JCA2007 4-6 June
The Japan-China Joint Conference of Acoustics 2007

ACOUSTIC CHARACTERISTICS OF NON-AUDIBLE MURMUR

Tatsuya Hirahara, Shota Shimizu and Makoto Otani
Faculty of Engineering, Toyama Prefectural University

ABSTRACT

Acoustic characteristics of Non-Audible Murmur (NAM) are clarified. NAM is a very weak whispered voice which can be detected by a NAM sensor attached to the surface of the skin close behind the ear. NAM is inaudible body-conducted sound whereas normal voice is audible air-conducted sound. NAM signals are recorded from 6 male and 7 female speakers using a soft-silicon type NAM microphone. The signal-to-noise ratio of the NAM signal is around 15 dB. The bandwidth of the NAM signal is 3 kHz. Long-term averaged spectra (LTAS) of the NAM signal show energy concentration around 500 to 800 Hz and a -17 dB/octave roll off. The calibration of NAM microphones reveals that a soft-silicon type NAM microphone has a low sensitivity above 3 kHz, whereas a urethane-elastomer type NAM microphone has a good sensitivity up to 10 kHz. After compensating the NAM microphone response, the LTAS of the raw NAM signal show energy concentration around 200 to 300 Hz and a -23 dB/octave roll off, which is roughly parallel to the numerical simulation of transfer characteristics of sound in head and neck.

INTRODUCTION

The ordinary voice is the air-conducted sound, which is the vocal-tract resonance of the vocal cords vibration due to the air stream from the 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 others' voices conducted through the air. This is why our voice sounds different when it is played back from a recording. These diverse types of voices can be classified based on two axes (Fig. 1).

One axis is the phonation type of the voice sound source. The sound source is either periodic glottal flow or random noise flow provided by the air stream passing through the glottis. Stronger air stream is needed to shout and to sing, while weaker air stream is enough to whisper. The sound source is then filtered by the vocal-tract configuration and radiated into the air from the mouth. This axis also represents the energy of voice.

Another axis is the conduction type of the voice sound. Typically, the voice sound we use is air-conducted sound, which is small and fast vibration of the air. Meanwhile, the small and fast vibration of the air in the vocal-tract vibrates the vocal-tract wall. Some sound energy passed through tissues and bones of the neck. This bone-conducted voice can be picked up with a bone-conduction microphone. As is the bone-conduction, the body-conducted voice, that is the voice running through neck muscles, can be sensed with a special microphone invented by Nakajima [1-3]. He also found that weak murmured voice, which is usually unheard by people nearby, can be detected by the

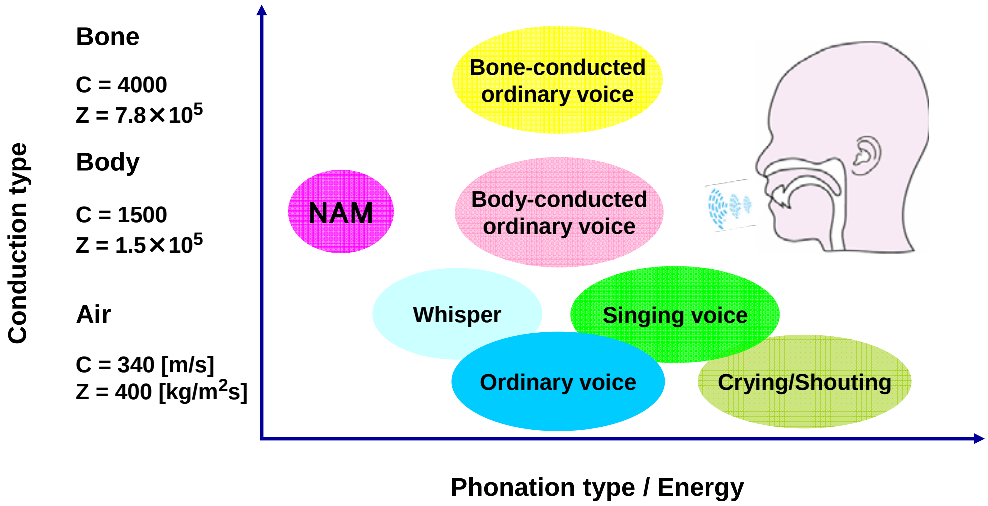

Fig. 1. Diverse types of voices.

special microphone attached to the surface of the skin close behind the ear. Namely, voice sounds propagate through not only the air but also the bone and the body. Such body-conducted weak murmured voice sounds are named non-audible murmur (NAM).

