PAPER

Numerical Simulation of Air Flow through Glottis during Very Weak Whisper Sound Production

Makoto OTANI[†a)] *and* Tatsuya HIRAHARA[††b)], *Members*

SUMMARY A non-audible murmur (NAM), a very weak whisper sound produced without vocal fold vibration, has been researched in the development of a silent-speech communication tool for functional speech disorders as well as human-to-human/machine interfaces with inaudible voice input. The NAM can be detected using a specially designed microphone, called a NAM microphone, attached to the neck. However, the detected NAM signal has a low signal-to-noise ratio and severely suppressed high-frequency component. To improve NAM clarity, the mechanism of a NAM production must be clarified. In this work, an air flow through a glottis in the vocal tract was numerically simulated using computational fluid dynamics and vocal tract shape models that are obtained by a magnetic resonance imaging (MRI) scan for whispered voice production with various strengths, i.e. strong, weak, and very weak. For a very weak whispering during the MRI scan, subjects were trained, just before the scanning, to produce the very weak whispered voice, or the NAM. The numerical results show that a weak vorticity flow occurs in the supraglottal region even during a very weak whisper production; such vorticity flow provide aeroacoustic sources for a very weak whispering, i.e. NAM, as in an ordinary whispering.

key words: non-audible murmur, whispered voice, whispered voice production mechanism, computational fluid dynamics, air flow through glottis

1. Introduction

A non-audible murmur (NAM) is a very weak vocal-tract resonance sound produced without a vocal fold vibration. The weak sound propagates through the body and it can be detected by using a sensor attached to the neck behind the ear, resulting in a NAM signal typically represented by voltage value. The NAM is inaudible, but, after applying appropriate signal amplification and processing to the detected NAM signal, it can be presented to a listener as an audible speech through an electroacoustic transducer, or to a speech recognition system, thus enabling silent-speech interfaces [1]–[3].

The speech sound we typically use is an air-borne sound, which is a small and fast vibration of the air. The vibration of the air column in the vocal-tract vibrates the wall of the vocal tract, and some sound energies pass through the tissues of the neck or chest. This body-conducted sound that travels through the neck tissue can be sensed with a custom sensor modified from a microphone. Actually, speech sounds propagate not only through the air or the bone but also via the body tissue including muscles. Nakajima et al. found that a weak murmured speech, which is usually unheard by nearby listeners, can be detected using a specially designed microphone, called a NAM microphone [2]. Shimizu et al. clarified frequency characteristics of several types of the NAM microphones [4]. The NAM microphone is a small condenser microphone covered with a soft impression material such as a soft silicon or a urethane elastomer, which provides a better impedance matching between biological soft tissues and the condenser microphone diaphragm. With the NAM microphone attached to the neck surface close behind an ear, a NAM can be detected as a body-conducted voice. The NAM, which is inaudible even for people nearby, can be audible when detected by the NAM microphone. This enables the development of human-to-human and human-to-machine interfaces whose inputs are inaudible voices, thereby providing a "silent" communication tool. The NAM microphone will also be capable of reviving the speech communication of those with vocal fold problems caused by a laryngeal cancer, nerve disorders and muscle diseases.

The detected NAM, however, has a low signal-to-noise ratio (SNR) and severely suppressed higher-frequency components, resulting in a poor clarity compared to an air-conducted ordinary voice [5]. To improve such degraded speech quality, the NAM propagation mechanism was numerically investigated. The results showed that, compared to an air-conducted sound detected in front of a mouth, a body-conducted, i.e., a soft-tissue conducted, sound attenuates 50 dB at 1 kHz. This consists of a 30-dB full-range attenuation due to an air-to-body transmission loss and a -10-dB/octave spectral decay due to a propagation loss through the body. This decay agrees well with the spectral characteristics of the measured NAM [6].

Following the definition of NAM [1], all speech sounds that are inaudible to nearby people can be regarded as a NAM, but all the NAM sounds that have been used in the past were very weak whispered sounds. However, the "weakness" of the whispered sound is determined only by an auditory sensation, namely a sensation level below 0 dB. No physical criterion exists for the "weakness" of the NAM. In that sense, definition of NAM is ambiguous. In spite of such ambiguous definition, a NAM is, in fact, detectable by the NAM microphone, and is usable as a speech signal, indicating that some physical phenomena occur within a speech organ. It is an intriguing question: What is the physical phe-

nomenon, namely, a production mechanism of the peculiar speech, NAM?

As mentioned above, all the NAM sounds used in the past were very weak whispered sounds. This fact indicates that the NAM is a kind of whispered voice whose intensity is very small so as to be inaudible even to nearby people. Therefore, it is expected that the production mechanism of NAM is basically similar to one of whisper: An air flow from a glottis to a vocal tract produces turbulent noise sources in the vocal tract near the glottis; the noise source resonates in the vocal tract; and the resonated sound radiates from the mouth. Furthermore, it is also expected that, for the NAM production, an air flow rate from the glottis to the vocal tract is smaller than that for an ordinary whispering. This hypothesis raises two questions: Does the weak air flow yield a similar turbulent flow as an ordinary whispering? Is the vocal tract shape with the very weak air flow similar to that for an ordinary whispering?

