2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Intelligibility of bone-conducted speech detected on the scalp assessed by mono-syllable articulation and speech transmission index*

Satoshi Nanri, Taishi Shinobu, *Member, IEEE*, Sho Otsuka, and Seiji Nakagawa, *Member, IEEE*

Abstract— **Bone-conduction microphones (BCMs) can detect speaker's voices with high signal-to-noise ratio even under extremely noisy environments. However, it is sometimes accompanied by discomfort and esthetic problems because BCMs are ordinarily attached to the front of the neck (larynx). In order to solve such problems, we have been developing a novel BCM systems built in a hard hat [2]. To develop this BCM system, characteristic of bone-conducted speech detected on the scalp need to be clarified. In this study, intelligibilities of bone-conducted speech detected at several locations on the head and neck were assessed by mono-syllable articulation tests and the speech transmission index (STI), objective measure of signal transmission quality. The results obtained indicated that the forehead and the vertex showed better articulation and STI than the mastoid process of the temporal bone, the mandibular condyle and occiput. Additionally, the larynx, commonly used in existing BCM systems, showed lower scores than others. These results suggest that attenuation of high-frequency components are smaller at the forehead and the vertex, and indicate the practicability of these locations as the BCM placement.**

I. Introduction

Under extremely noisy environments like a machine factory or an engine room of a watercraft, it is difficult to make use of ordinary microphones which detect vibrations of the air. Such air-conduction microphones (ACMs) detect not only speech but also noise. On the other hand, bone-conduction microphones (BCMs) can detect only speaker's voice with high signal-to-noise ratio even under the extremely noisy environments. Some BCM systems are already being made into a product [1] (Fig. 1).

However, BCMs are sometimes accompanied by discomfort and esthetic problems because BCMs are ordinary attached to the fore part of the neck (larynx). In order to solve these problems, we have been developing a novel BCM systems built in a hard hat [2]. This BCM system detect speech on the scalp, therefore, characteristic of bone-conducted speech recorded on the scalp need to be clarified. However, few former studies have been carried out on bone-conducted speech detected on the scalp even in quiet conditions [3]. In this study, mono-syllable articulation of bone-conducted speech detected at several locations on the head and neck were measured in quiet conditions and hearing tendency was summarized. Also, the speech transmission index (STI), objective measure signal transmission quality, was calculated.

Fig. 1 An example of the bone-conduction microphone (Panasonic Corp.[1]). The microphone is placed at around the larynx.

All the experiments were approved by the Institutional Review Board on Life Science Research of Chiba University. Necessary information about the experiments was given to the participants and informed consent was obtained from each participant before the experiments.

II. Experiment I: Recordings of air-conducted and bone conducted speech

A. Participants

A native speaker of Japanese (Tokyo dialect) participated (male, 21 years).

B. Recordings of the air-conducted and bone-conducted speech

The participant was requested to utter 100 Japanese mono-syllables in an anechoic room with the intensity of a normal conversation. Utterances of the speaker were recorded through a microphone (Crown CM-311A) placed in front of participant's mouth and an acceleration pickup (Ono Sokki NP-3110) placed at six locations on the speaker's head (Fig. 2). The detected speech was A/D converted with a sampling frequency of 25.6 kHz and low-pass filtered at 10 kHz.

C. Location of speech detection

The microphone and the acceleration pickup was placed at following locations (Fig. 2):

a) The larynx

S. Nanri and T. Shinobu are graduate students at the Department of Medical Engineering, Graduate School of Science and Engineering, Chiba University, Chiba 263-8522, Japan.

S. Otsuka is an Assistant Professor at the Center for Frontier Medical Engineering, Chiba University, Chiba 263-8522, Japan.

S. Nakagawa is a Professor at the Center for Frontier Medical engineering, Chiba, Chiba University, Chiba 263-8522, Japan (phone: +81-43-290-3114; fax: +81-43-290-3116; e-mail: s-nakagawa@chiba-u.jp).

S. Otsuka and S. Nakagawa are concurrently serve for Graduate School of Engineering, Chiba University, and Med-Tech Link Center, Chiba University Hospital.

