

Characteristics of Frequency Resolution and Speech Perception Using Bone Conduction on Different Human Facial Parts*

Ko Uemura, *Student Member, IEEE*, Sho Otsuka, and Seiji Nakagawa, *Member, IEEE*

Abstract— In conventional bone-conduction (BC) devices, a vibrator is typically attached to the mastoid process of the temporal bone or the condyle process of the mandible. However, BC-sound presentations to facial parts such as the nose and cheek have also been investigated recently. As the face is among the most complex structures of the human body, transmission of sounds using BC on different facial parts are likely to show different perception and propagation characteristics than those presented to conventional parts. However, the characteristics of BC sound presented to different part of the face have not yet been studied in detail. To test the frequency discrimination ability, we measured difference limens for frequency (DLFs). We also conducted monosyllable articulation tests in Japanese to assess the speech-perception characteristics when BC sounds are presented to various facial (nasal, infraorbital region, zygomatic, jaw angle, and chin) and conventional (mastoid and condyle process) parts of a normal-hearing subject. The results suggest that, at least in the parts investigated in the current study, the frequency resolution and intelligibility of the facial parts were about the same as those of the conventional parts. These results indicate that practical frequency information and speech perception are possible with BC devices attached to different facial parts.

I. INTRODUCTION

Normally, sound reaches the auditory organs through air. This process is known as air conduction (AC). However, sound can also be perceived as the vibration of bio tissues through a process called bone conduction (BC). As BC sounds have a component that directly acts on the inner ear [1], they have been applied to hearing aids for people with conductive hearing loss. In addition, because BC devices do not cover the ear canal, BC sounds can be easily heard even when earplugs are worn. Therefore, BC has been applied to voice-communication devices in noisy environments.

In BC hearing aids, the vibrator is attached to the osseous parts of the temporal bone such as the mastoid process. In addition, voice-communication devices, such as BC earphones, are often attached to the condyle process of the mandible [2]. This requires the vibrator to be pressed against the skin. Therefore, BC is often accompanied by pain and aesthetic problems. In addition, there are difficulties associated with keeping the vibrator steadily on the round surface of the mastoid process. To address this issue, a new

method for attaching vibrators has been explored, and basic studies have been conducted on BC sounds presented to any part of the human body, such as the neck, trunk, and upper limb [3,4].

On the other hand, we previously suggested that BC-sound presentation on the facial parts may also be a potential solution to the wearability problems of BC devices [5,6]. However, details of the perception mechanisms of BC sounds when presented to the face remain unclear. The mastoid and condyle processes, which are used in conventional BC devices, are closer to the auditory organ, such as the ear canal or cochlea. Therefore, BC sounds can be efficiently transmitted [2,7]. However, the face has a very complicated structure, and slight changes in the presentation part will likely also change BC-sounds perception and propagation characteristics.

We have studied the basic perceptual (the hearing threshold) and propagation characteristics (the ear-canal sound pressures and head vibrations) of BC sounds presented to facial parts and have discussed the differences among presentation parts [5,6]. Here, we measured difference limens for frequency (DLFs), which reflect the frequency resolution of BC-sound perception, to examine the frequency discrimination ability of BC sounds presented to the face. In addition, we conducted monosyllable articulation tests in Japanese to assess the speech-perception characteristics. For a more detailed evaluation, the results of the monosyllable articulation test were used to calculate the confusion matrices (the probability that a presented phoneme is perceived as a different phoneme) of the component phonemes of a monosyllable.

In the current study, BC sounds were presented to several parts of the face (hereafter referred to as “facial parts”), such as the nasal (nasal bone), infraorbital region (maxillary bone), zygomatic (zygomatic bone), jaw angle (mandible), and chin (mandible). We compared them with those of parts conventionally used with BC devices (the mastoid process of the temporal bone, and the condyle process of the mandible; hereafter referred to as “conventional parts”) and AC sounds.