These *new* voices extend the horizon of the voice fields. The use of the body-conducted voice, in particular, expands the capacity of voice technologies. As the NAM sounds are unheard, the NAM can be applied for automatic speech recognition (ASR) system and telephone system that require privacy. As the air conducted noise sounds are not mixed with sounds picked up by the NAM microphone, the device can eliminate noise problems of the ASR system. The NAM microphone also will be able to revive the speech communication of those with vocal cord problems caused by larynx cancer, nerve disorders and muscle diseases.

The nature of NAM, however, is less well understood because it is only recent that the NAM has ever recorded. Not only the acoustic characteristics of the NAM signal but also the frequency response of the NAM microphone has yet to be revealed. This paper clarifies acoustic characteristics of the NAM signal by analyzing large amount of data, and also characterizes the frequency response of NAM microphones.

CALIBRATION OF NAM MICROPHONE

NAM microphones are calibrated using an audio analyzer (Brüel & Kjær: PULSE 3560C), an accelerometer (Ono Sokki: NR-3211), a bone conduction vibrator (Rion: BR-41) and an urethane-elastomer cylinder of 75 mm in diameter and 50 mm in height.

The vibrator is placed beneath the cylinder. A NAM microphone and an accelerometer are placed on the top of the cylinder. The accelerometer is covered with silicone cover so as to align the height to the NAM microphone. A weight of 1 kg is added to hold them down (Fig. 2).

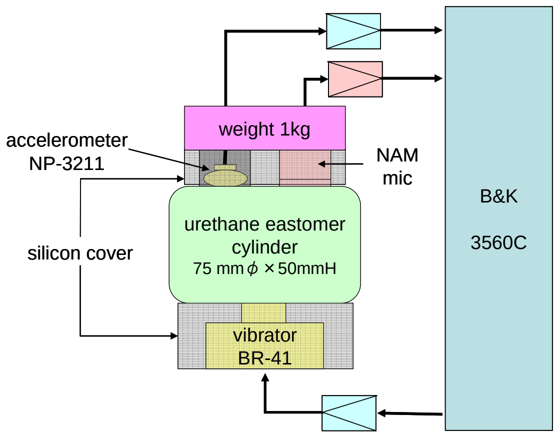

Fig.2. Microphone calibration system.

Figure 3 show the soft-silicon type NAM microphone made by Mitsumi Electric Corp. and the home-brew urethane-elastomer type NAM microphone. The signal sensed by the accelerometer was amplified by a sensor amplifier (Ono Sokki: SR-220), whose frequency response is flat. The signal sensed by a NAM microphone was amplified by a NAM microphone amplifier (Mitsumi: NAM-01A), whose frequency response is broad band-pass characteristics as shown in Figure 4.

Let measured frequency response of the accelerometer be A(ω), then $A(\omega)/\omega^2$ gives the frequency response for the displacement. The accelerometer has flat response for acceleration below 10 kHz. As NAM microphones are displacement sensors, comparing a measured frequency response of a NAM microphone D(ω) with $A(\omega)/\omega^2$ gives the calibrated frequency response of the NAM microphone. The frequency response of the NAM microphone amplifier was compensated.

Their frequency responses are shown in Figure 5. Both NAM microphones have similar response below 1.5 kHz, but they differ in higher frequency region. The response of the urethane-elastomer type NAM microphone is fairly flat up to 10 kHz, while that of the soft-silicon type NAM microphone gradually rolls off at a -17 dB/octave between 2 kHz and 9 kHz. Sensitivities of the two types of NAM microphone at 159.2 Hz are 15 dB higher than that of the accelerometer.

Fig. 3. A soft-silicon type (left) and a urethane-elastomer type (right) NAM microphones.

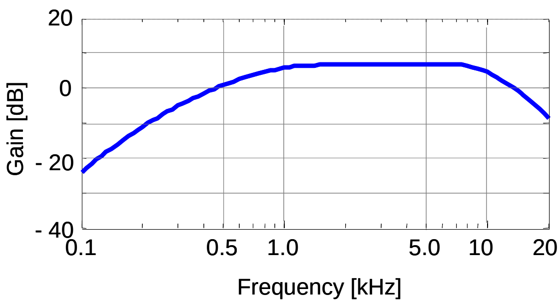

Fig. 4. Response of the NAM microphone amplifier.