Many researchers have investigated a mechanism of voice production, and recently, numerical simulations of fluid dynamics and aeroacoustics become a useful tool for clarifying the mechanism. For examples, as for an ordinary voice production, numerical simulations were performed to model the vocal fold vibration due to an air flow [7]–[11]. Some researchers have also performed numerical simulations of a turbulence production at the glottis to investigate the turbulent noise source during the whispered voice production [12]–[15]. These previous works, however, were for the production of audible voices either with or without a vocal fold vibration, not much weaker whispered voice.

In this work, to clarify aeroacoustic phenomena arising in the vocal tract for the NAM production, an air flow through a glottis during a very weak whisper production is numerically investigated. A computational fluid dynamics (CFD) simulation is applied to a very weak air flow in a vocal tract shape for the very weak whispering that is three-dimensionally captured by a magnetic resonance imaging (MRI) scan. For comparisons, the CFD simulation is also applied to air flows through a glottis and vocal-tract shapes for the ordinary strong and weak whisperings.

2. Modeling of Vocal Tract Shape

Vocal tract shapes were obtained for whispered voice productions with three strengths, i.e. strong, weak, and very weak, by using MRI scans to construct 3D vocal tract shape models for the three production types.

2.1 Method

Three adult male subjects between 22 and 30 years old participated in the MRI scan. The subjects uttered strong, weak, and very weak whispered voices of the vowel /i/, which was chosen because it is the easiest one for subjects to stabilize an articulatory position. In frontal vowels, the tongue is held in contact with the palate, thereby increasing the articulation repeatability [16]. The vocal tract shape

during each whispered voice production was scanned using phonation-synchronized MRI scans [17]. SHIMADZU-Marconi MAGNEX ECLIPS 1.5T at ATR-BAIC was used. The scanning sequence was RF-FAST. The echo time (TE) was 3 ms, and the repetition time (TR) was 12 ms. For subjects 1 and 2, MR images were recorded on a sagittal plane; each image was 512×512 pixels; field of view (FOV) was 256×256 mm; slice thickness was 1 mm; and spatial resolution was $0.5 \times 0.5 \times 1.0$ mm. For subject 3, MR images were recorded on a coronal plane; the image size was 512×512 pixels; FOV was 128×128, slice thickness was 1 mm; and spatial resolution was $0.25 \times 0.25 \times 1.0$ mm.

Because the MRI device emits noises of a very large amplitude that prevent subjects from having a feedback of their own voices, it is difficult for subjects to confirm that their voices are correctly produced especially for very weak whispered voices. Whisper sounds were recorded during the MRI scans using a special microphone for an air-conducted sound that was developed for a use in an MRI device. However, from the recorded sound signals, it is difficult to confirm that the subjects produced very weak whispered voices correctly, because the high-level MRI noises were also recorded along with them. Note that, although a NAM microphone is tolerant to an air-conducted noise, it is unavailable in the MRI device because it contains metals that cannot be used in the MRI device.

Therefore, for just before the MRI scans, the subjects were trained to be able to produce the whispered voices correctly even during the MRI scans. For a very weak whispering, subjects were instructed to utter whispered voices so that they were weak enough to be inaudible even to nearby people, which was confirmed by the experimenter in the training. Nevertheless, even with such training, it is possible that the subjects could not produce the whispered voices correctly under high-level MRI noise because of Lombard effect [18]. This problem should be solved in a future work.

2.2 3D Vocal Tract Shape Model

We constructed 3D vocal tract shape models from the MRI images of one subject who showed apparent variations in a larynx position and a glottis aperture among the three whispered voice production types. The procedure was as follows. First, the MR images were edge-reinforced. Next, they were binarized based on the edge-reinforced sagittal, coronal, and horizontal images. Finally, 3D vocal tract shape models were constructed by interpolating from the binarized volume data.

Compared to the original MR images, the 3D vocal tract shape model of a strong whispered voice had 0.42-mm error in the maximum height and 0.43-mm error in the maximum width of the piriform fossae. These errors are smaller than the MR spatial resolution, which is 0.5 mm, indicating that 3D models reasonably reproduce vocal tract shapes.

