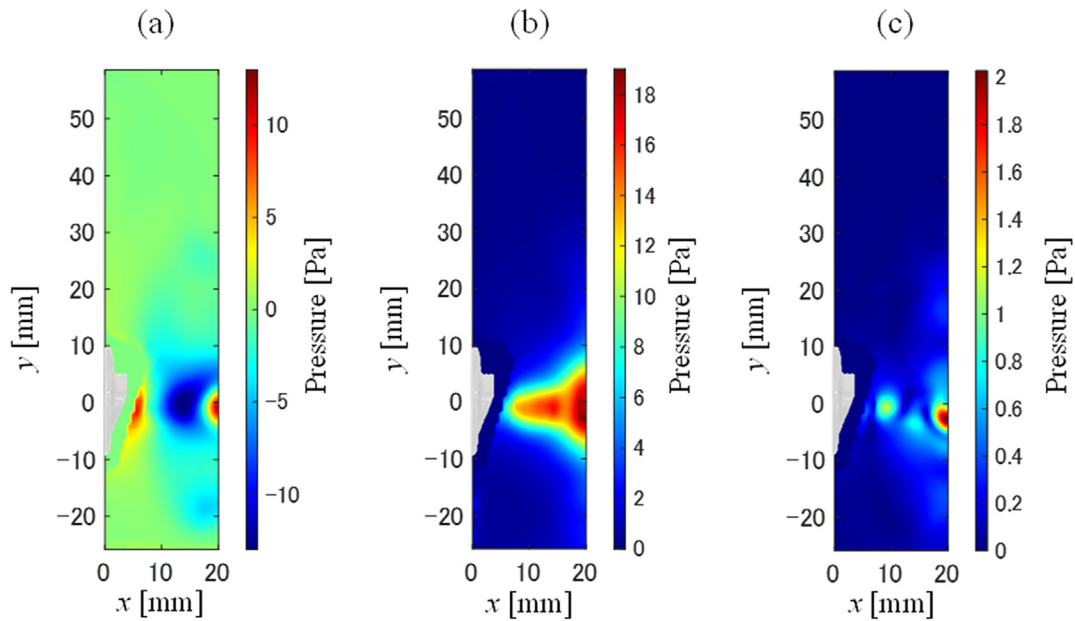


FIG. 14. Statistical distribution of pressure in the esophageal section: (a) time-averaged distribution, (b) RMS distribution, and (c) distribution of the frequency spectrum of pressure at 117 Hz.

IV. CONCLUSIONS

The study sought to clarify the effects of downstream flow of the voice prosthesis (Provox Vega) on the esophageal wall and the mechanism of sound generation in the esophagus during shunt vocalization. Expiratory flow in the esophagus during shunting was modeled by a rectangular duct that mimicked the esophagus downstream of the voice prosthesis. Flow in the esophageal section where the voice prosthesis was installed was measured using PIV, and the spatial and temporal

characteristics of the velocity and pressure distributions were analyzed, along with the characteristics of the valve of the voice prosthesis. The generated sound was estimated from the PIV-measured velocity field based on the FW–H equation and was compared with the generated sound measured by a microphone to examine the dominant factors in sound generation. The results can be summarized as follows:

1. In case of velocities of flow ranging from $q_v = 5$ l/min to $q_v = 20$ l/min, the valve of the voice prosthesis oscillated in the range of frequency of the shunt sound from 100 to 120 Hz, and the velocity of flow downstream of the voice prosthesis fluctuated at the same frequency as the oscillations of the valve inside it. The fundamental frequency of vibrations of the valve had a linear relationship with the applied flow rate in the range from $q_v = 5$ l/min to $q_v = 15$ l/min.
2. The velocity of flow passing through the voice prosthesis fluctuated in the range of frequency of speech. This creates a strong time-varying shear region and caused fluctuations in pressure in the range of frequency of speech on the esophageal wall near the point of impingement. Because of the high shear region near the valve inside the voice prosthesis, the pressure was always high and acted as a kind of force to close the valve.
3. Under expiratory flow ($q_v = 15$ l/min), the main source of aerodynamically generated sound in the esophageal section was the temporal fluctuation in pressure caused by the impingement of flow downstream of the voice prosthesis on the esophageal wall. Its frequency ranged from 100 to 120 Hz.