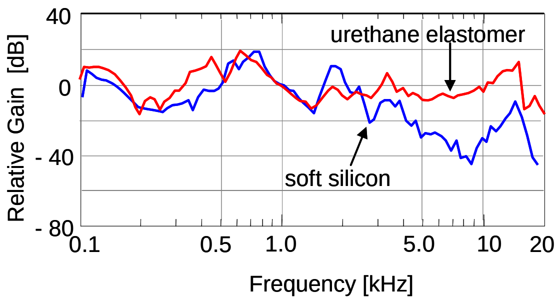

Fig. 5. Responses of the two types of NAM microphone.

NAM SIGNAL RECORDING

NAM signals were recorded in a sound proof room. 6 male and 7 female adult subjects read 50 ATR phoneme balanced Japanese sentences in NAM style. A soft-silicon type NAM microphone and the NAM microphone amplifier were used in the recording. The microphone was attached to the surface of the skin close behind the ear and fixed with a neck band. The setting position of the NAM microphone was chosen carefully so as to detect good NAM signal. The NAM signal were recorded by a solid state recorder (Marantz: PMD670) with sampling rate of 48 kHz and 16 bits quantization. During the recording, the NAM phonation style and signal level were monitored by an expert to keep the recording signal quality. When large noise was involved, talkers were asked to reread the sentence. In total, 80 minutes NAM signals of 640 sentences were recorded.

ACOUSTICAL CHARACTERISTICS OF NAM SIGNAL

Waveform and spectrogram of an NAM signal sensed by a soft-silicon type NAM microphone are shown in Figure 6. Speech period is identifiable on the signal waveform. But background noise level is high and a number of impulsive pop noises are overlapped with NAM signal, suggesting that the signal-to-noise ratio (SNR) of the NAM signal is low. The spectrogram also suggests that the SNR of the signal is low and that the bandwidth of the NAM signal is narrow.

The pop noises are generated either by a rustle of the NAM microphone and the skin or by articulatory organs' contacts when producing plosives. Detailed waveforms of these pop noises are depicted in Figure 7. Some of the noise can be reduced by a threshold processing, as the amplitude of the noise is usually larger than that of the NAM signals. After reducing some pop noises, speech periods of the NAM signals are detected using their power and delta cepstrum coefficients.

Mean signal-to-noise ratio (SNR) is 15 dB for the NAM signals of 48 kHz sampling rate. Lowering the sampling rate improves the SNR and it reaches 21 dB in maximum at sampling rate of 2 kHz. The SNR of the NAM signals, however, is much worse than that of ordinary speech signal sensed by a normal microphone.

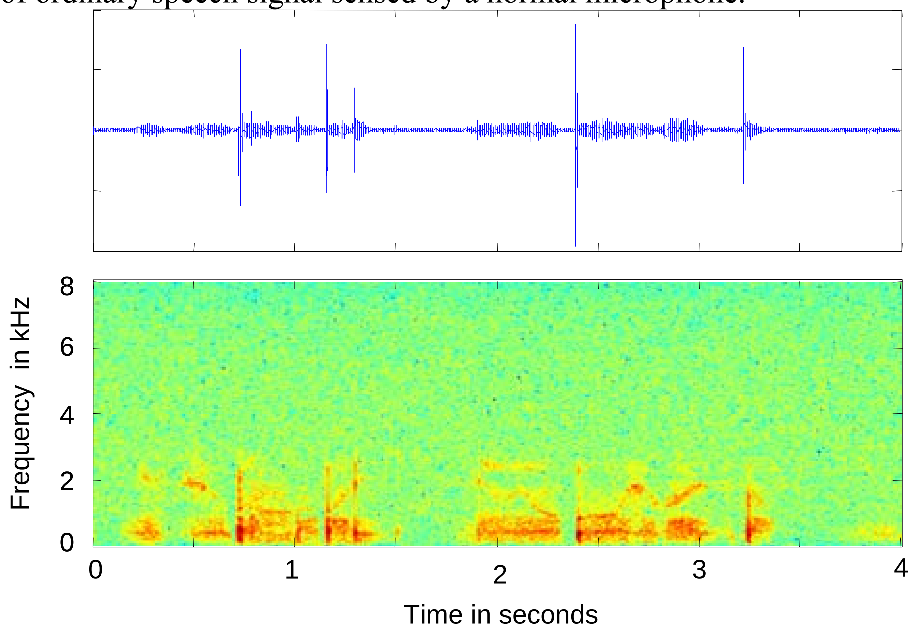

Fig. 6. Waveform and spectrogram of a NAM signal sensed by a soft-silicon type NAM microphone.