Figure 1 illustrates the 3D vocal tract shape models of strong, weak, and very weak whispered voices, hereafter labeled as *SW*, *WW*, and *VWW*, respectively. As the
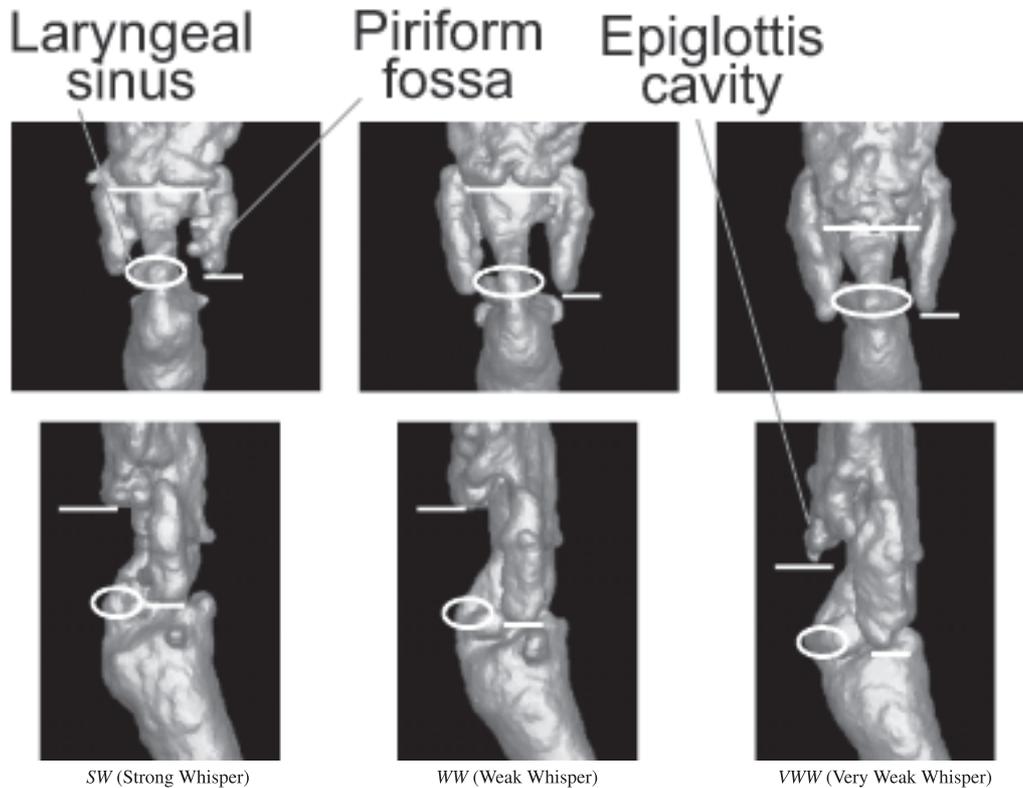
Fig. 1   3D vocal tract shape models of strong and weak whispered voices, and NAM (upper: coronal view, lower: sagittal view).

whispered voice production style varies from *SW* to *WW* to *VWW*, namely, as the intensity of the whispering decreases, it is clearly observed that the laryngeal sinus opens, the laryngeal sinus moves downward, the piriform fossae extend downward, and the epiglottis opens, showing a systematic change of the vocal tract shape with the decreasing intensity of whispering. Compared to the case of *SW*, the laryngeal sinus moves downward by 4.5 and 9.0 mm in the cases of *WW* and *VWW*, respectively. The piriform fossae extend downward by 5.0 and 7.5 mm, respectively.

3.   Numerical Simulation

A three-dimensional incompressible simulation was performed using a commercial 3D computational fluid dynamics (CFD) solver (FLUENT 6.3.26, ANSYS, Inc. [19]) based on the finite volume method.

The governing equations are the Navier-Stokes equations. The turbulence model used was large-eddy simulation (LES, Smagorinski Lilly, $C_s = 0.1$). The time difference scheme was the implicit second-order method (time step: 50 [μs]). The spatial difference scheme used was bounded central differencing. Tetrahedron cells were automatically generated from the vocal tract shape models by a mesh generator (ICEM-CFD, ANSYS, Inc.). The edge length was between 1.0 and 1.5 mm. The boundary layer adjacent to the vocal tract wall was discretized as trilaminar triangular prism meshes, the thickness of each being

0.2 mm. The boundary condition at the lung-side opening of the vocal tract was given as a temporally constant flow velocity, whose value was determined so that it matched the values of the volume flow rate described below. The boundary condition at the mouth-side opening of the vocal tract exit was given as pressure P (= 0). The air density ρ and viscosity μ were 1.225 [kg/m^3] and 1.79×10^{-5} [Pa·s], respectively. Figure 2 illustrates the numerical condition.

