Acoust. Sci. & Tech. **29**, 2 (2008)

ACOUSTICAL LETTER

Vocal tract shapes of non-audible murmur production

Makoto Otani*, Shota Shimizu and Tatsuya Hirahara

Faculty of Engineering, Toyama Prefectural University, Kurokawa 5180, Imizu, Toyama, 939–0398 Japan

(*Received 30 October 2007, Accepted for publication 29 November 2007*)

Keywords: Non-audible murmur, Vocal tract shape, Whispered voice, Magnetic resonance imaging

PACS number: 43.70.Aj [doi:10.1250/ast.29.195]

1. Introduction

Researches on non-audible murmur (NAM), a very weak speech sound produced without vocal cord vibration, have been conducted to develop a new speech communication tool. NAM can be detected using a specially designed microphone, called a NAM microphone [1]. This device is a condenser microphone covered with a soft impression material such as soft silicon or urethane elastomer, which provides better impedance matching between biological soft tissues and condenser microphone diaphram. With the NAM micophone attached to the neck surface close behind an ear, the NAM can be detected as a body-conducted voice. The recorded NAM, however, indicates that the signal exhibits low S/N ratio and severely suppressed higher-frequency component, resulting in a lack of clarity, compared to an air-conducted ordinary voice [2].

A numerical simulation clarified that the suppressed higher-frequency components are due to a transmission to biological soft tissues and a propagation through them [3]. However, the NAM production mechanism is yet to be revealed. Vocal tract shapes of NAM production are not clarified, whereas those of whispered voice production were revealed using magnetic resonance imaging (MRI) [4,5]. This letter reports MRI-scanned vocal tract shapes of whispered voice and NAM production and comparisons among them to investigate the NAM production mechanism.

2. Method

Three adult male subjects between 22 and 30 years old participated in the MRI scan. Subjects uttered strong and weak whispered voice (SW and WW), and NAM of vowel /i/. A vowel /i/ was chosen because it is the easiest one for subjects to stabilize an articulatory position. In frontal vowels, the tongue is hold by contact with the palate, thereby increasing the articulation repeatability [6]. A vocal tract shape of each voice production was scanned using phonation-synchronized MR imaging scans [7]. SHIMADZU-Marconi MAGNEX ECLIPS 1.5T at ATR-BAIC was used. The scanning sequence was RF-FAST. TE (Echo Time) was 3 ms, and TR (Repetition Time) was 12 ms. For subjects 1 and 2, MR images were recorded on a sagittal plane; image size was 512×512 pixel; field of view (FOV), 256×256 mm; slice thickness, 1 mm; spatial resolution, $0.5 \times 0.5 \times 1.0$ mm. For subject 3, MR images were recorded on a coronal plane; image size was 512×512 pixel; FOV, 128×128 mm; slice thickness, 1 mm; spatial resolution, $0.25 \times 0.25 \times 1.0$ mm.

3. Results

3.1. MR images

Images on the right of Fig. 1(a) show an example of subject 1's MR images on a coronal plane. A white line in an MR image on a sagittal plane, on the left of Fig. 1(a), represents the position of coronal plane examined. In the coronal plane images, a glottis and piriform fossae are bounded by white lines. It is observed that the glottis gradually opens; the glottis moves downward, as voice production style varies from a strong whisper to a weak whisper, and finally to the NAM (SW $\rightarrow$ WW $\rightarrow$ NAM). Cavities, which are found beside the glottis for the case of a weak whisper and NAM, are piriform fossae. It is also observed that the piriform fossa extend downward, as voice production style varies from SW to NAM.

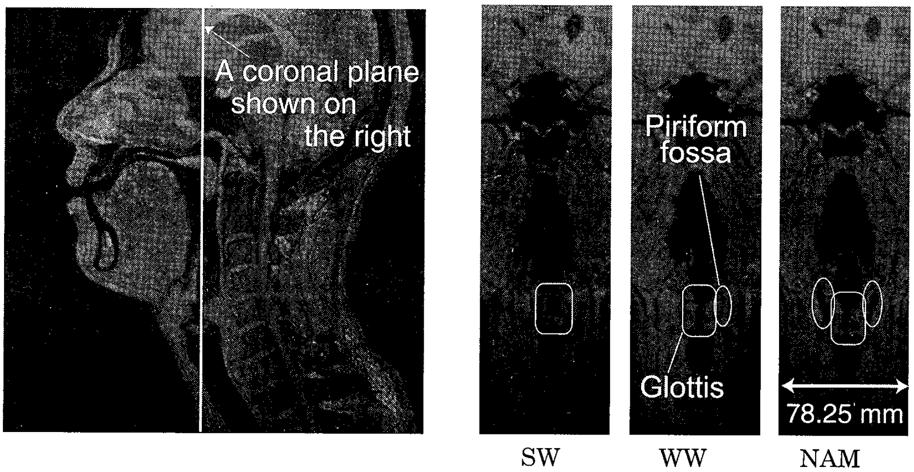
Figure 1(b) shows the corresponding results of subject 2. It is observed that the glottis gradually opens; the glottis, however, does not move downward; the piriform fossae do not extend downward.