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

*This work was supported by JSPS KAKENHI Grant Numbers JP19K22950 and JP20H04497 for SN.

K. Uemura is a graduate student at the Dept. of Medical Eng., Chiba Univ., Chiba 263-8522, Japan (e-mail: k_uemura1112@chiba-u.jp).

S. Otsuka is an Assistant Professor at the Center for Frontier Medical Eng., Chiba Univ., Chiba 263-8522, Japan (Otsuka.s@chiba-u.jp).

S. Nakagawa is a Professor at the Center for Frontier Medical Eng., Chiba Univ., Chiba 263-8522, Japan (phone/fax: +81-43-290-3263; e-mail: s-nakagawa@chiba-u.jp).

S. Otsuka and S. Nakagawa concurrently serve for Dept. of Medical Engineering, Chiba Univ., and Med-Tech Link Center, Chiba Univ. Hospital.

the participants, and informed consent was obtained from each participant before the experiments.

II. METHOD

A. Experimental setup

BC sounds were presented to the following body parts (Fig. 1): the mastoid process (temporal bone), condyle process (mandible), nasal (nasal bone, the root of the nose between the two eyes), infraorbital region (maxillary bone, below the orbital region), zygomatic (zygomatic bone, bony prominence protruding anteriorly on the cheek), jaw angle (mandible, horns on the posterior border of the mandible), and chin (mandible, ridge running longitudinally against the midline of the mandible). The BC sounds were presented to the left side of the head for the mastoid process, condyle process, nasal, infraorbital region, zygomatic, and jaw angle. All experiments were conducted in an anechoic room. The subjects were seated in easy chairs and wore silicone earplugs (Insta-Moid products, INSTA-PUTTY) in both ears to observe the perceptual characteristics of only the BC components. Moreover, they were instructed to rest and not make large facial movements such as opening the mouth wide, raising the corners of the mouth, and clenching the teeth.

The BC sounds were presented using a BC vibrator (Radioear, B-81). The vibrator was fixed using a sponge and an elastic band (Asics, Kinesis KJ1). In addition, a jig was embedded in the sponge to fix the pressure sensor and measured pressure (Fig. 2). The presentation pressure was set at 2.5 ± 0.3 N for the nasal and infraorbital regions and 3.0 ± 0.3 N for the other parts. Before performing the experiments, the presentation pressure was measured using a load cell (Kyowa Electronic Instruments, LMB-A-20N) and an instrumentation amplifier (Kyowa Electronic Instruments, WGI-400A). Fig. 3 shows the frequency response for the B-81 in this study, both with and without the fixture at different presentation pressures for a constant input voltage of $1\text{-}V_{\text{RMS}}$. For the measurement of AC sounds, the same sounds as the BC were presented to both ears using headphones (Sennheiser, HD 660 S). Each sound was generated by a PC using MATLAB (MathWorks) at a sampling frequency of 192 kHz and fed to the vibrator via a 16-bit digital-to-analog converter (ECHO, Audiofire12).

B. Measurement of DLFs

Five male subjects with normal hearing participated in the study. They ranged in age from 21 to 24 years. As stimuli, tone bursts with a duration of 400 ms each, including rising/falling ramps of 30 ms, were presented at seven center frequencies (f_c : 125, 250, 500, 1000, 2000, 4000, and 8000 Hz).

Before measuring DLFs at each f_c , the hearing thresholds of BC and AC sounds were measured for 125–8000 Hz tone bursts (duration of 500 ms each, including rising/falling ramps of 50 ms) using a one-up one-down adaptive yes/no recognition procedure. During this procedure, participants were requested to respond "yes" if they heard the presented sounds and "no" if they did not. We started the measurement at a sufficiently perceptible stimulus intensity (the initial sounds intensity) and changed the intensity adaptively according to subjects' responses. The measurements were taken until 12 reversals of intensity change were obtained, and the mean of the last 8 reversals was used as the thresholds.

