Evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of speaker gender and age information

Takayuki Kagomiya[1,2], *Seiji Nakangawa*[2]

[1]Center for Research Resources, National Institute for Japanese Language nad Linguistics, Japan
[2]Health Research Institute,
National Institute of Advanced Industrial Science and Technology (AIST), Japan
t-kagomiya@ninjal.ac.jp, s-nakagawa@aist.go.jp

Abstract

Human listeners can perceive speech signals in a voice-modulated ultrasonic carrier from a bone-conduction stimulator, even if the listeners are patients with sensorineural hearing loss. Considering this fact, we have been developing a bone-conducted ultrasonic hearing aid (BCUHA). The purpose of this study was to assess the usefulness of the BCUHA in transmission of speakers' physical attributes: gender and age. The evaluation used gender and age-identification experiments. The experiments were also conducted under air-conduction (AC) and cochlear implant simulator (CIsim) conditions. The results showed that: the BCUHA can well transmit speakers' gender information; the BCUHA can transmit speaker age information better than CIsim.

Index Terms: ultrasound, bone-conduction, hearing aid, paralinguistic information, speakers' attribute.

1. Introduction

We have developed a bone-conducted ultrasonic hearing aid (BCUHA) for sensorineural hearing-impaired patients [1]. A BCUHA consists of two components: an amplitude-modulated ultrasound processor and a bone-conduction vibrator.

Ultrasound is defined as sound with a frequency higher than the limitation of human perception (about 15 kHz). However, humans can perceive sound transmitted as ultrasound through a bone-conduction vibrator [bone-conducted ultrasound (BCU)] [2]. Moreover, if the ultrasound is amplitude modulated by speech sounds, the original speech sounds can be perceived in addition to the carrier sound by both normal-hearing (NH) and hearing-impaired listeners. We based the development of our BCUHA on these observations.

The cochlear implant (CI) was developed for and widely adopted by sensorineural hearing-impaired patients. A CI also consists of two components: speech signal processors and electrodes mounted in the cochlea. However, despite their widespread use, some problems with CIs have been reported. The biggest problem is that a CI requires surgical positioning, which causes irreversible damage to the cochlea. Another problem is that, because performance is limited by the number of electrodes, the CI transmits only partial or reduced information, not the entire sound. Nowadays, the number of CI electrodes is limited to between 12 and 24, and frequency resolution is limited according to the number of electrodes. This number may be sufficient for transmitting linguistic messages; however, it is reported that CI users have great difficulty perceiving music, speaker identity, emotional state, etc. [3, 4, 5, 6].

In contrast, the BCUHA does not require surgical fitting; users simply attach the bone-conduction vibrator with a hair band-like device (Figure 1). Furthermore, the BCUHA does not have the frequency limitations of CIs. However, in addition to speech signals, BCUHA listeners perceive high-frequency sound owing to the carrier signal [1]. Therefore, the carrier-originated sound may prevent clear perception of speech sounds. Consequently, the speech signal transmission performance of the BCUHA has been assessed using various techniques, such as using monosyllables [7] or word intelligibility scores [1]. These studies found that syllable articulation scores using BCU were over 60% [7] and that word intelligibility scores for high-familiarity words were over 85% [1]. The patterns of confusion in speech perception using BCU have many points in common with air conduction (AC) [7]. However, speech sounds convey not only linguistic messages but also indexical information about the speaker, such as gender, age, identity, and emotional state. As mentioned above, it is difficult for CI users to perceive such messages; thus, if the BCUHA performs better in this regard, it has a great advantage over the CI.

In this study, we focused on the ability to transmit speaker information. Previously, we compared the BCUHA with CI simulator (CIsim) based on speaker discrimination [8]. The results of this previous study showed that the speaker discrimination performance of the BCUHA is as good as that of the CI. However, this evaluation was performed using a discrimination task; the participants were simply requested to judge if the speaker

10.21437/SpeechProsody.2014-80

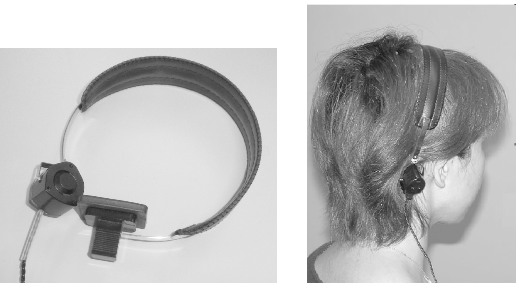


Figure 1: *Ceramic vibrator of the BCUHA attached to the mastoid with a hair band-like device*

of two sounds was "the same" or "different," and the result did not illustrate what types of speaker attributes were transmitted. Therefore, we assessed the ability of the BCUHA to transmit a speaker's physical attributes, including gender and age by conducting a series of listening experiments.