Figure 1(c) shows the results of subject 3. Note that, for subject 3, MR images were recorded on a coronal plane at a smaller region but in a higher resolution than subjects 1 and 2. For a strong whisper, the glottis image seems to be distorted, probably due to an instability in the subject's production style or posture during MRI scans. Hence it is difficult to find the actual glottis position for a strong whisper. As for other production styles, it is observed that, compared to a weak whisper, a NAM production yields a narrower glottis opening, which is inconsistent with other subjects. Further, the piriform fossae do not show a significant variation in their positions between a weak whisper and the NAM.

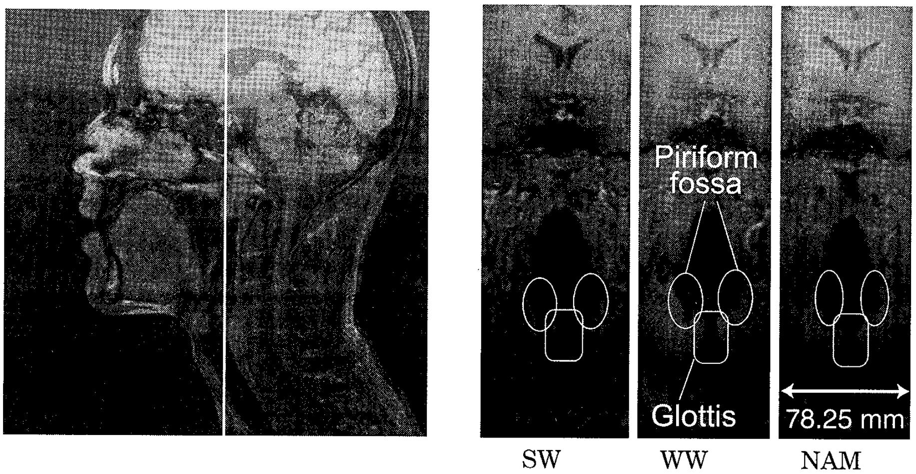
3.2. 3D model of vocal tract shapes

In order to investigate vocal tract shapes of whispered voice and NAM productions in detail, 3D vocal tract shape models were constructed from subject 1's MR images by using MECHANICAL FINDER (Research Center of Computational Mechanics, Inc). The procedure is as follows. Firstly, the MR images were edge-reinforced. Next, MR images 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.

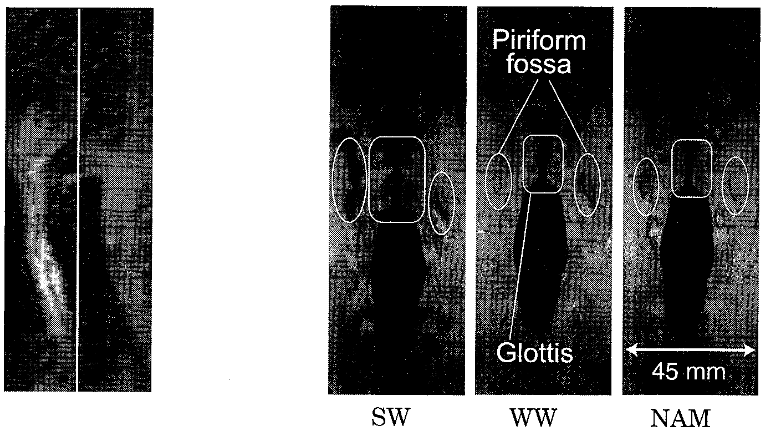
*e-mail: otani@pu-toyama.ac.jp



(a) Subject 1



(b) Subject 2



(c) Subject 3

Fig. 1 MR images on coronal planes, which correspond to white lines on the sagittal plane image on the left.