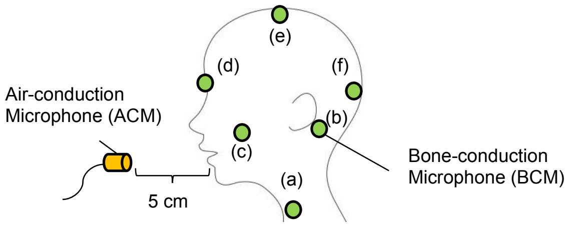

Fig. 2. The locations of the acceleration pickup: (a) the larynx, (b) the mastoid process of the temporal bone (c) the mandibular condyle, (d) the forehead, (e) the vertex, and (f) the occiput.

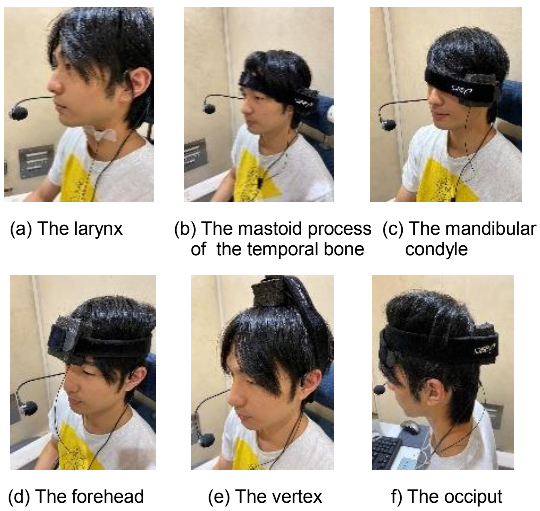

Fig. 3 Settings of the acceleration pickup in the experiment: (a) the larynx, (b) the mastoid process of the temporal bone (c) the mandibular condyle, (d) the forehead, (e) the vertex, and (f) the occiput.

b) The mastoid process of the temporal bone

c) The mandibular condyle

d) The forehead (Fpz in the international 10-20 system used in EEG measurement; the same hereinafter)

e) The vertex (Cz)

f) The occiput (Oz)

Settings of the acceleration pickup are shown in Fig. 3. The acceleration pickup was held at the six locations on the speaker's head using an elastic band made of polystyrene, and at the larynx using double-sided and medical surgical tapes.

III. Experiment II: Mono-syllable Articulation Tests

A. Participants

8 adults (5 males and 3 females, 21-24 years) who has no history of deficits of hearing functions participated.

B. Stimulus and Procedure

100 Japanese mono-syllables recorded in the experiment I were used as stimuli. The stimuli were presented binaurally every three seconds using a headphone (Sennheiser HD650). Participants were requested to answer the presented stimulus was as they heard. The intensities of stimuli were adjusted to the most clearly perceiving level.

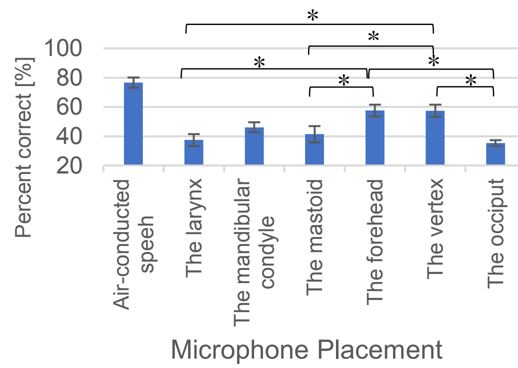

Fig. 4 Percent corrects of the mono-syllable articulation tests.

C. Result (1): Percent correct

Fig. 4 shows the mean of the percent correct across all participants. In the experiment, scores of all bone-conducted speech were inferior to that of air-conducted speech, but the forehead and the vertex shows practical scores close to 60%. The scores of the larynx and the occiput are lower than others.

An analysis of variance (ANOVA) showed that the significant effect of the detection location ($p < 0.05$). The post-hoc test (Bonferroni test) showed that correct rate of the air-conducted speech was significantly higher than all of bone-conduction speech ($p < 0.01$). The forehead and the vertex showed higher scores ($p < 0.05$) and the larynx showed lower scores than other locations on the head.

