Numerical Simulation of Transfer and Attenuation Characteristics of Soft-tissue Conducted Sound Originating from Vocal Tract

Makoto Otani [1], Tatsuya Hirahara [1], Shota Shimizu [1],

and Seiji Adachi [2]

[1] Faculty of Engineering, Toyama Prefectural University,

Kurokawa 5180, Imizu, Toyama, 939-0398, Japan

[2] Fraunhofer Institute for Building Physics,

Nobelstrasse 12, 70569 Stuttgart, Germany

Corresponding author: Makoto Otani
Tel: +81-766-56-7500 ext.542
Fax: +81-766-56-8030
E-mail: otani@pu-toyama.ac.jp

ABSTRACT

A non-audible murmur (NAM), a very weak speech sound produced without vocal cord vibration, can be detected by a special NAM microphone attached to the neck, thereby providing a new speech communication tool for functional speech disorders as well as human-to-machine and human-to-human interfaces with inaudible voice input for use with unimpaired. The NAM microphone is a condenser microphone covered with soft-silicone impression material that provides good impedance matching with the soft tissues of the neck. Because higher-frequency components are suppressed severely, however, the NAM detected with this device can be insufficiently clear. To improve NAM clarity, the mechanism of NAM production as well as the transfer characteristics of the NAM in soft neck tissues must be clarified. We have investigated sound propagation from the vocal tract to the neck surface, using a finite difference time domain method and a head model based on magnetic resonance imaging scans. Numerical results show that, compared to air-conducted sound detected in front of a mouth, soft-tissue-conducted sound attenuates 50 dB at 1 kHz, which consists of 30 dB full-range attenuation due to air-to-soft-tissues transmission loss and -10 dB/octave spectral decay due to a propagation loss in soft tissues. The decay agrees well with the spectral characteristics of the measured NAM.

Keywords: non-audible murmur, finite difference time domain method, soft tissue, spectral decay

PACS: 43.35.Mr

1. INTRODUCTION

The ordinary voice is an air-conducted sound, which is a vocal-tract resonance of a vocal cord vibration due to an air stream from lungs. Other than the ordinary voice, we are using diverse types of voice in daily life such as singing voice, shouting voice, laughing voice, crying voice, small voice, whispered voice, and so on. Further, we hear our own voice through bone conduction, but other's voices through air conduction.

Typically, the voice sound we use is air-conducted sound, which is vibration of the air. Meanwhile, vibration of the air in the vocal tract vibrates the vocal-tract wall. Some sound energy transmits to tissues and bones of the neck. The bone-conducted voice can be picked up with a bone-conduction microphone. As is the bone-conducted voice, a body-conducted voice, that is a voice propagating through neck muscles, veins, and other biological tissues, can be detected with a special microphone invented by Nakajima [1]. He also found that weak murmured voice, which is usually unheard by people nearby, can be detected using the special microphone attached to the neck surface close behind the ear [1-3]. Such body-conducted weak murmured voice sounds are named non-audible murmur (NAM).

A specially designed microphone to detect the NAM is named NAM microphone. This device is a condenser microphone covered with soft impression material such as soft silicon and urethane elastomer, which provides good impedance matching with the soft tissue of the neck. 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-machine and human-to-human interfaces whose inputs are inaudible voice, thereby providing a "silent" communication tool. The NAM microphone will also be able to revive the speech communication of those with vocal cord problems caused by laryngeal cancer, nerve disorders and muscle diseases.

The NAM was first recorded in 2003 by Nakajima [1]. Subsequently, Hirahara *et al.* recorded the NAM to clarify its acoustical characteristics; characterization of the NAM indicates that the signal exhibits severely suppressed higher-frequency components, and does not have sufficient clarity [4]. In order to improve the clarity of NAM detection, the NAM production mechanism as well as the sound transfer characteristics of the NAM in soft neck tissues must both be investigated. To reveal transfer and attenuation characteristics of NAM propagation through soft tissues, we investigated sound propagation from the vocal tract to neck surface, using a head model constructed on the basis of magnetic resonance imaging scans and the finite difference time domain (FDTD) method.