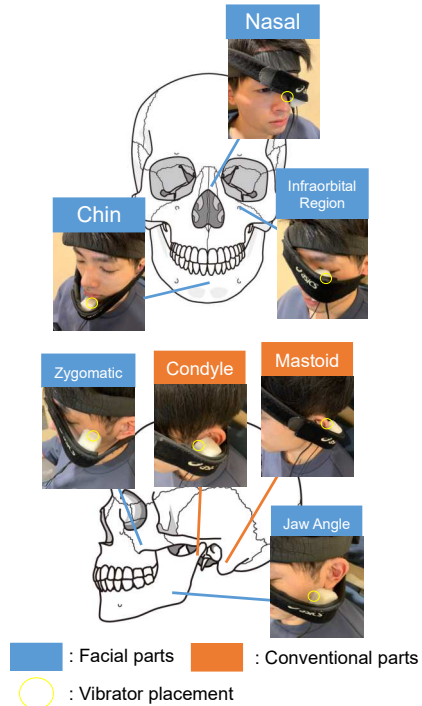


Figure 1. Stimulus locations of BC sounds in the experiment.

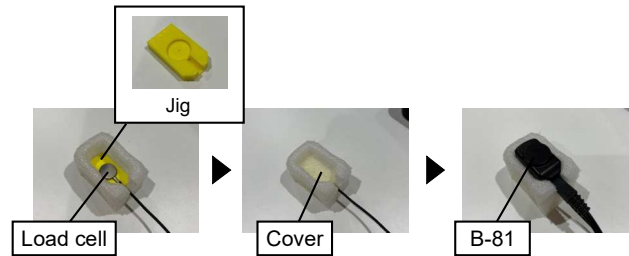


Figure 2. A Jig for fixing the load cell and BC vibrator used in all experiments.

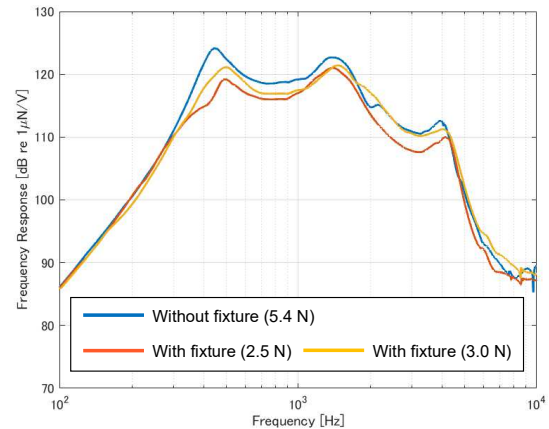


Figure 3. Frequency response for the B-81 with and without the fixture and at different presentation pressure.

The DLFs were measured using a one-up two-down two-alternative forced-choice adaptive procedure. The stimulus output levels were set to 30 dB SL for each part and AC sounds. In each trial, two tone bursts that were equally spaced in linear frequency on either side of the f_c were presented, and the participants were requested to select the stimulus of higher

frequency. The deviation from the f_c (Δf) at which the participants could sufficiently discriminate between two stimuli, and Δf varied adaptively. The measurements were conducted until 11 reversals of the Δf change were obtained, and the geometric mean of the $\Delta f/f_c$ of the last eight reversals was used as the discrimination values.

C. Monosyllable articulation tests

Five male subjects and one female subject with normal hearing participated in this test. They ranged in age from 21 to 24 years.

Before performing monosyllable articulation tests at each part, the hearing thresholds of the BC and AC sounds were measured for 1000 Hz pure tones (equal energy per unit time to speech sounds) using a one-up one-down adaptive yes/no recognition procedure.

A commercially available database was used to take 100 Japanese monosyllables recorded in a female voice (NTT-AT FW03). Before the tests, sound output levels were set to 40 dB SL for each part and AC. In each test, speech sounds were presented randomly. After each speech sound, participants were asked to type in the sounds they heard using a keyboard.