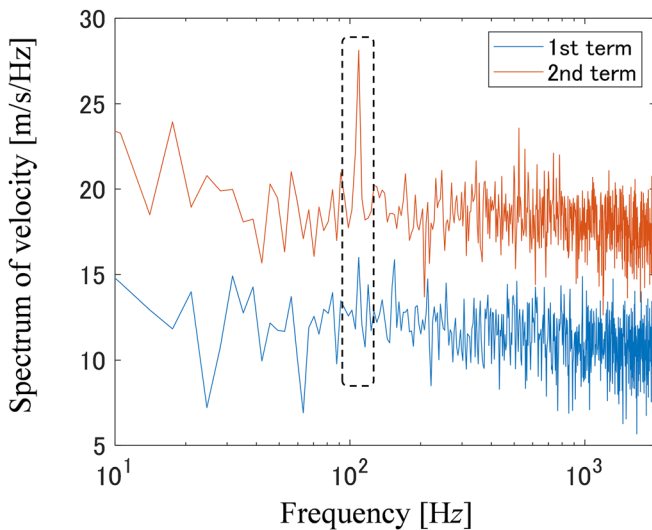


FIG. 15. Frequency spectrum of sound pressure estimated from the FW–H equation. Blue solid line: aerodynamic source (first term), red solid line: fluctuations in pressure in the esophageal wall (second term).

ACKNOWLEDGMENTS

This research was partly funded by the Nitto Science Foundation (2020 Research Grant, No. 3). We thank Mr. Kazuhiro

Murata, a graduate student, for his indispensable assistance in conducting the experiments for this study.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Seiya Kosako: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Resources (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Mariko Hiramatsu:** Formal analysis (supporting); Methodology (supporting); Resources (supporting). **Yasushi Fujimoto:** Formal analysis (supporting); Methodology (supporting); Resources (supporting). **Yoshiyuki Tsuji:** Formal analysis (supporting); Funding acquisition (lead); Investigation (supporting); Project administration (lead); Supervision (lead).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: INFLUENCE OF TRACER PARTICLES ON THE SOUND GENERATED

The spectra of frequencies of the generated sounds with and without tracer particles (DEHS) at a rate of flow of $q_v = 15\text{ l/min}$ in the preliminary experiment are shown in Fig. 16. No significant differences were noted in the frequency spectra with and without the tracer particles.

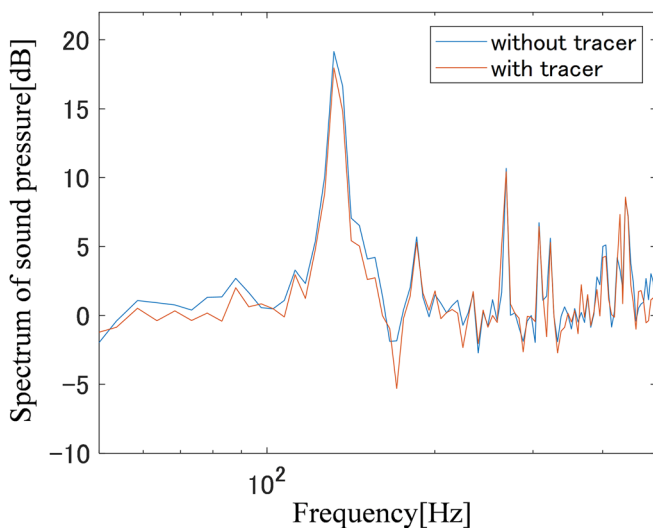


FIG. 16. Influence of tracer particles on the sound generated.

