1Pb5-7

Intelligibility of bone-conducted speech detected on the scalp

Satoshi Nanri[1†], Sho Otsuka[1,2,3], and Seiji Nakagawa[1,2,3‡] ([1]Dept. of Medical Eng., Graduate School of Sci. & Eng., Chiba Univ.; [2]Ctr. for Frontier Medical Eng., Chiba Univ.; [3]Med Tech Link Center, Chiba Univ. Hospital)

1. Introduction

An air-conduction microphones (ACM) detect sound as a vibration of air. However, it is difficult to use an ACM in extremely noisy environments such as a machine factory or engine room of a watercraft, because it detects both ambient noises and the voices of speakers. On the other hand, a bone-conduction microphone (BCM) can detect only speaker's voices with a high signal-to-noise ratio (SNR), even in extremely noisy environments. Recently, BCM systems for speech communication in noisy environments becomes commercially available [1] (Fig. 1).

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

However, BCMs are sometimes associated with discomfort and esthetic problems because existing BCMs are ordinary attached to the fore part of the neck (the larynx). To solve these problems, we develop a novel BCM system built in a hard hat [2]. However, since transmission characteristics of bone-conducted (BC) speech detected on the scalp are not clarified.

In this study, to evaluate the intelligibilities of BC speeches detected on the scalp, mono-syllable articulation of BC speech detected on the several locations of the head and the neck were conducted. Furthermore, Speech Transmission Index (STI), an objective measure of signal transmission quality, were calculated for BC speech.

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

2. Exp. I: Recording of air-conducted and bone-conducted speech

2.1 Method

Four native speakers of Japanese (21-24 years, Tokyo dialect) participated in these recordings. Their utterance of 100 Japanese mono-syllables were recorded in an anechoic room. A microphone (CM-3114A, Crown) was used to capture air-conducted (AC) speeches. Also, BC speeches were captured at same time by accelerometers (352A24, PCB). Fig.2 shows the locations of the microphone and the accelerometers. The accelerometer was fixed at six locations: (a) the larynx (Lar), (b) the mandibular condyle (Che), (c) the mastoid process of the temporal bone (Mas), (d) the forehead (Fpz), (e) the vertex (Cz), and (f) the occiput (Oz), using surgical tape and an elastic band made of polystyrene. And, to improve the coupling, a piece of wax was placed between the accelerometer and surface of the speaker's head and the neck.

3. Exp. II Mono-syllable articulation tests

To assess the intelligibilities of BC speech recorded in Exp. I, mono-syllable articulation tests were conducted.

Seven native listeners of Japanese (21-23 years) participated. Mono-syllables, recorded as AC and BC speech in Exp. I, were presented binaurally every three seconds using the headphone (HD660S, Sennheiser). The intensity of the stimuli was set to the most clealy perceiving level for each test.

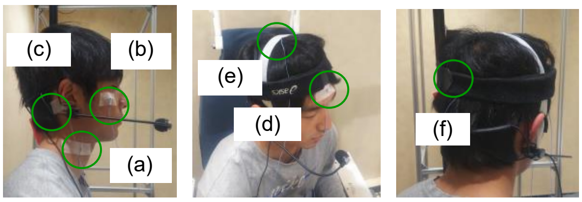

Fig. 2 Locations of the microphone for AC and the accelerometers for BC: (a) the larynx (Lar), (b) the mandibular condyle (Che), (c) the mastoid process of the temporal bone (Mas), (d) the forehead (Fpz), (e) the vertex (Cz), and (f) the occiput (Oz), using surgical tape and an elastic band made of polystyrene.

†nanri.s@chiba-u.jp, ‡s-nakagawa@chiba-u.jp

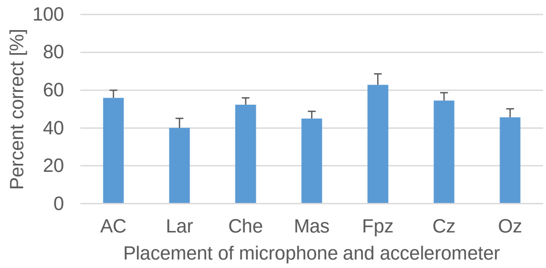

Fig.3 Percent corrects of the mono-syllable articulation tests.

