An Evaluation of Bone-conducted Ultrasonic Hearing Aid regarding Perception of Paralinguistic Information

Takayuki KAGOMIYA[1], *Seiji NAKAGAWA*[1]

[1]Institute for Human Science and Biomedical Engineering,
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{t-kagomiya, 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). This paper reports an evaluation of the BCUHA; the evaluation was carried out by considering the transmission of paralinguistic information, especially information on speakers' intentions. Intention identification experiments were carried out using the BCUHA. Multidimensional scaling (MDS) analysis was carried out, and the results indicated that the BCUHA shows good performance in transmitting information on the speaker's intention. Thus, this research shows that the BCUHA has the capability of being an effective tool in expressive oral communications.

Index Terms: ultrasound, bone-conduction, hearing aid, paralinguistic information, multi-dimensional scaling

1. Introduction

For patients with acute sensorineural hearing loss who are not able to hear using a normal hearing aid, we have been developing a bone-conducted ultrasonic hearing aid (BCUHA) [1].

Ultrasound waves are defined as sound waves that travel at such a high frequency that they cannot be heard by humans. However, ultrasound generated by a bone-conducted stimulator (bone-conducted ultrasound, BCU) is perceived by human listeners [2]. In addition, if BCU signals are amplitude-modulated by speech signals, listeners can perceive the original speech signals [2]. These voice-modulated BCU signals enable patients with acute sensorineural hearing loss perceive speech signals [1]. The BCUHA being developed is based on these observations.

The performance of the BCUHA has been evaluated in the past by considering syllable articulation [3] and word intelligibility [1]. These studies found that syllable articulation scores when using BCU were over 60% [3] and that word intelligibility scores for words with high familiarity were over 85% [1]. The patterns of confusion in speech perception in the case of BCU have many points of similarities with those for air conduction (AC) [3].

Although the usability of the BCUHA has been evaluated as mentioned above, the evaluations have been restricted to the transmission of linguistic information, or textural messages. In other words, little attention has been paid to the transmission of paralinguistic or nonlinguistic information.

The purpose of this study is to evaluate the usability of the BCUHA by considering paralinguistic information trans-mission, especially the transmission of speakers' intention or attitude. In daily oral communications, listeners perceive speakers' intention besides linguistic or textual information. For example, the English word "Really" can represent various intentions or attitudes like "I don't believe it" and "I'm surprised to know that." Paralinguistic information can be expressed by changing the pronunciation [4], even if the textual information is the same. Moreover, paralinguistic information enriches oral communications and makes it more expressive than written language. Thus, it is important to evaluate the performance of the BCUHA by considering the transmission of paralinguistic information. In this study, we conduct a series of listening experiments that involve paralinguistic information identification.

2. Method

The participants were presented short passages spoken with various intentions, and they were asked to identify the intention of the speaker.

To examine whether differences were observed in the identification of intentions between AC and BCU conditions, the same tasks were conducted under both conditions.

2.1. Sound Recording

For providing sound stimuli, a series of short passages were recorded.

2.1.1. Speaker

The speaker was a native Japanese woman in her thirties with no reported hearing or speech defects.

2.1.2. Intention Types

The specified intentions comprise *admiration* (A), *suspicion* (S), *disappointment* (D), *indifference* (I), *neutral* (N), and *focused* (F). Intentions "A," "S," "D," and "I" are listed as representative "emotions" in a textbook for Japanese learners, which focuses on effective oral communications [5]. Further, these intentions or attitudes have been adopted in some studies on paralinguistic information transmission [6, 7, 4].

2.1.3. Sentences

Three sentences were selected; the criterion used to select them was that they should convey the intentions described above. The selected sentences were the following: "So-desu-ka" (Is that so?/That is so), "Anata-desu-ka" (Is that you?/That is you), "Hontou-desu-ka" (Is that true?/That is true).

original sound

DSB-TC modulated sound
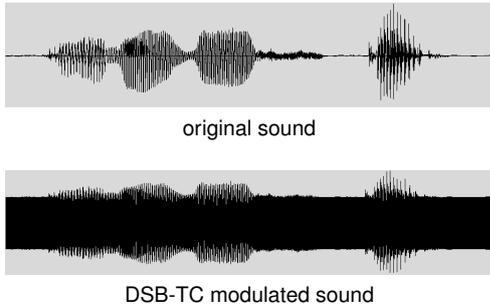
Figure 1: DSB-TC modulation of sound

Figure 2: Ceramic vibrator of the BCUHA attached to the mastoid with a hair-band-like device

2.1.4. Instruction to the Speaker

The intentions and speaking styles were explained to the speaker by rephrasing the intended paralinguistic message as a textual message, for example, "A" was explained as "That's great. I love it" and "S" was described as "I doubt it and I don't believe it." These textual messages were identical to those used in previous studies [6, 7, 4], and full texts can be accessed from the appendix of [6].