2. A WAVE PROPAGATING THROUGH BIOLOGICAL TISSUES

A wave propagating through biological tissues can be modeled using a linear viscoelastic wave equation [5]. Assuming anisotoropy, physical parameters, which characterize the biological tissues, are the density ρ and the first and second complex Lamé constants $\lambda = \lambda_1 + i\omega\lambda_2$ and $\mu = \mu_1 + i\omega\mu_2$, where λ_1 is the bulk modulus; λ_2, the bulk viscosity; μ_1, the shear modulus; μ_2, the shear viscosity; i, the imaginary unit; ω, the angular frequency. A sound wave in the air corresponds to a linear viscoelastic wave with $\mu_1 = 0$.

3. HEAD MODEL

Three-dimensional geometrical data of a human head uttering vowel /e/ NAM were obtained using phonation-synchronized magnetic resonance imaging scans [6]. A two-dimensional image of the median sagittal plane was extracted (Fig. 1a). Subsequently, for simplicity, a homogeneous head model was generated (Fig. 1b), *i.e.*, the head is approximated as being composed only of soft tissue that is muscle tissue. The vocal tract was replaced by a simplified model with a rectangular cross-section 30 mm wide. This latter approximation was made because a scanned vocal tract is extremely narrow in the vicinity of the vocal cords, and would require too fine of a grid to be simulated by the FDTD method.

4. NUMERICAL SIMULATION

Cyber Logic Wave 2000 Pro [7] was used as a two-dimensional FDTD solver. The simulated region is a rectangular area, which includes the head model of the following dimensions: 414 mm (width) x 292 mm (height), with a 2-mm grid. The edges of the region are an infinite boundary, *i.e.* absorptive boundary. Physical parameters of soft tissue were set as follows; density ρ, 1,100 kg/m^3; bulk modulus λ_1, 2,600 MPa; bulk viscosity λ_2, 0.001 Pa·s; shear modulus μ_1, 0.025 MPa; shear viscosity μ_2, 1,500 Pa·s, according to the reference [5]. Note that μ_2 was 100 times larger than the value found in the reference, in order to avoid numerical divergence. Also note that the value of λ_2 is determined from those of other soft materials, because the value of λ_2 for soft tissue is not known. Corresponding parameters of air are as follows; ρ, 1.24 kg/m^3; λ_1, 0.147 MPa; λ_2, 0.13 Pa·s; $\mu_1 = \mu_2 = 0$.

Figure 2 shows the geometry of the simulated region including the head, a sound source, and the receivers. Ideally, the sound source would be located near the vocal cords, as in the real human head; unfortunately, due to a limitation of numerical modelling, it cannot be located there. Instead, the sound source is located 100 mm away in front of the mouth, assuming a reciprocal theorem. The sound source is a pulse-driven vibrating plate modelled as a line of 60 mm length. Reflection from the vibrating plate would not have a significant effect on the transfer functions, because the

source distance is relatively large enough compared to the width of the mouth opening. Therefore, it is assumed that the pipe-shaped vocal tract has open end at the mouth. The receivers are located at the inlet and outlet of the vocal tract and at the neck surface close behind the ear, *i.e.*, where the NAM microphone would be attached. These receivers are referred to as "**out**", "**in**", and "**nam**", respectively. Note that **nam** is located inside the head in 2-D modelling, although it is located on the neck surface in the real human situation.