APPENDIX B: INFLUENCE OF THE PRESENCE OR ABSENCE OF ESOPHAGEAL WALL VIBRATION ON THE GENERATED SOUND

The esophageal wall vibrates in shunt vocalization owing to the downstream flow of the voice prosthesis, but the test section is made of acrylic and does not itself vibrate. The effect of the presence (or absence) of vibrations in the esophageal wall on the generated sound was verified. A schematic view of the test section is provided in Fig. 17. The test was conducted by using different materials attached to the forward wall of the channel in the esophageal section, indicated by the materials of the wall in Fig. 17. The materials used were acrylic, silicone rubber (SW950D, gomusuke), and agar (9002-18-0, Hayashi Pure Chemical Ind., Ltd.). The thickness of material was defined as x_w and was set to 5 mm as shown in Fig. 17. The rate of flow was set to $q_v = 15\text{ l/min}$, and the following two measurements were performed:

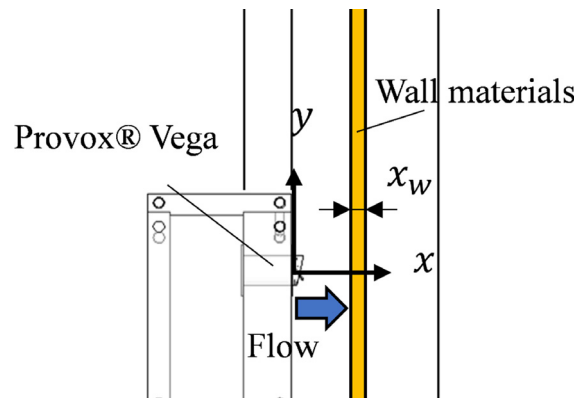


FIG. 17. Overview of the test section wall vibration testing.

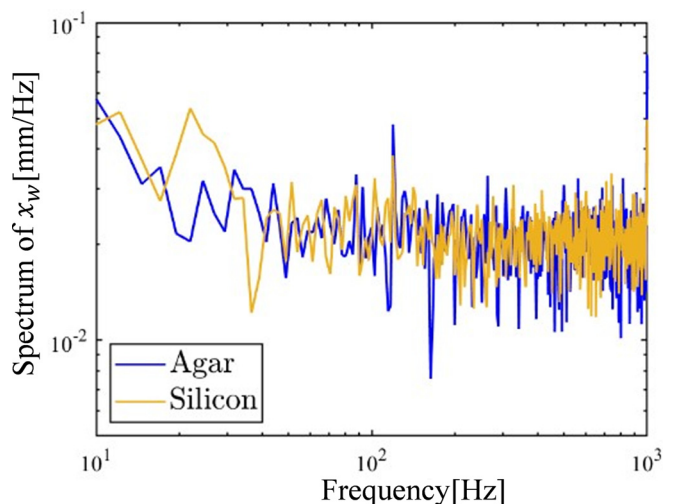


FIG. 18. The frequency spectrum of the variation of x_w for each material.

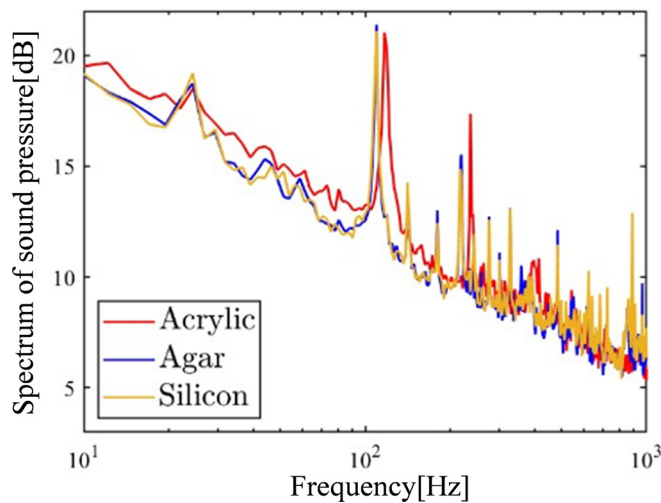


FIG. 19. The frequency spectrum of the generated sound for each material.