2.1.5. Apparatus

The speaker pronounced the series of sentences into a microphone in an anechoic room. The speech samples were recorded at 16 kHz/16 bit resolution and stored in a personal computer.

2.2. Experiments

2.2.1. Participants

Eight native Japanese speakers with no reported hearing or speaking defects participated in the experiments. Their ages were in the range 19-32 years.

2.2.2. Sound Presentation

The sound stimuli mentioned above were presented through a headphone (Sennheiser HD650) under AC conditions.

The stimuli presented under BCU conditions were 30 kHz ultrasounds that were amplitude-modulated by speech signals. The amplitude modulation method applied in this study was the double sideband-transmitted carrier (DSB-TC) method since previous studies had found this method to be capable of speech modulation for BCU [1, 3] . With the DSB-TC method, the modulated speech signals $U(t)$ are given by the following expression:

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

Table 1: Confusion matrix for each intention

AC	A	D	F	I	N	S
A	0.917	0.000	0.000	0.028	0.000	0.000
D	0.000	0.986	0.000	0.049	0.000	0.007
F	0.069	0.000	0.361	0.035	0.125	0.007
I	0.000	0.000	0.062	0.722	0.160	0.007
N	0.014	0.000	0.576	0.146	0.715	0.007
S	0.000	0.014	0.000	0.021	0.000	0.972
Correct perception rate: 0.779						
BCU	A	D	F	I	N	S
A	0.854	0.007	0.000	0.160	0.007	0.000
D	0.000	0.903	0.028	0.056	0.014	0.014
F	0.125	0.000	0.243	0.160	0.083	0.007
I	0.000	0.042	0.160	0.389	0.326	0.014
N	0.014	0.028	0.562	0.194	0.535	0.035
S	0.007	0.021	0.007	0.042	0.035	0.931
Correct perception rate: 0.642						

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, Figure 1).

The sampling frequency of the original signal and the bandwidth of the DSB-TC modulated signals were 16 kHz. Since the carrier frequency was 30 kHz, all frequency components in the DSB-TC modulated signals were over 20 kHz.

The stimuli under BCU conditions were presented using a custom-made ceramic vibrator (Figure 2). Bone-conducted ultrasound can be perceived when it is applied to various parts of our body, and the mastoids are among the locations where the perception is high. Therefore, we applied the vibrator to the left or right mastoid of the subject using a hair-band-like supporter (Figure 2).

2.2.3. Procedures

Each stimulus was presented at random, and participants were asked to identify the intention underlying the stimulus. Each stimulus was presented six times. All listeners who participated in the BCU experiments also took part in the AC experiment, which was conducted a few days later.

2.2.4. Apparatus

Both the presentation of the stimuli and the recording of the responses were executed using a personal computer. Further, the stimuli were provided using a FireWire-based audio interface (Echo Audiofire 12) attached to the personal computer.

The sound levels of the stimuli were adjusted to the most comfortable levels for each participant.

The experiments carried out under both AC and BCU conditions were conducted in a soundproof chamber.

3. Analysis

3.1. Broad View

The response rates were calculated for each intention and a confusion matrix was created. Table 1 shows the confusion matrix; the presented stimuli are the column headings and the participants' choices are the row headings. According to Table 1, intentions "A", "D," and "S" have a higher correct perception rate (more than 0.850) under both AC and BCU conditions, i.e., these intentions were correctly perceived by the listeners under both conditions.
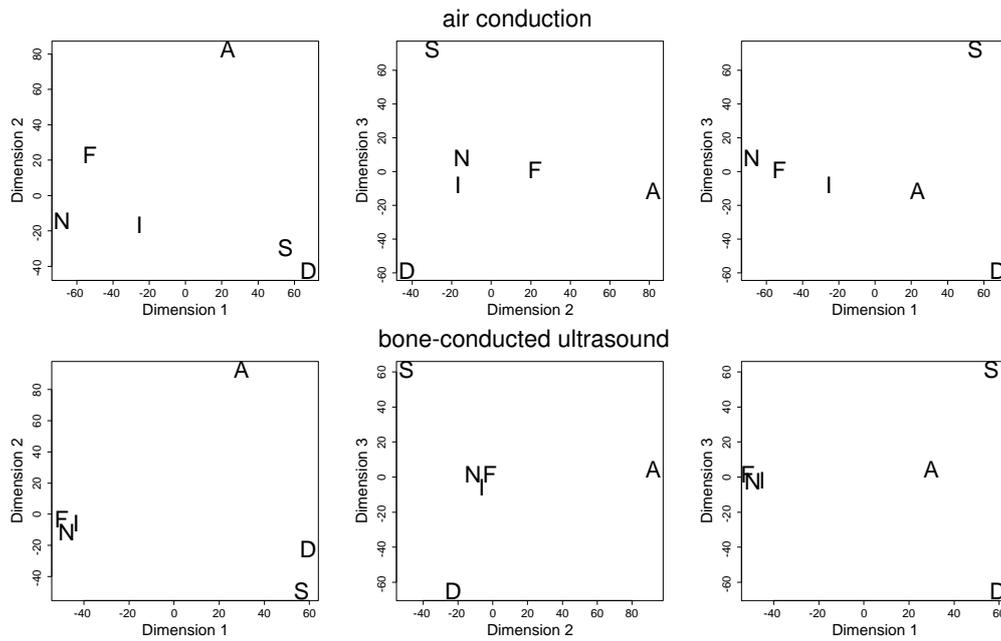
Figure 3: Distribution of intentions, determined from the results of MDS