5. RESULTS

Figure 3 shows the sound intensity level observed at each receiver. Sound intensity levels were approximately obtained by dividing the square of sound pressure at each receiver by acoustical impedance at each medium. $\boldsymbol{I_{out}}$, $\boldsymbol{I_{in}}$, and $\boldsymbol{I_{nam}}$ denote the sound intensity levels at **out**, **in**, and **nam**, respectively. $\boldsymbol{I_{out}}$ and $\boldsymbol{I_{in}}$ are relatively suppressed at lower frequencies, which reflect the frequency characteristics of the particle velocity pulse used as a sound source signal. $\boldsymbol{I_{nam}}$, however, is flat. $\boldsymbol{I_{out}}$ is larger at higher frequencies than $\boldsymbol{I_{in}}$ due to characteristics of sound radiation from the mouth. $\boldsymbol{I_{nam}}$ has spectral peaks at 6.3 and 9.7 kHz. These peaks can be attributed to eigenfrequencies of the head [8]. Figure 3 also shows that $\boldsymbol{I_{nam}}$ has some spectral zeros, for examples, at 100 Hz, 600 Hz, and 6.8 kHz. It is likely that these spectral zeros might be produced due to interference of sound in the entire head structure, although the current results does not provide sufficient information to clarify what causes these spectral zeros.

In order to clarify a transmission loss of sound from air to soft tissue and a propagation loss through soft tissues, $\boldsymbol{I_{in}}$ / $\boldsymbol{I_{out}}$ and $\boldsymbol{I_{nam}}$ / $\boldsymbol{I_{out}}$ were calculated (Fig. 4). Assuming the reciprocal theorem, $\boldsymbol{I_{in}}$ and $\boldsymbol{I_{out}}$ correspond to the transfer functions from **in** and **out**, respectively, to the source position. Consequently, $\boldsymbol{I_{in}}$ / $\boldsymbol{I_{out}}$ corresponds to the vocal tract transfer functions, showing spectral peaks, *i.e.* formants, at 0.5, 1.4, 2.1, and 3.0 kHz. Further, ($\boldsymbol{I_{nam}}$ / $\boldsymbol{I_{in}}$) / ($\boldsymbol{I_{out}}$ / $\boldsymbol{I_{in}}$), which equals to $\boldsymbol{I_{nam}}$ / $\boldsymbol{I_{out}}$, corresponds to the sound intensity level observed at **nam** if a sound radiated from the mouth has flat characteristics. Compared to the sound intensity level detected at **out**, those detected at **nam** ($\boldsymbol{I_{nam}}$ / $\boldsymbol{I_{out}}$) is approximately 50 dB lower at 1 kHz. Furthermore, the sound intensity level at **nam** shows -10 dB per octave spectral decay at higher frequencies.

6. DISCUSSION

Sound energy originating in the vocal tract propagates from the air inside the vocal tract to the soft tissues surrounding the vocal tract. Subsequently, it propagates inside the soft tissues and reaches to the NAM microphone, traveling in soft tissues. The numerical simulation demonstrates 50 dB attenuation of 1-kHz sound detected at the NAM microphone, compared to that at the outlet of the vocal tract. Within the soft tissues in the current 2D numerical model, the shortest propagation distance is 70 mm. According to the physical parameters that are found in [5], the acoustic impedance of air Z_{air} and soft tissues (muscle) Z_{tissue} are approximately 420 and 1,690,000

kg/m^{2}s, respectively. These values yield transmission loss from air to soft tissues ($4\ Z_{air}\ Z_{tissue}\ /\ (Z_{air}\ +\ Z_{tissue}\ )^2$) of about 30 dB attenuation in sound intensity level at all frequencies. Furthermore, the sound attenuates during propagating in the soft tissues by 20 dB at 1 kHz due to a propagation loss which has a -10 dB/octave spectral decay, as observed in the numerical results. These transmission and propagation loss can reasonably explain the attenuation characteristics of the NAM detected at the neck surface.