Expiratory volume flow rates for the three whispered voice production types were determined according to previous research [20], [21], [23]. Netsell et al. reported that the expiratory volume flow rate of an ordinary voice production is between 100 and 200 [ml/s] [21]. Rubin et al. reported that the expiratory volume flow rate of weak and strong whispered voices are double and triple that of the ordinary voice, respectively [23]. Therefore, the volume flow rates of whispered voices were determined to be 500 and 333 [ml/s] respectively for *SW* and *WW*. Because there is no experimental data for the very weak whispering *VWW*, the volume flow rate for *VWW* having much weaker energy was set to be 100 [ml/s] so that the value is smaller than those for strong and weak whispered voices. Although the volume flow rate for *VWW* should be experimentally measured for more realistic simulation, the volume flow rate of 100 [ml/s] would not deviate largely from the actual value because it is as much as the lower bound of volume flow rate for ordinary voice production [21].
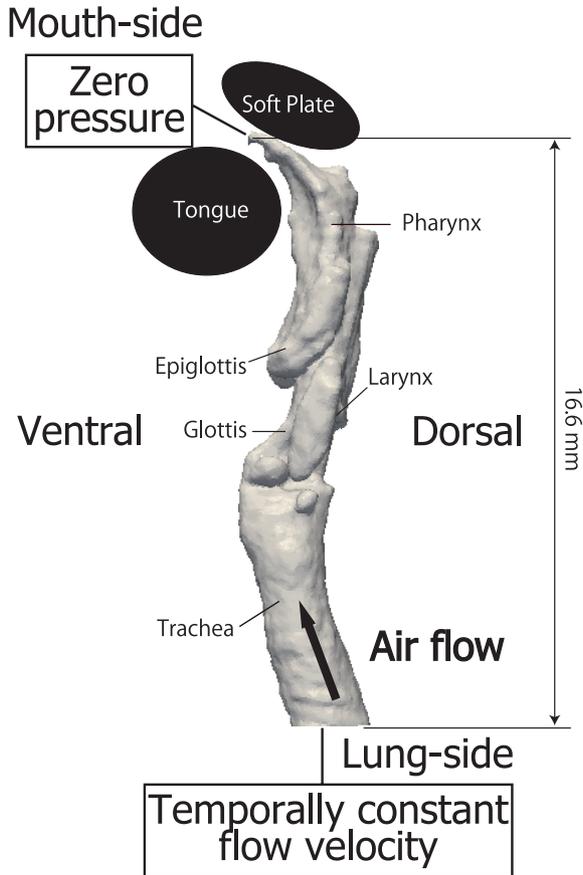
Fig. 2 Sagittal view of vocal tract shape model. Air flows from lung-side to mouth-side. Temporally constant flow velocity given at lung-side boundary. Zero pressure given at mouth-side boundary.

Table 2 Observed Reynolds number and Mach number.

	SW	*WW*	*VWW*
L [mm]	5.4	6.5	8.8
$|\mathbf{u}|_{max}$ [m/s]	11.7	7.79	1.66
Re	4.3×10^3	3.5×10^3	1.0×10^3
M_a	0.034	0.023	0.005

Table 1 Model details.

	SW	*WW*	*VWW*
Number of cells$(\times 10^3)$	363	321	365
Number of nodes$(\times 10^3)$	119	112	125
Volume flow rate [ml/s]	500	333	100

Table 1 describes the number of cells and nodes, and the volume flow rate for each whispered voice production style.

4. Results

Reynolds number at the glottis for each whispered voice production type was estimated from the model geometry and simulation results. The Reynolds number is defined as $Re = \rho |\mathbf{u}|_{max} L / \mu$, where L is a characteristic length that corresponds to the minimum width of the glottal constriction; $\mathbf{u}$ is the flow velocity vector; and $|\mathbf{u}|_{max}$ is the maximum velocity at the glottal region. Each value for *SW*, *WW*, and *VWW* is shown in Table 2. The values of L show that the minimum width of a glottal constriction increases and the maximum flow velocity decreases as the whispered voice production type varies from *SW*, *WW* to *VWW*. As a result, Re are estimated as $O(10^3)$ for all the whispered voice pro-

duction styles. Air flows with such values of Re are reasonably modeled by the fluid dynamics simulation method used in this work. Although critical Reynolds numbers for the air flow through a glottis are unknown, the values of Re observed in the simulation indicate that turbulent flows would occur in the supraglottal region for all the whispered voice production styles. Mach numbers M_a for the air flow through the glottis are also calculated from the simulation results, as shown in Table 2. The simulation shows that, for each whispered voice production types, the observed Mach numbers are much smaller than 0.3 [22], indicating that the air flows through the glottis can be modeled by an incompressible flow simulation utilized in this work.

Figure 3 demonstrates an instantaneous flow velocity distribution on sagittal and coronal sections for each whispered voice production style at 15 ms after the start of the simulation. Orientations and lengths of the arrows in the figure indicate flow directions and velocity amplitudes, respectively. The results show that, in the case of *SW*, the jet separates from a vocal-tract wall at a boundary between the glottis and larynx, but proceeds along a side wall; a circulation region is produced beside the jet on another side. In the case of *WW*, the jet separates from the wall and bifurcates: One proceeds straight; another proceeds along the side wall. Then, both the jet streams disappear and produce a turbulent flow. In the case of *VWW*, the jet bifurcates but proceeds along the wall until it reaches the entrances of the piriform fossae where it separates from the vocal-tract wall. The jets turned into a turbulent flow just after separating from the wall. The results show that, for each whispered voice production type, a turbulent flow is produced at the supraglottal region. It can be seen that vortices occur in the turbulent flow region.