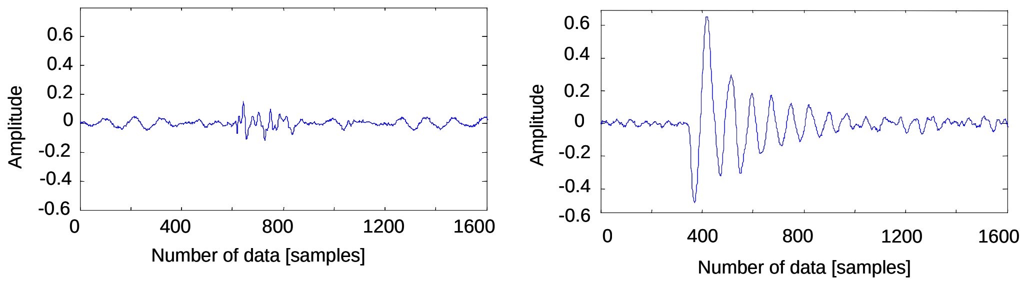

Fig.7. Waveform of the pop noise caused by an articulatory organs' contact (left) and by a rustle of the NAM microphone and the skin (right) with a magnified time axis.

Long-term averaged spectra (LTAS) of the NAM signals resampled at 16 kHz are shown in Figure 8: blue and red solid lines indicate mean LTAS of 6 male and 7 female NAM signals and dotted line shows the noise floor. LTAS show energy concentration around 500 to 800 Hz both for male and female NAM signal. Their spectrum roll-off above 800 Hz is at a -17 dB/octave and that below 500 Hz is at a 10 dB/octave. The LTAS also show the bandwidth of the NAM signal is 3 kHz. Noise floor level of the recording system and the NAM signal strength limit the bandwidth.

The frequency responses of the NAM microphone as well as the NAM microphone amplifier are involved in the NAM signals to calculate the LTAS. The LTAS of the raw NAM signal, thus, can be obtained by subtracting those responses from the LTAS. Figure 9 shows the LTAS of the raw NAM signal. The LTAS show energy concentration around 200 to 300 Hz. Their spectrum roll-off above 300 Hz is at a -23 dB/octave.

According to the source-filter theory of speech production, the LTAS of the ordinary voice has an approximately -6 dB/octave roll-off. As the radiation characteristics from the lips is +6 dB/octave, spectral tilt is estimated -12 dB/octave for the ordinary voice in the vocal tract. Although the spectral tilt of whispered voice is less sloped than that of ordinary voice, the spectral tilt of raw NAM signal is far less sloped than that of air-conducted whispered voice and ordinary voice.

Numerical simulation of transfer characteristics of sound in head and neck [4] suggests that body conducted sound has fewer higher-frequency components than an air- conducted sound and that the decay of the transfer characteristics is approximately -10 dB/octave in the audible range. Adding this decay to the -12 dB/octave is -22 dB/ octave, which is roughly parallel to the spectral characteristics of the raw NAM signal.

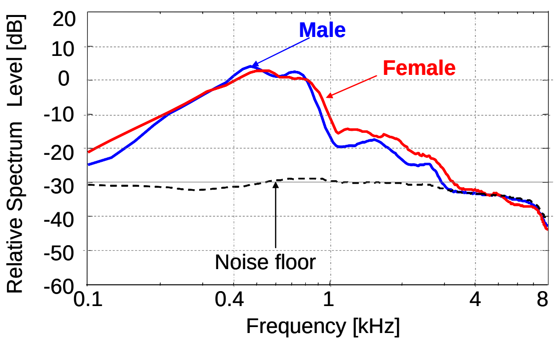

Fig.8. LTAS of the NAM signals.

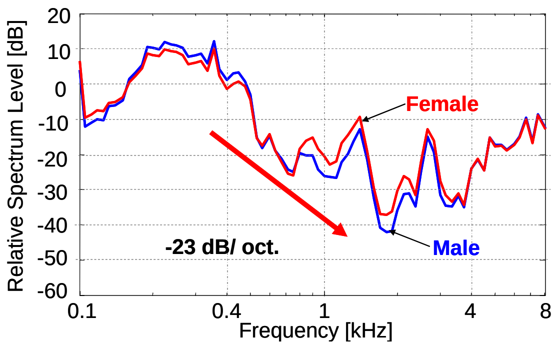

Fig.9. LTAS of the raw NAM signals.