Figure 5 shows the long-term averaged spectrum (LTAS) of the recorded NAM uttered by six speakers. Subjects read 50 ATR phoneme balanced Japanese sentences in NAM style [9]. A soft-silicon type NAM microphone was used to record NAM signals. In total, about 40 minutes NAM signals were recorded to compute LTAS. According to this result, LTAS of the raw NAM signals shows a -23 dB/octave spectral decay, which involves glottal sound-source characteristics having a -12 dB/octave spectral decay [4]. This result indicates that the measured NAM involves -11 dB/octave spectral decay due to propagation loss in the soft-tissue, which is roughly parallel to the spectral decay observed in the current numerical result, *i.e.* -10 dB/octave. An LTAS for normal speech mainly reflects spectra of voiced sound, such as vowels, liquids, nasals *etc*, since they frequently appear and have high energy. Contribution of unvoiced consonants, such as /p/, /t/, /k/, /s/, /sh/, /h/ *etc*, is relatively small. Like the case of normal speech, unvoiced vowels, whose source locates around the glottis, mainly contribute to the

LTAS of NAM, whereas unvoiced consonants, whose source locates other than the glottis, have relatively small effect on it. Therefore, the LTAS of NAM can be directly compared with the numerical results, although non-glottal source would produce different spectrum.

The frequency dependency of the soft-tissue propagation loss is inconsistent with the spectral decay derived from an absorption coefficient of soft tissues in the ultrasonic range, *i.e.* 0.5–2 dB/cm/MHz [10]. Applying this absorption coefficient to the audible range, there would be virtually no spectral decay due to soft-tissue propagation. However in the audible range, an absorption coefficient of soft tissues is yet to be clarified. This issue remains to be solved in the future.

7. CONCLUSIONS

In order to clarify the attenuation characteristics of the NAM detected by a NAM microphone, voice sound propagation characteristics from the vocal tract to the neck surface is numerically simulated. Compared to the sound detected in front of the lips, the sound detected at the NAM microphone attenuates by 50 dB at 1 kHz and has a -10 dB/octave spectral decay at higher frequencies. The attenuation of 50 dB at 1 kHz consists mainly of the air-to-soft-tissue transmission loss of 30 dB and the -10 dB/octave spectral decay due to propagation in the soft tissue, showing good agreement with the acoustic analysis of the measured NAM.

ACKNOWLEDGEMENTS

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

References

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

2. Nakajima Y. Development and evaluation of soft silicone NAM microphone. Technical Report of Institute of Electronics, Information and Communication Engineers 2005; 105(97):7–12.

3. Nakajima Y, Kashioka H, Campbell N, Shikano K. Non-audible murmur (NAM) recognition, Institute of Electronics, Information and Communication Engineers, Transactions on Information and Systems 2006; E89-D(1):1–8.

4. Hirahara T, Shimizu S, Otani M. Acoustic characteristics of non-audible murmur. Proceeding of Japan-China Joint Conference on Acoustics 2007; P-2-30.

5. Oestreicher HL. Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics. Journal of the Acoustical Society of America 1951; 23(6):707–714.

6. Nota Y, Kitamura T, Honda K, Takemoto H, Hirata H, Shimada Y, Fujimoto I, Shakudo Y, Masaki S. A bone-conduction system for auditory stimulation in MRI. Acoustical Science and Technology 2007; 28(1):33–38.

7. http://www.cyberlogic.org/

8. Fujisaka Y, Nakagawa S, Ogita T, Tonoike M. Analysis of wave propagation for bone-conducted ultrasonic in the heterogeneous human head model. Technical Report of Institute of Electronics, Information and Communication Engineers 2004; 103(608):13–17.

9. Abe M, Sagisaka Y, Umeda T, Kuwabara H. Speech database user's manual. ATR Interpreting Telephony Research Laboratories 1990; TR-I-0166.

10. Goldman DE, Hueter TF. Tabular data of the velocity and absorption of high-frequency sound in mammalian tissues. Journal of the Acoustical Society of America 1956; 28(1):35–37.