2. Methods

2.1. Stimuli

The experiments were designed as gender- and age-identification tasks, and stimuli were selected using the following procedures.

2.1.1. Speech material

To develop speaker gender- and age-identification tasks, we extracted speech material spoken by a large number of gender- and age-balanced speakers from "The Corpus of Spontaneous Japanese" (CSJ) [9]. From this corpus, we selected a phrase pronounced by various speakers: a short passage reading task "DNA." From the "DNA" task, the first phrase "di:-enu-e:" (DNA) was selected as the speech material. This phrase consists of voiced sounds and contains a nasal consonant. These types of sounds contribute to speaker identification [10]; thus, this phrase is expected to be suitable for the speaker attribute identification task.

Another reason for selecting this phrase was that it was spoken by 229 speakers in the CSJ. The speakers' genders were balanced, and the speakers' ages ranged from the low 20s to the high 60s. The numbers of speakers categorized by gender and age are listed in Table 1, in which "20L" represents low 20s, "20H" represents high 20s, etc.

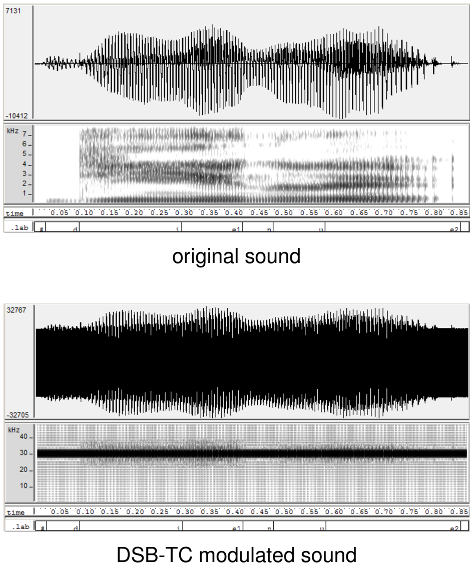
Table 1: *Number of speakers categorized by gender and age*

	20L	20H	30L	30H	40L	40H	50L	50H	60L	60H	total
M	12	11	12	12	10	13	13	10	10	11	114
F	8	18	12	9	14	6	13	10	17	8	115
total	20	29	24	21	24	19	26	20	27	19	229

2.2. Experiments

2.2.1. Bone-conducted ultrasound

The stimuli were converted to BCU stimuli in the form of amplitude-modulated 30 kHz sinusoid waves. A double-sideband transmitted-carrier (DSB-TC) amplitude-modulation method (Figure 2) was applied for this study because previous studies revealed it to be the best amplitude-modulation method for the BCUHA [1, 7]. With the DSB-TC method, the modulated speech signals $U(t)$ are represented by the following expression:



original sound

DSB-TC modulated sound

Figure 2: *DSB-TC amplitude-modulation*

$$U(t) = [S(t) - S_{\min}] \times \sin(2\pi f_c t) \qquad (1)$$

where $S(t)$ is the speech signal, $S_{\min}$ is the minimum amplitude of $S(t)$, and f_c is the carrier frequency (30 kHz).

The BCU stimuli were presented using a custom-made ceramic vibrator (Figure 1). Bone-conducted ultrasound can be perceived when it is applied to various parts of the body, and the mastoids are among the locations where such perception is high. Therefore, we applied the vibrator to the subject's left or right mastoid using a hair band-like device (Figure 1).

2.2.2. Cochlear implant simulator

To generate CI-simulated sounds, the Cochlear Implant Simulation (http://www.ugr.es/˜atv/web_ci_SIM/en/ci_sim_en.htm) developed by the University of Granada was adopted in this study, and the software was config-