Compared to the original MR images, the 3D model constructed for strong whispered voice has 0.42 mm error in maximum height and 0.43 mm error in maximum width in the part of piriform fossae. These errors are less than the MR spatial resolution, that is 0.5 mm, thereby indicating that 3D models accurately reproduce the vocal tract shapes.

Figure 2 shows the results. As the voice production style varies from strong whispered voice, weak whispered voice, to NAM (SW → WW → NAM), it is clearly observed that a laryngeal sinus opens; a laryngeal sinus moves downward; piriform fossae extend downward; an epiglottis opens. Compared to the case of the strong whispered voice, the laryngeal sinus moves downward by 4.5 mm and 9.0 mm in the cases of the weak whispered voice and NAM, respectively. The piriform fossae extend downward by 5.0 mm and 7.5 mm, respectively.

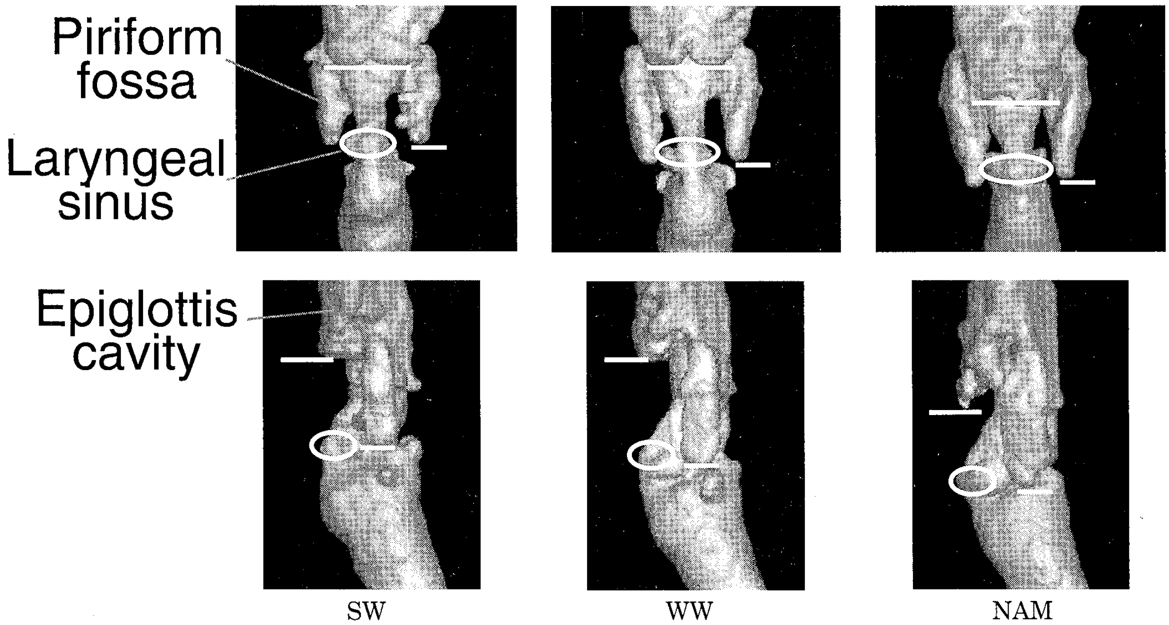


Fig. 2 Subject 1's vocal tract shapes for strong and weak whisper, and NAM (upper: coronal view, lower: sagittal view).

4. Discussion

Some researches were conducted to reveal differences in vocal tract shapes of normal and whispered voice productions. From MR images of single subject, Matsuda *et al.* reported that, compared to a normal voice production, a whispered voice production of vowel /i/ yields a downward movement of a laryngeal sinus [4]. Further, also from MR images of single subject, Honda *et al.* reported that whispered voice production of vowel /a/ narrows a laryngeal cavity, and shrinks piriform fossae, compared to a normal voice production [5]. From subject 1's results in the current work, it is observed that the NAM production yields a downward movement of a larynx sinus, a glottis opening, and downward extension of piriform fossae, compared to a whispered voice production. Accordingly, as to subject 1's results, it can be said that the vocal tract shape of the NAM production is more similar to that of a normal voice production than a whispered voice one. Other subjects, however, show inconsistent results; subject 2's vocal tract shape of NAM production does not exhibit a downward movement of glottis and a downward extension of piriform fossae, thereby showing few similarities to that of a normal voice production; subject 3's vocal tract shape of NAM production also does not exhibit a downward movement of glottis, a downward extension of piriform fossae, and a glottis opening, indicating fewer similarities to that of a normal voice production. These results indicate a diversity of NAM production style among subjects in the current work.