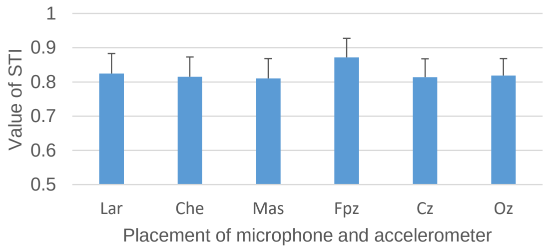

Fig.4 The values of STI for each placement.

4. Exp. III: Speech Transmission Index (STI)

To objectively assess the intelligibilities of bone-conducted speeches, the speech transmission index (STI), an objective measure of signal transmission quality, was calculated.

The STI is an objective measure based on the contribution of a few frequency bands within the frequency range of speech signal [3-5]. The values of the STI vary between 0 and 1. The larger the STI value, the higher the intelligibility of speech.

5. Results

Fig.3 shows the percent corrects for a ACM and BCMs in Exp. II. The correct rates of BC speeches of the forehead and the vertex were significantly higher than that of the larynx, the mastoid and the occiput ($p < 0.01$). Especially, score of the forehead tended to be higher than that of AC, reached about 63%. Also, The mandibular condyle showed higher score than that of the larynx ($p < 0.01$) and the mastoid ($p < 0.1$).

Fig.4 shows the mean of the STI values for each location in Exp. III. As well as the intelligibility score, the STI value of the forehead was the highest among all locations ($p < 0.001$).

6. Discussion and Conclusion

Generally, the percent correct of BC speech at each location was lower than that of AC speech. However, the intelligibility of the forehead reached about 63%, almost equivalent to that of telephones in the early 2000s [6]. Also, the forehead showed the highest STI value among all locations. The high intelligibility of the forehead observed in the experiment corresponds to a result of a former study [7]. Also, the mandibular condyle and the vertex showed relativity high intelligibility. On the other hand, the intelligibility of the larynx, which is used in existing BCM system, showed the lowest score.

Although it is not easy to elucidate the mechanisms of BC speech propagation inside the head, the results obtained suggest that the intelligibility of BC speech is affected by distance from the lips, which generates high-frequency components includes in consonants. Additionally, anatomies of the head also seem to affect the BC speech transmission. The head is composed of many bones and has cavities such as the frontal sinus, nasal cavity and paranasal sinuses. For example, the frontal sinus locates just under the forehead and connected to the oral cavity via a nasal cavity. Therefore, it can be considered that speech components were effectively transmitted to the forehead, or resonances of such cavities enhance the enhance the speech components.

6. Conclusion

In this study, we evaluated the intelligibility of our BCM system at five different on the scalp. The results showed that the scalp locations, especially the forehead, mandibular condyle, and the vertex are more suitable for detecting voices than the larynx, used for existing BCM systems. These location can be available in the novel BCM system built in a hard hat.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP19K22950 and JP20H04497, and a Research Grant for the Next Generation, Research Incubator from the Inst. for Global Prominent Research, Chiba University for SN.

References

1. Panasonic Crop., “Bone-conduction headset”, https://panasonic.biz/cns/invc/bone_conduction/, (2020).
2. T. Shinobu et al., Proceedings of Inter-Noise 2020, **5** (2022) pp.2651-2647.
3. H.Steeneken et al., Proceedings of the 11th ICA, **7** (1983) pp.85-88.
4. T. Houtgast et al, J. Acoust. Soc. Am, **77** (1985) pp.1069-1077.
5. H. Steeneken et al, “Basic of the STI-measuring method”, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.501.8775, (2002).
6. R. Jeff, Polycom Inc., https://www.paperzz.com/doc/6369680/ (2003).
7. P. Tran et al., ICALIP2008, (2008) pp.107-111.