D. Result (2): Confusion matrix

The results of the articulation tests show how well the subjects heard the BCU speech. The results, however, provide no information as to what kind of errors occurred. Confusion matrices have been used to investigate perceptual confusion by comparing presented syllables and subject's answers, according to vowels and consonants. We composed confusion matrices for each stimulus type using averaged scores of mono-syllable articulation across the participants (Fig. 5). Fig. 5 shows the confusion matrices for air-conducted and bone-conducted speech. The phonemes were classified into vowels, unvoiced consonants, voiced plosive and fricative consonants, and other voiced consonants. These matrices show similar patterns of confusion generally, however, some differences between detection locations were observed. For example, confusion among vowels and missing of consonants seems to occur more frequently at the larynx.

And correct answer rate was calculated by confusion matrices. In the category (3), correct answer rate of the forehead was significantly better than the larynx, the mandibular condyle and the occiput ($p < 0.05$), and significantly better tendency than the mastoid process of the temporal bone ($P < 0.10$).

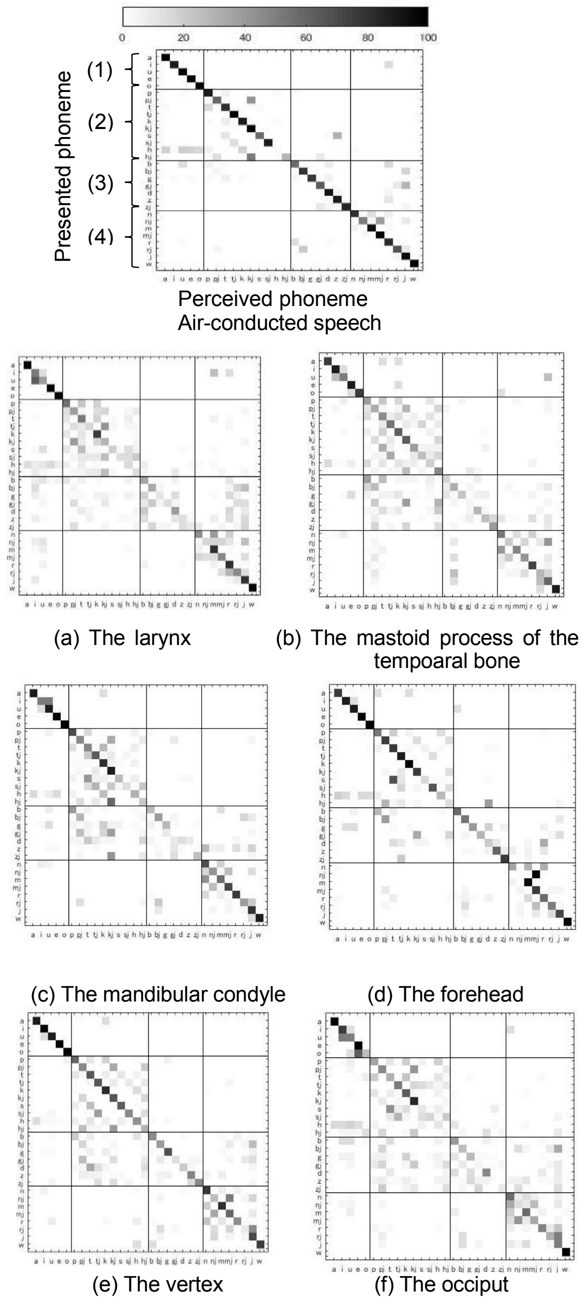

Fig. 5 Confusion matrices

IV. Speech Transmission Index (STI)

A. Overview

STI, an objective measure of signal transmission quality, is calculated from modulation transfer functions (MTFs) which indicate modulation of the envelope of a transmitted signal from an original undistorted signal. Values of STI varies between 0 and 1. Generally, the larger STI value is, the higher the intelligibility is.

B. Estimation procedure

In this experiment, air-conducted and bone-conducted speech were used as the original and transmitted signals, respectively. Fig. 6 shows a procedure of calculation of the MTFs. Each parameter is described as follows

Apparent masking strength: $I_{AMi,j}$ were calculated each octave-band i.

$$I_{AMi,j} = \bar{I}_{B(i-1),j} * AMF \tag{1}$$

Here, AMF, meaning "the auditory masking attenuation factor", is 0.0003.