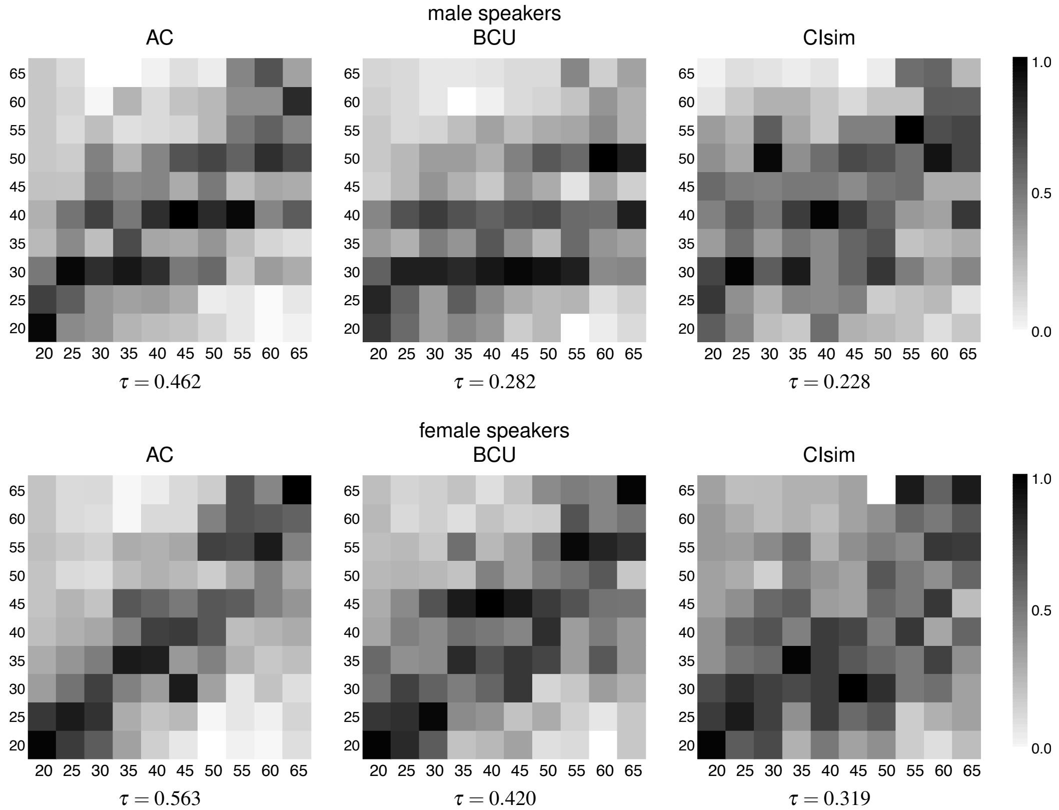


Figure 3: *Confusion ratios and Kendall's rank correlation coefficients (τ) obtained from the age-identification task*

Table 2: *Configuration of CI simulator*

length of implant	26.4 mm
number of channels	12
n-of-m	12 (CIS strategy)
interaction	2.4 mm
pulse rate	1515 pps/ch (18180 pps)

Table 3: *Confusion ratios from the gender-identification task*

	AC		BCU		CIsim	
	M	F	M	F	M	F
M	0.996	0.004	0.981	0.019	0.969	0.031
F	0.005	0.995	0.011	0.989	0.141	0.859

ured to simulate the MEDEL COMBI 40+ and TEMPO+ systems (see Table 2).

2.2.3. Participants

Seven native Japanese speakers (age: 19-41 years) with no reported hearing or speech defects participated in the experiments.

2.2.4. Procedures

All experiments were conducted in a soundproof chamber, and the sound levels of the stimuli were adjusted to the most comfortable level for each participant. For the AC and CIsim conditions, the sound stimuli were presented in a counterbalanced order through a set of headphones (Sennheiser HDA200). The participants were then requested to identify the speaker's gender (male or female) and age (10 levels listed in Table 1).

3. Results and Analysis

3.1. Speaker gender information

Table 3 lists the results of the gender identification task. The responses of all participants were pooled. The rows represent stimuli, and the columns show the responses. As shown in Table 3, speaker gender information was well transmitted in each condition. In the AC condition, the percentage of correct perceived ratios was higher than 99.5%, and in the BCU condition, it was higher than 98%. However, a small number of male-to-female errors (14%) were observed in CIsim condition.

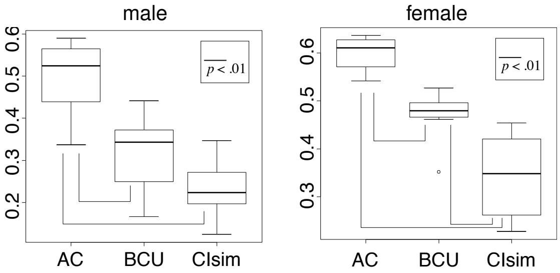


Figure 4: *Kendall's rank correlation coefficients by subjects*

3.2. Speaker age information

Figure 3 shows the confusion ratio and Kendall's rank correlation coefficients for speaker age. The responses of all participants were pooled. The columns represent stimuli, and the rows show the response. "20" represents low 20s, and "25" indicates the high 20s. Black-to-white contrast indicates the response ratio; darker cells indicate higher responses. The upper part shows the results of the male speaker condition, and the lower shows the female speaker condition.

From Figure 3 and the correlation coefficients, a large number of errors were observed. The rank correlation coefficients (τ) were lower than 0.57 (female voice) and 0.43 (male voice) even in the AC condition. In the BCUHA condition, the correlation coefficients were lower than 0.42 (female) and 0.29 (male) and 0.32 (female) and 0.23 (male) in the CIsim condition.