III. RESULT

A. DLFs

Fig. 4 shows the DLFs for each stimulated part and AC sounds. The DLFs were significantly increased up to 125 Hz and to 8000 Hz, regardless of the stimulus type or stimulated parts. Moreover, there were no significant differences among the conventional part, facial parts, and AC sounds at each frequency. A two-way analysis of variance (ANOVA) with the stimulated parts and frequencies as factors revealed that the main effect of the frequencies was significant ($p < 0.01$), but not the main effect of stimulated parts or interaction effects.

B. Monosyllable articulation tests

Fig. 5 shows the percent corrects of the articulation test for each stimulated part and AC sounds. The intelligibility of each part tended to be lower than those of the AC sounds. In the BC sounds, relatively high intelligibility was obtained in the conventional parts (the mastoid and condyle processes), zygomatic and jaw angle. A one-way ANOVA with the stimulated parts as a factor revealed that the main effect was significant ($p < 0.001$). A multiple comparison of the stimulated parts revealed that the percent corrects for stimulated parts except for the mastoid process and zygomatic were significantly lower than for the AC sounds (condyle process: $p < 0.01$, nasal: $p < 0.001$, infraorbital region: $p < 0.001$, jaw angle: $p < 0.01$, chin < 0.001). In contrast, there were no significant differences among the mastoid process, zygomatic, and AC sounds, or among each stimulated part of BC.

Fig. 6 shows the confusion matrices for each stimulated part and the AC sounds obtained from the articulation tests. In the facial parts, the confusion of voiced consonants increased, particularly in the nasal, infraorbital region, and chin. A one-way ANOVA with the stimulated parts as a factor was conducted for each phoneme, and a main effect was

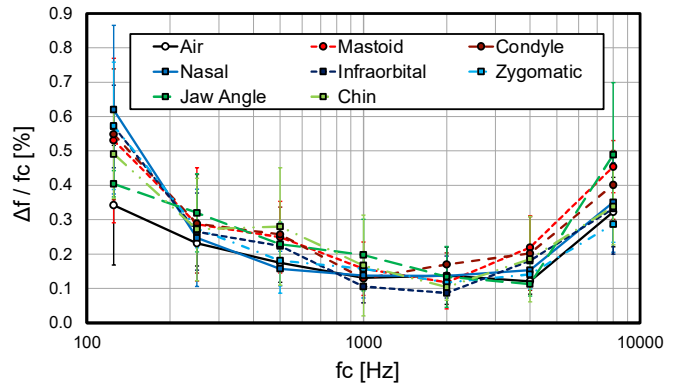


Figure 4. DLFs for each part.

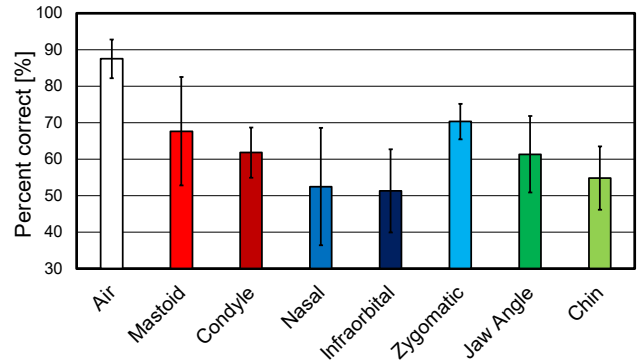


Figure 5. Percent correct for each part.