Powell suggested that vorticity ω [1/s] (= rot($\mathbf{u}$)) relates to a vortex sound source, namely, a sound wave arises where a vortex exists [24]. Figure 4 shows instantaneous vorticity magnitude $|\omega|$ distributions on sagittal and coronal sections for *SW*, *WW*, and *VWW* at the same instant for Fig. 3. The gray colors represent the values of the vorticity magnitude, ranging from 0 to 4,000 for *SW* and *WW*, and from 0 to 1,000 for *VWW*, as the gray-scale bars indicate. The results generally show that the vorticity magnitude is larger at just beside the jet and the vocal-tract wall whereas its magnitude depends on the whispered voice production types, namely, it is larger in *SW* and smaller in *VWW*. The result for *SW* shows that vorticity is injected into the flow in the same manner as the velocity amplitudes shown in Fig. 3; however, the vorticity amplitude is large at the boundary between the jet and a stagnant region, namely just beside a
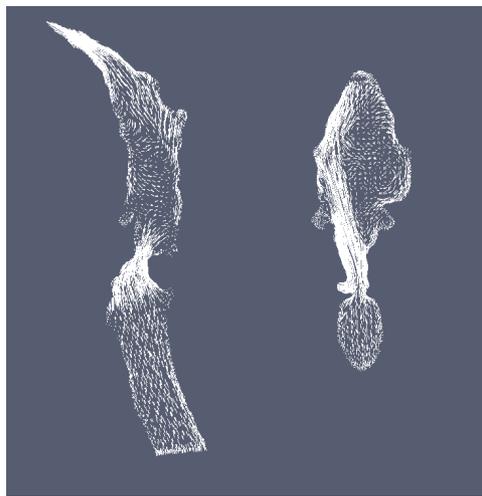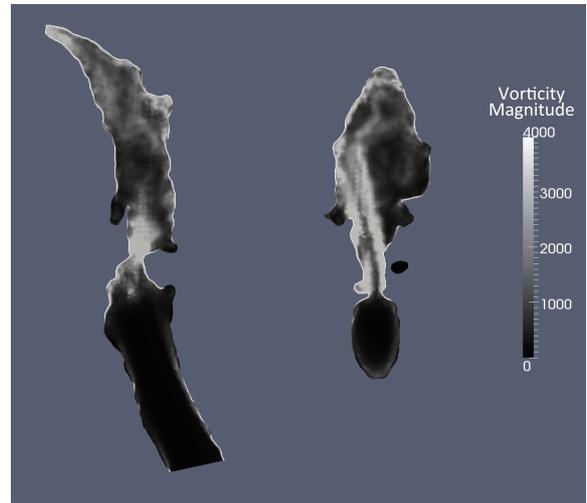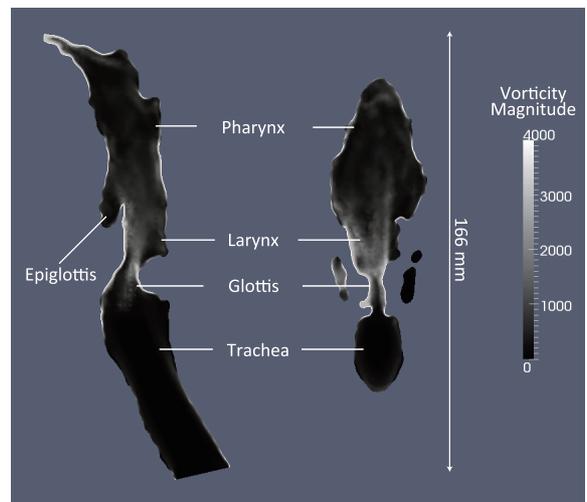
SW (Strong Whisper)
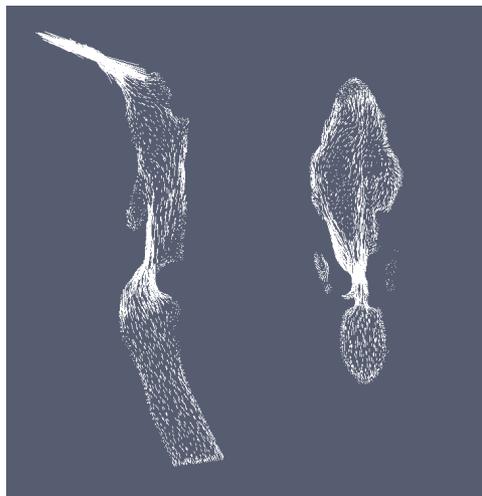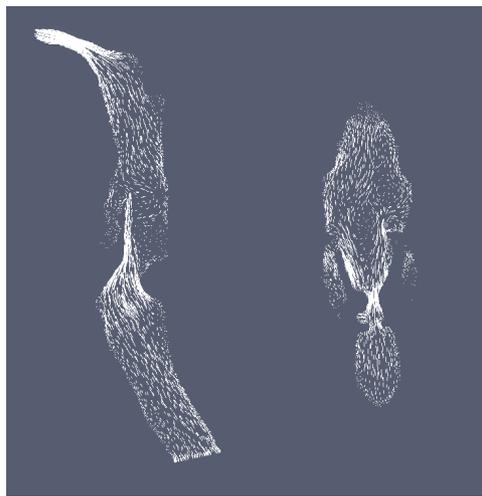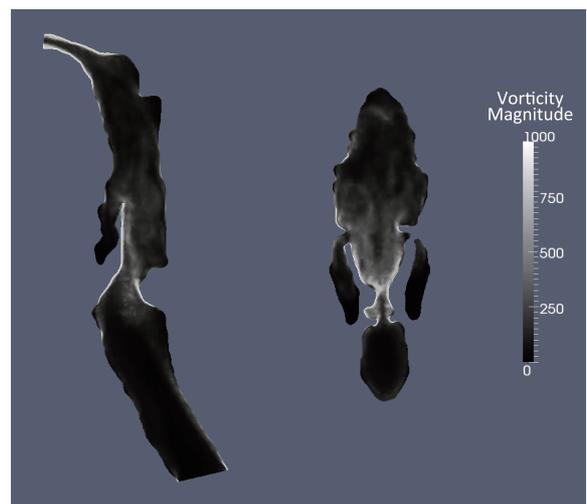