The auditory correction factor: $ACF_{i,j}$ were calculated each octave band i.

$$ACF_{i,j} = \bar{I}_{i,j} / (\bar{I}_{i,j} + I_{AMi,j}) \tag{2}$$

Corrected $m_{i,j}$: $m'_{i,j}$ were calculated.

$$m'_{i,j} = ACF_{i,j} * m_{i,j} \qquad (0 \leq m'_{i,j} \leq 1) \tag{3}$$

Apparent S/N ratio were calculated by $m'_{i,j}$.

$$SNR_{i,j} = 10\log_{10}\{m'_{i,j} / 1 - m'_{i,j}\} \tag{4}$$

The transmission index: $TI_{i,j}$ were calculated by normalizing $SNR_{i,j}$.(Range of equivalent S/N ratio to consider: R=30, Shift factor: S=-15)

$$TI_{i,j} = (SNR_{i,j} - S)/R \qquad (0 \leq T_{i,j} \leq 1) \tag{5}$$

The average value of $TI_{i,j}$:MTI_i ware calculated.

$$MTI_i = \frac{1}{14}\sum_{j=1}^{14} TI_{i,j} \tag{6}$$

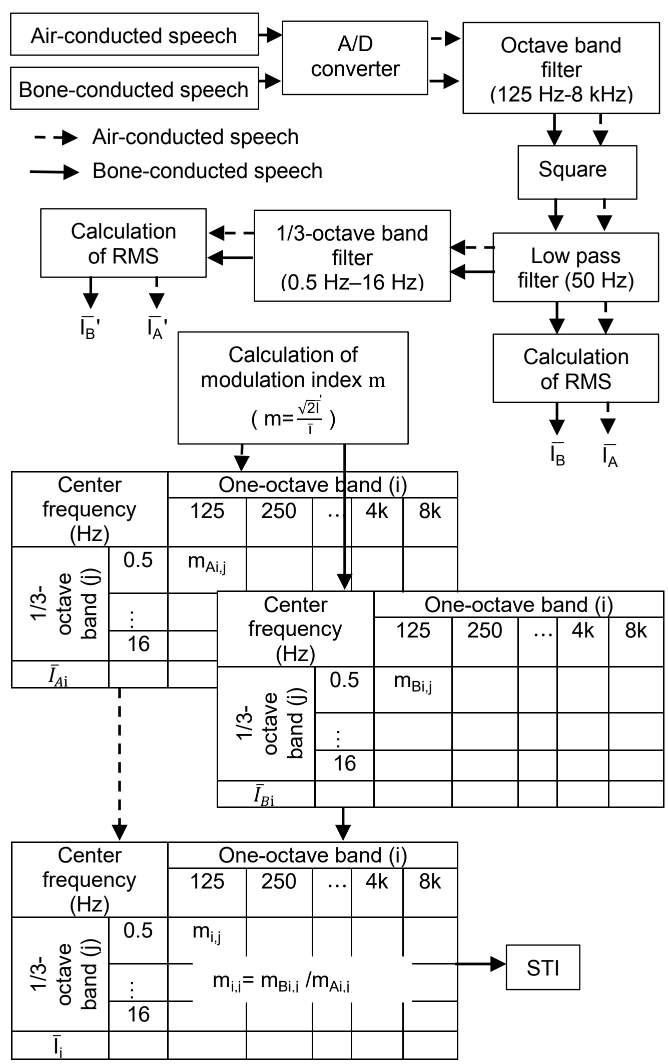

Fig 6 Calculation procedure of STI.

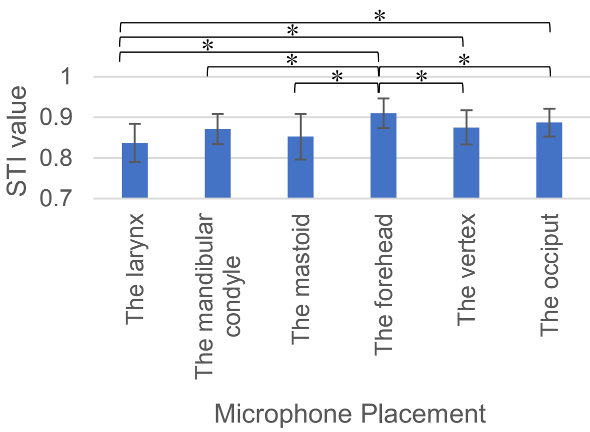

Fig. 7 Values of STI

STI was calculated using the weighting coefficients α and β, which are the contributions of each octave band to the intelligibility.