Figure captions

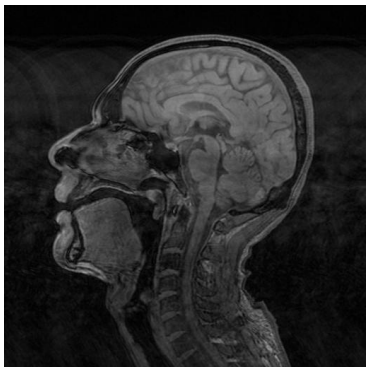
Fig. 1: (a) MR image at median saggital plane; (b) Two-dimensional head model

Fig. 2: Simulated region including the head, vibrating plate, and receiving points.

Fig. 3: Sound intensity level at outlet and inlet of vocal tract, and NAM microphone.

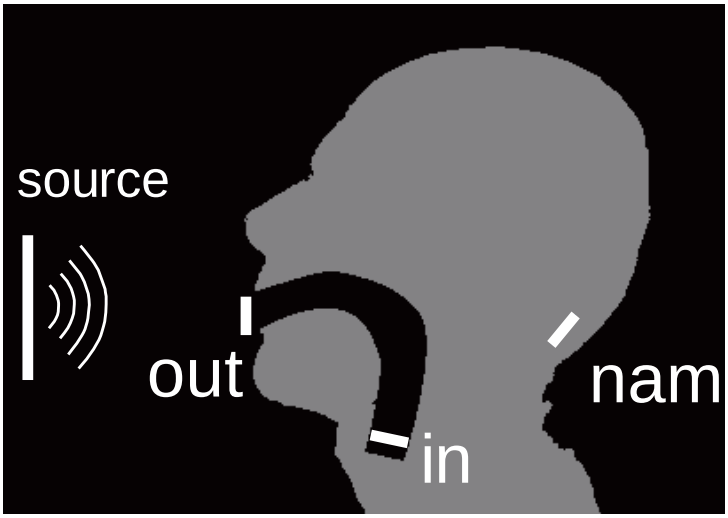
Fig. 4: Attenuation characteristics at NAM microphone.

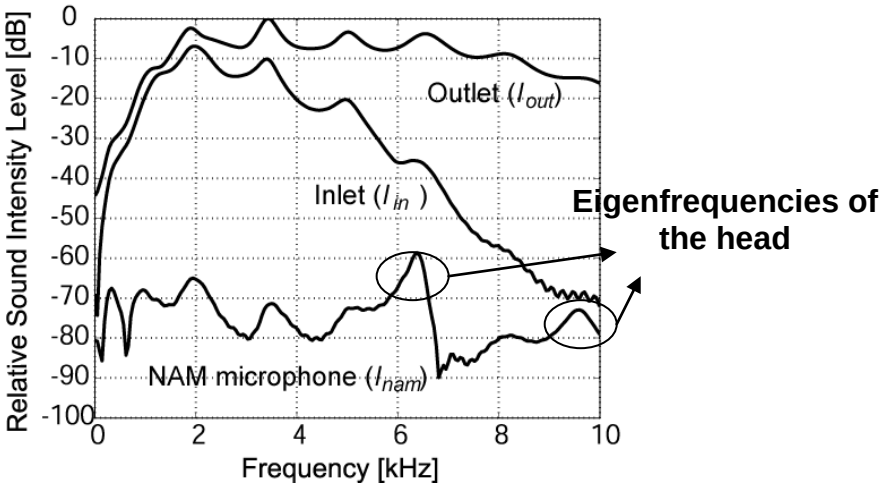
Fig. 5: Long term averaged spectrum of the recorded NAM.