WW (Weak Whisper)

VWW (Very Weak Whisper)

Fig. 3　Instantaneous flow velocity distributions on sagittal and coronal sections for strong, weak, and very weak whispering. An orientation and length of the arrows indicates an air flow direction and velocity amplitude, respectively. Note that the arrow length for *VWW* is 4 times longer of that for *SW* and *WW*.

SW (Strong Whisper)

WW (Weak Whisper)

VWW (Very Weak Whisper)

Fig. 4　Instantaneous vorticity magnitude (in $[1/s]$) distributions on sagittal and coronal sections for strong, weak, and very weak whispering. Note that color bars for *SW* and *WW* range from 0 to 4,000, whereas that for *VWW* ranges from 0 to 1,000.

central axis of the jet, whereas it is small at the central axis of the jet. Also in the case of *WW* and *VWW*, vorticity is injected along with the jet and its amplitude is large at just beside the central axes of the bifurcated jets.

Recent work by Krane [25] indicated important features of a vorticity flow that are relevant for speech production. He clarified that an aerodynamic force is generated when an air flow velocity is not parallel to a velocity disturbance caused by a change in a vocal-tract shape that finally results in a non-rectilinear stream of vorticity flow; such an aerodynamic force is relevant for an aeroacoustic source strength. Subsequently, he pointed out that vorticity on the vocal-tract wall does not contribute to a sound production because the wall shape does not change drastically in that region; the convection of vorticity through a nonuniform vocal tract produces sound. As shown in Fig. 4, vorticity is large at the vocal-tract wall in the supraglottal region but such vorticity does not contribute to a sound generation because of above-mentioned reason. On the other hand, vorticity injected into the supraglottal space along with the jet produces a sound by interacting with the air flow velocity disturbance caused by the nonuniform vocal-tract shape.

Such vorticity flows are observed for the simulation results for all the whispered voice production types including *VWW*, although generated vorticity magnitudes are much smaller than the ordinary whispering. Considering Krane's suggestions, these results clearly show that the vorticiy flow in the supraglottal region produces turbulent noise sound sources for strong and weak whispering. The same aerodynamic mechanism produces noise sources for the very weak whispered voice production even with much weaker energy.

5. Discussions

As mentioned in 2.1, it is difficult to confirm, during and after an MRI scan, that subjects certainly uttered an "inaudible" very weak whispered voice, i.e. NAM, during an MRI scan because of high MRI noises and difficulties of a sound recording in an MRI device. However, subjects were instructed to produce an "inaudible" very weak whispered voice which is weaker than weak whispered voice. Actually, from the MRI scan results as described in 2.2, it is observed that the vocal tract shapes change systematically as the whispered voice production type varies from *SW* to *WW* to *VWW*, in other words, as the intensity of the whispering decreases. These results indicate that, for *VWW*, subjects certainly uttered a whispered voice that is weaker than *WW*. Although this does not guarantee that the subjects uttered an "inaudible" whispered voice during the MRI scans, it is at least likely that the captured vocal tract shape for *VWW* would not deviate largely from one for an inaudible very weak whispering that would be found in an ordinary situation.