There are two possible explanations for the diversity of NAM production style among subjects. Firstly, a high noise level, to which subjects were exposed during MRI scans, might prevent them from obtaining one's own vocalized sound feedback, resulting in diverse production styles. It is hard for subjects to hear their own NAM and even whispered voice by themselves during MRI scans. This issue should be taken into account although it is difficult to solve it unless quieter MRI device is developed. Secondly, NAM production style could be essentially diverse among subjects. For example, NAM can be produced by both a normal and a whispered voice production style with significantly reduced expiratory airflow. Nakajima defined NAM, "Non-Audible Murmur," as "a non-voiced speech sound, created by turbulent airflow generated in the glottis and articulated in the vocal tract by speech-like movements of the tongue, lips, and jaw. It is similar to whisper, but is generally inaudible to persons other than the speaker" [8]. According to this definition, any specific production style is not required for NAM production; only if one utters a very weak voice that is unheard by nearby people, the voice is the NAM. As well as controlling of a voice production style during MRI scans, the diversity of NAM production style among subjects should be taken into account in future experiments.

Honda *et al.* reported that the laryngeal sinus shrank and almost closed when a subject uttered whispered vowel /a/ [5]. However, the corresponding 3D vocal tract shape models, as shown in Fig. 2 SW and WW, do not exhibit such a significant laryngeal sinus shrinkage. This discrepancy is possibly caused by a difference in 3D model construction technique between [5] and the current work, and/or a variation in voice production style among subjects.

5. Summary

Vocal tract shapes of a whispered voice and NAM production were obtained by MRI scans. Results show that the vocal tract shapes are diverse among subjects. One subject's results show that a NAM production yields opening and downward movement of glottis as well as downward extension of piriform fossae, compared to a whispered voice production.

Finally, the 3D models of vocal tract shapes, which were constructed in this work, can be applied to computational aeroacoustics (CAA) in order to clarify the NAM production mechanism.

Acknowledgements

This work is supported by SCOPE-S of the Ministry of Internal Affairs and Communications of Japan.

References

[1] Y. Nakajima, H. Kashioka, N. Campbell and K. Shikano, "Non-audible murmur (NAM) recognition," *IEICE Trans. Inf. Syst.*, **E89-D**, 1–8 (2006).

[2] T. Hirahara, S. Shimizu and M. Otani, "Acoustic characteristics of non-audible murmur," *Proc. Jpn.-Chin. Jt. Conf. Acoustics*, P-2-30 (2007).

[3] M. Otani, T. Hirahara and S. Adachi, "Numerical simulation of attenuation characteristics of soft-tissue conducted sound originated from vocal tract," *Proc. 19th Int. Congr. Acoustics*, CAS-04-006, Madrid (2007).

[4] M. Matsuda, H. Mori and H. Kasuya, "Formant structure of whispered vowels," *J. Acoust. Soc. Jpn. (J)*, **56**, 477–487 (2000).

[5] K. Honda, H. Takemoto, Y. Nakajima, S. Adachi and T. Hirahara, "Vocal tract shapes for whispered vowels with varyingpitches," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 349–350 (2006).

[6] T. Kitamura, H. Takemoto, K. Honda, Y. Shimada, I. Fujimoto, Y. Syakudo, S. Masaki, K. Kuroda, N. Oku-uchi and M. Senda, "Difference in vocal tract shape between upright and supine postures observations by an open-type MR scanner," *IEICE Tech. Rep. Speech*, 104(149) pp. 1–6 (2004).

[7] Y. Nota, T. Katsumura, T. Honda, K. Takemoto, H. Hirata, I. Fujimoto, Y. Shakudo and S. Masaki, "A bone-conduction system for auditory simulation in MRI," *Acoust. Sci. & Tech.*, **28**, 33–38 (2007).

[8] Y. Nakajima, "NAM interface communication —Development and evaluation of flesh conduction voice sensors as the basis—," *Doct. Diss. Nara Inst. Sci. Technol.* (2005).