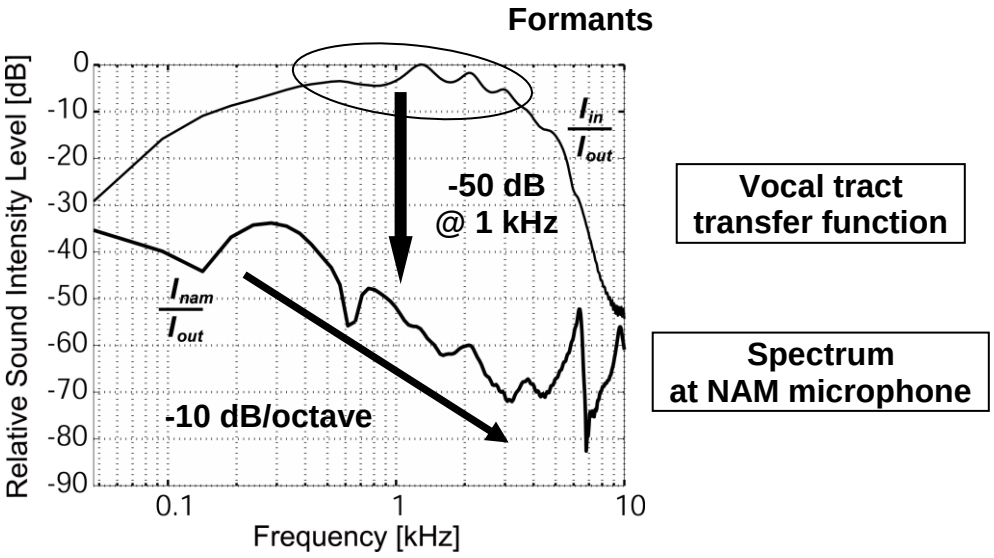


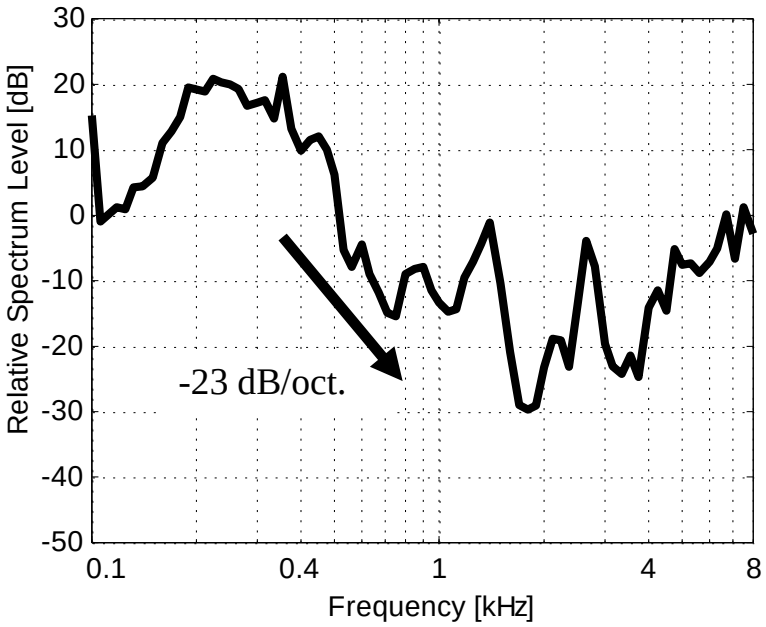
(a)

(b)


source
out
in
nam


Relative Sound Intensity Level [dB]
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
0
2
4
6
8
10
Frequency [kHz]
Outlet (l_{out})
Inlet (l_{in})
NAM microphone (l_{nam})
Eigenfrequencies of the head


Formants
Relative Sound Intensity Level [dB]
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
0.1
1
10
Frequency [kHz]
$\frac{I_{in}}{I_{out}}$
-50 dB
@ 1 kHz
$\frac{I_{nam}}{I_{out}}$
-10 dB/octave
Vocal tract
transfer function
Spectrum
at NAM microphone


30
20
10
0
-10
-20
-30
-40
-50
Relative Spectrum Level [dB]
-23 dB/oct.
0.1
0.4
1
4
8
Frequency [kHz]