$$\mathrm{STI}=\sum_{i=1}^{7}\alpha_i{*}\mathrm{MTI}_i+\sum_{i=1}^{6}\beta_i{*}\sqrt{\mathrm{MTI}_i{*}\mathrm{MTI}_{i+1}} \qquad (7)$$

(α_1=0.085,α_2=0.127,α_3=0.230,α_4=0.233,α_5=0.309, α_6=0.224,α_7=0.173,β_1=0.085,β_2=0.078,β_3=0.065, β_4=0.011,β_5=0.047,β_6=0.095)

C. *Result*

Fig. 7 shows mean of STI values. Kruskal-Wallis test showed a significant effect of the detection location ($p < 0.05$). The post-hoc test (Mann-Whitney U test corrected by Bonferroni) showed that the STI value of the forehead was better than all others ($p < 0.01$). and that of the vertex and the occiput were significantly better than the larynx ($p < 0.05$).

V. Discussion

A. *Mono-syllable articulation tests*

Percent corrects of bone-conducted speech detected on the forehead and the vertex reached about 60%. Since it is almost same as the intelligibility of AM radio, the results indicate the practicabilities of these locations as BCM placement. On the other hand, intelligibility of the larynx, commonly used in existing BCM systems, was much lower than others.

These results indicated intelligibilities of bone-conduction speeches may be depended on distance from the lips that generate high frequency components. However, despite distance from the lips to vertex is long, intelligibility of the vertex was high. Since previous study has indicated that a major pathway in bone conduction stimulation to the inner ear is via CSF [7], it is a possible that the propagation path of bone-conducted speech to the vertex is CSF.

And, intelligibility of vowel part of the larynx was significant lower than almost all of the others. Furthermore, the consonant part tended not to be perceived by confusion matrices. This is due to the larynx is close to the vocal cord which generate sound source and it is easy to detect the voice before it is tuned in the oral cavity, the lips and so on.

The high intelligibility of the forehead was consistent with a previous study [3]. On the other hand, low intelligibility was observed for the vertex because it was covered by hair and difficult in making good contact condition between BCM and the scalp [3] On the other hand, in this study, the vibrator was firmly secured using an elastic band and it seems that good coupling had been kept during measurements.

B. *Speech transmission index (STI)*

Values of STI are almost consistent with mono-syllable articulation tests. However, value of STI of the occipt was high despite percent correct was low. There is a possibility that this cause was the air-conducted speech measured at same time when the bone-conducted speech of the occiput was measured had deteriorated.

VI. Conclusion

In this study, intelligibilities of bone-conducted speech detected at several locations on the head and neck were assessed by mono-syllable articulation tests and the speech transmission index (STI).

The results obtained indicated that the forehead and the vertex showed better articulation and STI than the mastoid process of the temporal bone, the mandibular condyle and occiput. Additionally, the larynx, commonly used in existing BCM systems, showed lower scores than others. These results suggest that attenuation of high-frequency components are smaller at the forehead and the vertex, and indicate the practicability of these locations as the BCM placement. Utterances of only one speaker was used in this study. Further studies using utterances obtained from multiple speakers are needed to obtain more general results.

References

[1] Panasonic Corp., “Bone conduction headset”, <https://panasonic.biz/cns/invc/bone_conduction/>

[2] T. Shinobu et al., “Effects of vibrator placement and ambient noise on perception of bone-conducted sound during earplugging”, Proceedings of Inter-Noise 2020.

[3] P. Tran et al., ”Bone conduction microphone: Head sensitivity mapping for speech intelligibility and sound quality”, Proceedings of ICALIP, pp.107-111, 2008.

[4] H. Steeneken et al., “The temporal envelope spectrum of speech and its significance in room acoustics”, Proceedings of the 11th International Congress on Acoustics, vol. 7, pp.85-88, 1983.

[5] T. Houtgast et al., “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria”, The Journal of the Acoustical Society of America, vol. 77, pp.1069-1077, 1985.

[6] H. Steeneken et al., “Basic of the STI-measuring method”, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.501.8775

[7] H. Sohmer et al., “Bone conduction experiments in humans – a fluid pathway from bone to ear”, Hearing Research, vol. 146, pp. 81-88, 2000.