- (1) The vibrations of the material were filmed by using a high-speed camera (Photron, FASTCAM-Mini AX50). The high-speed camera was placed parallel to the yz plane ($x = 500$ mm, $y = 0$ mm, $z = 0$ mm) to capture the motion of voice prosthesis from the esophageal side of the test section. The frequency of imaging was set to 5 kHz, and 16 384 images were taken. Figure 18 shows the frequency spectrum of the variations in x_w for the three materials considered. Although there was no variation in x_w when acrylic was used, both agar and silicon led to oscillations of around 110 Hz. This peak frequency was close to the frequencies of the generated sound, vibrations of the valve, and fluctuations in the velocity of flow measured in the test section without materials for the wall. We can conclude that the vibrations in agar and silicone were caused by the downstream flow of the voice prosthesis.
- (2) The generated sound was measured by a microphone. An omni-directional microphone (ACO, Type 7016) was placed at $x = 1.3$ m, $y = 0$ m, and $z = 2$ m. The measurement frequency was set to 20 kHz, and the sound was recorded for 30 s. The frequency spectrum of the generated sound for each material is shown in Fig. 19. The fundamental frequency and sound pressure of the generated sound were similar for all three materials.

The presence (or absence) of vibrations in the wall had no significant effect on the generated sound. We can conclude that the material of the wall did not influence the results obtained when using acrylic in the test section.

REFERENCES

- Hilgers, F. J., Ackerstaff, A. H., van Rossum, M., Jacobi, I., Balm, A. J., Tan, I. B., and van den Brekel, M. W., "Clinical phase I/feasibility study of the next generation indwelling Provox voice prosthesis (Provox Vega)," *Acta Otolaryngol.* **130**(4), 511–519 (2010).
- Hilgers, F. J., Cornelissen, M. W., and Balm, A. J., "Aerodynamic characteristics of the Provox low-resistance indwelling voice prosthesis," *Eur. Arch. Otorhinolaryngol.* **250**(7), 375–378 (1993).
- Iida, A., "Aeroacoustic measurement," *Acoust. Soc. Jpn.* **59**(5), 282–287 (2003).
- Kaburagi, T., Masaki, S., Motoki, K., Matsuzaki, H., and Kitamura, T., *Computational Models and Visualization Techniques of Speech Production Processes* (Corona Publishing Co, Bunkyo-ku, Japan, 2010), pp. 105–109.
- Kosako, S., Hiramatsu, M., Fujimoto, Y., and Tsuji, Y., "Flow field structure formed around voice prosthesis during shunt vocalization," *Trans. Jpn. Soc. Mech. Eng.* **87**(901), 00157 (2021) (in Japanese).
- Kosako, S., Tsuneyoshi, T., and Tsuji, Y., "Influence of the flow field structure formed downstream of the human vocal cord model on the sound generated," *Trans. Jpn. Soc. Mech. Eng.* **87**(893), 00255 (2021) (in Japanese).
- Kujirai, K., "Relationship among intensity and fundamental frequency of voice and air flow rate during phonation," *Otol. Fukuoka* **29**(5), 756–769 (1983) (in Japanese).
- Kuo, B. and Urma, D., "Esophagus-anatomy and development," *GI Motil. Online* (published online, 2006).
- Sano, K., Fukushima, H., Kamiyama, R., and Mitani, H., "Pre/post-operative instruction for patients using Provox® voice prosthesis," *Head Neck Cancer* **46**(4), 322–327 (2020) (in Japanese).
- Sano, Y., "The outcome of voice restoration with Provox® after total laryngectomy," *J. Otolaryngol. Jpn.* **124**(2), 128–134 (2021).
- Tateda, M., Kuwashima, S., Yokoyama, T., and Sato, H., "Management of voice rehabilitation using voice prosthesis," *Jpn. J. Logopedics Phoniatrics* **51**, 266–268 (2010) (in Japanese).
- Tomifuji, M., "Voice prostheses," *J. Jpn. Broncho-Esophagol. Soc.* **71**(3), 283–284 (2020) (in Japanese).
- van Sluis, K. E., van der Molen, L., van Son, R., Hilgers, F., Bhairosing, P. A., and van den Brekel, M., "Objective and subjective voice outcomes after total laryngectomy: A systematic review," *Eur. Arch. Otorhinolaryngol.* **275**(1), 11–26 (2018).
- Weinberg, B. and Moon, J., "Aerodynamic properties of four tracheoesophageal puncture prostheses," *Arch. Otolaryngol.* **110**(10), 673–675 (1984).
- Williams, J. E. F. and Hawkings, D. L., "Sound generation by turbulence and surfaces in arbitrary motion," *Philos. Trans. R. Soc. London, Ser. A* **264**(1151), 321–342 (1969).