In the current simulation, the volume flow rate for a very weak whispering (*VWW*) was assumed to be approximately one third of that for a weak whispering (*WW*), resulting in 1.66 m/s of the maximum flow velocity at the glottis

region for *VWW* that is approximately one fifth of that for *WW* (7.79 m/s). The validity of these values for *VWW* is open to question. No research has been performed on the minimum flow rate or flow velocity that enables a whisper production. However, it is known that, for low flow velocity, a mono-pole source, caused by a velocity disturbance such as a vortex, is dominant in sound production; a sound intensity radiated from the mono-pole source varies with the fourth power of a flow velocity [26]. Therefore, one fifth of the flow velocity would result in an approximately 28-dB decrease in a sound intensity level compared with *WW*, indicating that sufficiently weak sound sources were assumed in the simulation for *VWW*.

Because a vocal tract has a nonuniform shape, there is no fully irrotational flow in it. Therefore, in practice, vorticity is always present whenever an air flows in the vocal tract [27]. Accordingly, there is no threshold of a volume flow rate or a flow velocity to produce vorticity and, subsequently, a sound. Thus, the discussion converges to a question: Does an air pressure fluctuation, which is produced by an air flow through the glottis and is resonated in the vocal tract, have a sufficient intensity for people to hear as a sound and for a NAM microphone to detect? Addressing this question would necessitate a more comprehensive simulation including fluid dynamics, acoustics, and vibration in addition to a further investigation of a NAM microphone's sensitivity, which are future works.

Other future works are to ensure a NAM production during an MRI scan with a sound recording technique available even in an MRI device and to increase the accuracy of a numerical simulation with a measurement of an unknown volume flow rate for very weak whispering or NAM. Note that turbulent noise sources in a whispered voice production do not always occur at a supraglottal region near the glottis; namely, sources for some types of consonant such as fricative or plosive, e.g. /s/ or /k/, are produced due to vortex flows arising at an alveolar or a velar. Clarifying how the produced vorticity flow relates to the frequency characteristics of the very weak whispering should be counted as another future work.

6. Conclusion

The air flow passing through the glottis was numerically simulated using computational fluid dynamics and three-dimensional vocal tract shape models that were obtained by MRI scans for three whispered voice production types, i.e., strong, weak, and very weak whisperings. The results of the MRI scan show that the vocal tract shape changes systematically as the intensity of the whispering decreases. The numerical results reveal that, even in the very weak whispered voice production, a weak vorticity flow occurs in the supraglottal space, as in the ordinary (strong and weak) whispering; a turbulent noise due to the vorticity flow is the sound source even for a very weak whispered voice. These results suggest that the NAM, an inaudible very weak whispered voice, is produced by turbulent noise sources due to vortic-

ity flows arising in the supraglottal space, as in the ordinary whispered voice.

Acknowledgments

References

[1] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell, "Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.708–711, 2003.

[2] Y. Nakajima, H. Kashioka, N. Campbell, and K. Shikano, "Non-audible murmur (NAM) recognition," IEICE Trans. Inf. & Syst., vol.E89-D, no.1, pp.1–8, Jan. 2006.

[3] T. Hirahara, M. Otani, S. Shimizu, T. Toda, K. Nakamura, Y. Nakajima, and K. Shikano, "Silent-speech enhancement using body-conducted vocal-tract resonance signals," Speech Commun., vol.52, no.4, pp.301–313, 2010.

[4] S. Shimizu, M. Otani, and T. Hirahara, "Frequency characteristics of several types of NAM microphones," Acoust. Sci. & Tech., vol.30, no.2, pp.139–142, 2009.

[5] T. Hirahara, S. Shimizu, and M. Otani, "Acoustic characteristics of non-audible murmur," Proc. Japan-China Joint Conference on Acoustics, P-2-30, 2007.

[6] M. Otani, T. Hirahara, S. Shimizu, and S. Adachi, "Numerical simulation of transfer and attenuation characteristics of soft-tissue conducted sound originating from vocal tract," Applied Acoustics, vol.70, no.3, pp.469–472, 2009.

[7] W. Zhao, C. Zhang, S. Frankel, and L. Mongeau, "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am., vol.112, no.5, pp.2134–2146, 2002.

[8] C. Zhang, W. Zhao, S. Frankel, and L. Mongeau, "Computational aeroacoutics of phonation, Part II: Effects of subglottal pressure, glottal oscillation frequency and ventricular folds," J. Acoust. Soc. Am., vol.112, no.5, pp.2147–2154, 2002.

[9] C. Tao, J. Jiang, and Y. Zhang, "Simulation of vocal fold impact pressures with a self-oscillating finite-element model," J. Acoust. Soc. Am., vol.119, no.6, pp.3987–3994, 2006.

[10] C. Tao, Y. Zhang, D. Hottinger, and J. Jiang, "Asymmetric airflow and vibration induced by the Coanda effect in a symmetric model of the vocal folds," J. Acoust. Soc. Am., vol.122, no.4, pp.2270–2278, 2007.