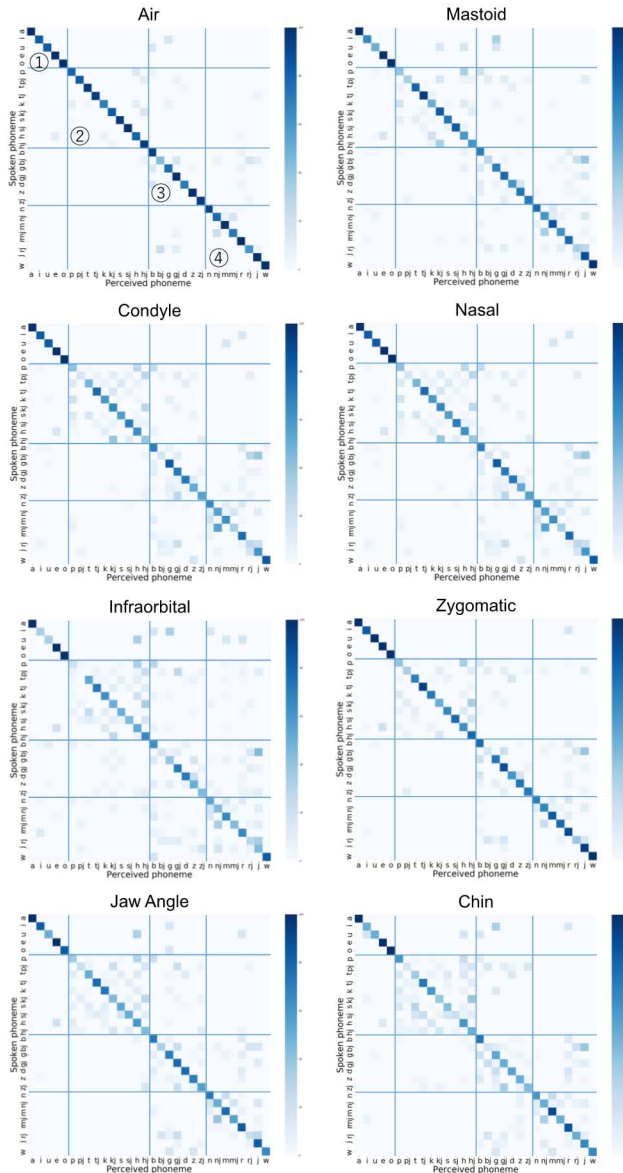
found for /s/, /g/, and /j/ (/s/: $p < 0.05$, /g/: $p < 0.01$, /j/: $p < 0.05$). A multiple comparison revealed that the confusion of /s/ of the chin ($p < 0.01$) increased compared to the mastoid process, and the confusion of /g/ of the infraorbital region ($p < 0.01$) increased compared to the condyle process. On the other hand, other phonemes showed no differences among each part.

IV. DISCUSSION

A. DLFs

It was confirmed that the DLFs for each facial part are approximately the same as for the conventional parts and AC sounds. These results of the DLFs reveal that the frequency resolution of BC sounds presented to the face is comparable to that of AC sounds and the resolution of BC sounds presented to conventional parts. Moreover, the current results of the DLFs are also similar to those of the DLFs of BC sounds presented more distant parts, such as the neck [3]. Although the hearing thresholds increase with the increase in distance of the stimulated parts from the head, there is no difference in DLF. On the other hand, differences in propagation distance and stimulated parts cause changes in sensitivity but do not affect the transmission of frequency information.

Additionally, the earplugs used in current study were reported to insulate 20 dB (125–1000 Hz) and 40 dB (4000 Hz) [8], and the possibility that the AC component was perceived cannot be ruled out in the DLFs measurement. More rigorous insulating methods and stimulus intensity settings may be necessary to study the frequency discrimination ability of exclusively BC components.



① Vowel ② Unvoiced ③ Voiced ④ Other
Figure 6. Confusion matrices for each part.