DISCUSSION

Two types NAM microphones show quite different frequency responses. Existing soft-silicon type NAM microphone senses traditional telephone band signals. Lately developed urethane-elastomer type NAM microphone can sense broader signals. Urethane elastomer is an industrial material having both the hardness of plastic and the elasticity of rubber. As the material can be poured, it is easy to fill the material into the condenser microphone. In addition to that, urethane elastomer has impact absorbing qualities with glutinosity, and is resistant to oil and chemicals. An urethane-elastomer type NAM microphone can, thus, be attached to neck without using a neck band. As it sticks to the neck skin, rustle of the NAM microphone and the skin decreases, reducing the pop noise generation. Waveform and spectrogram of a NAM signal sensed by a urethane-elastomer type NAM microphone are shown in Figure 10. As a talker and a

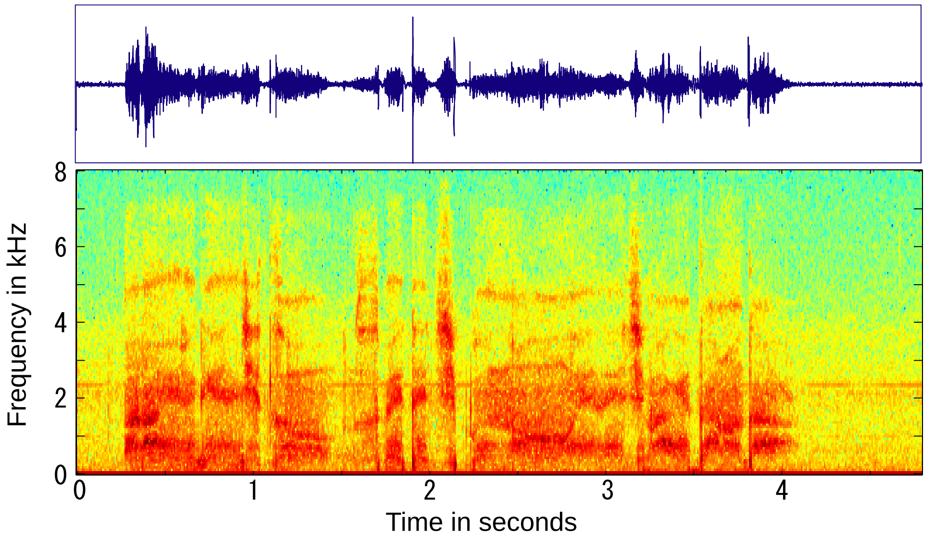

Fig. 10. Waveform and spectrogram of a NAM signal sensed by a urethane-elastomer type NAM microphone.

NAM microphone amplifier used are different, direct comparison is not fair. The NAM signal, however, apparently has better SNR and broader bandwidth. Fewer impulsive pop noises are observed in the waveform in Figure 10 than that in Figure 6. The urethane-elastomer type NAM microphone is promising.

CONCLUSIONS

Two types of NAM microphone were calibrated. A soft-silicon type NAM microphone only senses signals below 3 kHz. Meanwhile, a urethane-elastomer type NAM microphone can sense broader signals up to 10 kHz. Sensitivities of the two types of NAM microphone are 15 dB higher than that of the accelerometer.

Acoustic characteristics of NAM signals sensed by a soft-silicone NAM microphone were clarified by analyzing large amount of data. The mean SNR of the NAM signal is 15 dB and the bandwidth of that is 3 kHz. The LTAS of the NAM signal, including frequency responses of the NAM microphone as well as the NAM microphone amplifier, show energy concentration for frequencies between 500 to 800 Hz and a -17 dB/octave roll off. Meanwhile, after compensating the NAM microphone response, the LTAS of the raw NAM signal show energy concentration around 200 to 300 Hz and a -23 dB/octave roll off. This is roughly parallel to the numerical simulation of transfer characteristics of sound in head and neck.

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

REFERENCES

[1] Nakajima Y. *et al.*, "Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin, " Proc. ICASSP, **708-711**. (2003).

[2] Nakajima Y., "Development and Evaluation of Soft Silicone NAM Microphone," Technical Report of IEICE Speech, SP2005-7, **7-12**. (2005)

[3] Nakajima Y., Kashioka H., Campbell N. and Shikano K., "Non-Audible Murmur (NAM) Recognition," IEICE Trans. on information and systems**, Vol.E89-D, No.1, 1-8**. (2006)

[4] Otani M. and Hirahara T. and Adachi S., "Numerical simulation of sound originated from the vocal tract in soft neck tissues," *J.Acoust.Soc.Am.*, **Vol.120, No.5**, **3352**. (2006)