[11] H. Luo, R. Mittal, and S. Bielamowicz, "Analysis of flow-structure interaction in the larynx during phonation using an immersed-boundary method," J. Acoust. Soc. Am., vol.126, no.2, pp.816–824, 2009.

[12] H. Nomura and T. Funada, "Numerical simulation of unsteady flow through the rigid glottis," Acoust. Sci. & Tech., vol.27, no.3, pp.154–162, 2006.

[13] H. Nomura and T. Funada, "Sound source generation by unsteady flow ejecting from the vibrating glottis based on a distributed parameter model of the vocal cords," Acoust. Sci. & Tech., vol.28, no.6, pp.392–402, 2007.

[14] H. Nomura and T. Funada, "Effects of the false vocal folds on sound generation by an unsteady glottal jet through rigid wall model of the larynx," Acoust. Sci. & Tech., vol.28, no.6, pp.403–411, 2007.

[15] J. Suh and S. Frankel, "Numerical simulation of turbulence transition and sound radiation for flow through a rigid glottal model," J. Acoust. Soc. Am., vol.121, no.6, pp.3728–3739, 2007.

[16] T. Kitamura, H. Takemoto, K. Honda, Y. Shimada, I. Fujimoto, Y. Syakudo, S. Masaki, K. Kuroda, N. Okuchi, and M. Senda, "Difference in vocal tract shape between upright and supine postures: Observations by an open-type MR scanner," IEICE Technical Report, SP2004-29, 2004.

[17] Y. Nota, T. Katsumura, T. Honda, K. Takemoto, H. Hirata, I. Fujimoto, Y. Syakudo, and S. Masaki, "A bone-conduction system for auditory stimulation in MRI," Acoust. Sci. & Tech., vol.28, no.1, pp.33–38, 2007.

[18] E. Lombard, "Le signe de l'élévation de la voix," Annales des Maladies de L'oreille et du Larynx, vol.37, no.2, pp.101–119, 1911.

[19] ANSYS FLUENT Flow Modeling Software http://www.ansys.com/products/fluid-dynamics/fluent/

[20] E. Holmberg, R. Hillman, and J. Perkell, "Glottal air flow and transglottal pressure measurements for male and female speakers in soft, normal and loud voice," J. Acoust. Soc. Am., vol.84, no.2, pp.511–529, 1988.

[21] R. Netsell, W. Lotz, A. DuChane, and S. Barlow, "Vocal tract aerodynamics during syllable productions: Normative data and theoretical implications," J. Voice, vol.5, no.1, pp.1–9, 1991.

[22] J.D. Anderson, Fundamentals of Aerodynamics, 4th ed., pp.483–755, McGraw Hill, New York, 2007.

[23] A. Rubin, P. Veeraphol, G. Shirley, M. Cheryl, and S. Robert, "Laryngeal hyperfunction during whispering: reality or myth?," J. Voice, vol.20, no.1, pp.121–127, 2006.

[24] A. Powell, "Theory of vortex sound," J. Acoust. Soc. Am., vol.36, no.1, pp.177–195, 1964.

[25] M.H. Krane, "Aeroacoustic production of low-frequency unvoiced speech sounds," J. Acoust. Soc. Am., vol.118, no.1, pp.410–427, 2005.

[26] S. Yoshikawa and H. Wada, Aeroacoustics of Sound Sources, Corona Publishing, Tokyo, 2007.

[27] C. Pozrikidis, Fluid Dynamics: Theory, Computation, and Numerical Simulation, Springer, New York, 2009.

Makoto Otani received the B.Eng., M.Eng., and Dr. Eng. degrees in architectural engineering from Kyoto University, Kyoto, Japan, in 1999, 2001, and 2006, respectively. He was a postdoctoral researcher at Faculty of Engineering, Toyama Prefectural University, from 2006 to 2008, and a Global COE fellow at Graduate School of Engineering, Tohoku University, from 2008 to 2009. He has been an assistant professor from 2009 to 2011, and an associate professor since 2011, at Department of Information Engineering, Shinshu University. He is a member of the Acoustical Society of Japan, the Acoustical Society of America, and the Virtual Reality Society of Japan.

Tatsuya Hirahara received the B.Eng., M.Eng., and Dr. Eng. degrees in electronic engineering from Hokkaido University, Sapporo, Japan, in 1978, 1980, and 1983, respectively. He worked for NTT-ECL Yokosuka, ATR Auditory and Visual Perception Labs., NTT Basic Research Labs., ATR-International Corporate Planning Division, NTT Communication Science Labs., and ATR Human Information Science Lab. He was also a visiting researcher at the Speech Communication Group of the MIT-RLE and the Eaton-Peabody Laboratory of Auditory Physiology at Massachusetts Eye and Ear Infirmary. Since 2006, he has been Professor of Intelligent Systems Design Engineering at Toyama Prefectural University. Dr. Hirahara is a member of the Acoustical Society of Japan, the Acoustical Society of America, and the Virtual Reality Society of Japan.