B. Monosyllable articulation tests

In the zygomatic and jaw angle, the intelligibility was about the same as that of the conventional parts. In particular, intelligibility of the zygomatic surpassed that of the conventional parts and that was not significantly lower than those of the AC sounds. The results suggest that this part is capable of practical speech perception, especially in the facial parts. In contrast, the percent correct for the nasal, infraorbital region, and chin were slightly lower than those for the conventional parts, zygomatic, and jaw angle. For these three parts, the distance to the auditory organs is somewhat increased compared to the zygomatic and jaw angle. It is possible that the high-frequency components were attenuated during the propagation, resulting in decreased the intelligibility. Furthermore, it has also been reported that the intelligibility of the BC sounds decreases with the distance of propagation from the head and the intelligibility in the neck is comparable to that of the conventional parts [4]. In combination with the intelligibility of the facial parts, it can

be said that speech perception is likely to be affected by the distance of propagation and the associated distance attenuation. In addition, it is thought that speech perception of the presentation parts such as face, and neck is comparable or equivalent to that of the conventional parts.

In the facial parts, the confusion was found to be similar to that of the conventional parts, with the exception of some phonemes. However, the confusion of voiced consonants, including fricatives and affricates, was slightly increased in the nasal, infraorbital region, and chin. It is possible that relatively large distortion, including attenuation of high-frequency components, occurred during the propagation inside the head, and that the distortion may have affected the discrimination of voiced consonants with energy concentrated in the high-frequency ranges above 1 kHz. In addition, it is necessary to investigate the cause of the phonemes that have been identified as having confusion and study more detailed speech-perception characteristics based on the details of the propagation mechanism in the facial parts.

V. CONCLUSION

We measured DLFs to clarify the frequency discrimination ability presented to the face. In addition, to assess the speech-perception characteristics, we conducted monosyllable articulation tests in Japanese. The results suggest that, at least in the parts investigated in current study, the frequency resolution and intelligibility of the facial parts were about the same as or similar to those of the conventional parts. In particular, the zygomatic was capable of practical speech perception equivalent to that of AC sound. These results indicate that practical frequency information and speech perception are possible with BC devices presented to facial parts. This study also suggested that the distance attenuation due to increased propagation distance on the face does not affect frequency resolution but does affect speech perception to a small extent. Thus, distance attenuation should be taken into account when developing BC devices for facial parts.

REFERENCES

- [1] S. Stenfelt *et al.*, "Factors contributing to bone conduction: The outer ear," *J. Acoust. Soc. Am.*, 113 (2), pp.902- 13, 2003.
- [2] X. Qin *et al.*, "Estimation on the influence of placement on bone conduction transmission by ear-canal sound pressure," *Acoust. Sci. Tech.*, 41(1), pp.384-385, 2020.
- [3] H. Ishikawa *et al.*, "Threshold and frequency- and temporal resolutions of distantly presented bone-conducted sound in the audible-frequency range," *Jpn. J. Appl. Phys.* Vol. 61, SG1065, 2022.
- [4] H. Ishikawa *et al.*, "Characteristics of speech transmission of distantly-presented bone-conducted sounds in the audible-frequency range," presented at the 2021 Neuroscience 2021 (the 50th Ann. Meet. Soc. for Neurosci.), Chicago, IL.
- [5] K. Uemura *et al.*, "Perception and Propagation Characteristics of Bone-Conducted Stimuli Presented to the Human Facial Cranium," *Proc. of the 44th Intl. Conf. in the IEEE EMBS*, pp. 4163-4166, 2022.
- [6] A. Miwa *et al.*, "Occlusion effects by bone-conducted sound to the facial parts assessed by hearing threshold and ear-canal sound pressure measurements," *Proc. of the 43rd Symp. Ultrasonic Electronics*, 2022, 3Pa5-2.
- [7] P. Henry and T. Letowski, "Bone conduction: Anatomy, physiology, and communication," *Tech. Rep. ARL-TR-4138*, 2007.
- [8] X. Qin *et al.*, "Discussion on the Influence from Materials of Earplugs in Bone Conduction by Hearing Threshold and ECSP Measurements," *Ext. Abst. (44th MidWinter Meet., 2021); Association for Research in Otolaryngol.*, p. 114.