For each gender, the correlation coefficients were the best in the AC condition, then the BCUHA condition, and worst in the CIsim condition. To validate whether this tendency is statistically significant, correlation coefficients were calculated for each participant, and a series of ANOVA and post-hoc tests (multiple comparison with Holm's p adjustment method) was performed. The analyses were conducted for each speaker gender, and the results are shown in Figure 4. In the male speaker voice experiments, significant differences were observed between AC and BCU and between AC and CIsim ($p < 0.05$) and, in the female voice experiments, between all conditions ($p < 0.05$).

4. General Discussion

In each condition, speaker gender information was well transmitted. This result indicated that the BCUHA users can discriminate speaker gender as well as normal-hearing listeners can. This was also consistent with the results of the previous speaker discrimination tests [8], which showed that speaker errors were rarely observed in the AC, BCUHA, and CIsim conditions [8].

These high-accuracy gender identification results can be accounted for by good transmission of F0 information. It is well known that the F0 for an adult male voice is low and that for an adult female is medium. Moreover, BCUHA listeners can perceive the Japanese pitch accent [11] or prosodically salient paralinguistic information [12]. The results of these studies show that the BCUHA can transmit both local F0 modulations and global F0 range. Transmission of local F0 modulations allows pitch accents to be perceived, and global F0 range conveyance enables paralinguistic information, such as the speaker ' s gender, to be obtained.

In contrast, BCUHA listeners have difficulty identifying the speaker's age. This result indicates that some voice timbre information is lost during BCU listening. However, this was also observed in the CIsim condition, and the correlation ratios between stimuli and perceived age by BCUHA listeners were superior to those of the CIsim listeners. Thus, the results of this study indicate that the BCUHA performs at least as well as the CI for speaker information transmission and outperforms the CI in some aspects.

5. Summary and Conclusions

The ability of the BCUHA to transmit speaker information was evaluated by conducting a series of speaker gender- and age-identification experiments. The results showed that the ability of BCUHA to make speaker gender judgments reaches the level of AC, whereas the BCUHA outperformed the CIsim for transmitting age information; however, the correlation coefficients between speaker age and perceived age using BCUHA were not sufficient. Further investigations are required to solve this problem.

6. Acknowledgements

This research was supported by a Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (24500675, 25280063) and the Funding Program for Next-Generation World-Leading Researchers provided by the Cabinet Office, Government of Japan.

7. References

[1] S. Nakagawa, Y. Okamoto, and Y. Fujisaka, "Development of a bone-conducted ultrasonic hearing aid for the profoundly sensorineural deaf," *Transactions of the Japanese Society for Medical and Biological Engineering : BME*, vol. 44, no. 1, pp. 184–189, 2006.

[2] M. L. Lenhardt, R. Skellett, P. Wang, and A. M. Clarke, "Human ultrasonic speech perception," *Science*, vol. 253, pp. 82–85, 1991.

[3] Q.-J. Fu, S. Chinchilla, and J. J. Galvin, "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," *Journal of the Association for Research in Otolaryngology*, vol. 5, no. 3, pp. 253–260, 2004.

[4] X. Luo, Q.-J. Fu, and J. J. Galvin, "Vocal emotion recognition with cochlear implants," in *Proceedings of Interspeech 2006*, 2006, pp. 1830–1833.

[5] J. Gonzalez and J. C. Oliver, "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," *Journal of Acousitcal Society of America*, vol. 118, pp. 461–470, 2005.

[6] R. Müler, M. Ziese, and D. Rostalski, "Development of a speaker discrimination test for cochlear implant users based on the oldenburg logatome corpus," *Journal of Oto-Rhino-Laryngology, Head and Neck Surgery*, vol. 71, no. 1, pp. 14–20, 2009.

[7] Y. Okamoto, S. Nakagawa, K. Fujimoto, and M. Tonoike, "Intelligibility of bone-conducted ultrasonic speech," *Hearing Research*, vol. 208, pp. 107–113, 2005.

[8] T. Kagomiya and S. Nakagawa, "Evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of speaker discrimination information," in *Proceedings of Interspeech 2011*, 2011, pp. 2209–2212.

[9] K. Maekawa, "Corpus of Spontaneous Japanese: Its design and evaluation," in *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition*, Tokyo, 2003, pp. 7–12.

[10] K. Amino and T. Osanai, "Speaker identification using japanese monosyllables and contributions of nasal consonants and vowels to identification accuracy," *Japanese Journal of Forensic Science and Technology*, vol. 18, no. 1, pp. 13–21, 2013.

[11] T. Kagomiya and S. Nakagawa, "Perception of Japanese prosodical phonemes through use of a bone-conducted ultrasonic hearing-aid," in *Proceedings of Speech Prosody 2012*, vol. 1, 2012, pp. 35–38.

[12] T. Kagomiya and S. Nakagawa, "An evaluation of bone-conducted ultrasonic hearing aid regarding perception of paralinguistic information," in *Proceedings of Speech Prosody 2010*, 2010, pp. 100 867:1–4.