To obtain more information from Table 1, a series of multidimensional scaling (MDS) analyses were conducted. The response data shown in Table 1 were regarded as psychological distances between intentions, and Kruskal's nonmetric multidimensional scaling was performed for the results of both AC and BCU.

Stress values were checked to decide appropriate numbers of dimensions. Each stress value revealed that three dimensions were sufficient for the results of both AC (0.001) and BCU (0.011).

3.2. Results of MDS

Figure 3 shows the distribution of the intentions; the distributions are obtained from the results of MDS. The result for AC conditions is positioned in the upper row, while that for BCU conditions is in the lower row.

The result for AC conditions in Figure 3 shows that "D", "S," and "A" have large values for a dimension of 1, while "N", "F," and "I" have small values. "D", "S," and "A" are salient intentions, while "N", "F," and "I" are not. Thus dimension 1 is interpreted as dimension of "salient" [4].

On the diagram showing a combination of the dimensions 2 and 3, "A", "D," and "S" are distributed in peripheral regions and "N", "F," and "I" are located at the center. For dimension 2, "A" lies on the extreme right, while "D" and "S" are distributed on the opposite side. "N", "I," and "F" are positioned at the center. "A" is positive and indicates a pleasant expression, whereas "S" and "D" are negative and denote unpleasant expressions. "N", "I," and "F" are neutral and positive. Thus, dimension 2 is interpreted as representing "pleasure." Next, for dimension 3, "S" is located in an upper position, "D" lies on the opposite side, and the others are distributed at the center. "S" corresponds to an extrovert, excited, and aroused disposition. On the other hand, "D" indicates a diametrically opposite disposition. Thus, dimension 3 is interpreted as corresponding to "arousal." For the combination of dimensions 2 and 3, the

intentions are distributed according to the "pleasure" level and "arousal" level. This result corresponds to Russell's circumplex model of affect [8].

The same tendencies are found in the case of BCU conditions. Dimension 1 can be interpreted as representing "salient"; dimension 2, the "pleasure" level; and dimension 3, the "arousal" level.

However, differences were also observed between the results of AC conditions and those of BCU conditions. For AC conditions, distances between "N", "F," and "I" were small, but these intentions were still separated. In contrast, under BCU conditions, these intentions were distributed in a quite small area and were not clearly separated in all dimensions. This difference indicates that listeners were not sensitive to the acoustic parameters that help discriminate "N", "F," and "I" in the case of BCU conditions.

3.3. Factors Influencing Ability to Discriminate between "N," "F," and "I"

As described above, the results of MDS suggest that BCU listeners have difficulty in perceiving acoustic cues that help in discriminating between "N," "F," and "I."

To obtain the parameters that help distinguish between "N," "F," and "I," a series of discriminant analyses and variable selection were conducted. Acoustic parameters presented in Table 2 were used as predictor variables. These parameters had also been used in previous studies for the analysis of paralinguistic information transmission [4]. All parameters were estimated for each stimulus. Wilks' Λ was used for choosing the best predictor subsets.

3.4. Results of Discriminant Analysis

Using the procedures described above, a predictor subset consisting of "Intensity," "F0 mean," and "F0 range" was selected. Table 3 shows the canonical discriminant functions and Table 4 presents the canonical discriminant coefficients obtained

Table 2: Acoustic variables used for the series of canonical discriminant analyses

F_0 mean	F_0 mean of whole utterance
F_0 range	F_0 range of whole utterance
Intensity	Maximum RMS power of whole utterance
Duration	Duration ratio of whole utterance
F1 of /a/	F1 value of last vowel /a/
F2 of /a/	F2 value of last vowel /a/

Table 3: Canonical discriminant functions

Func.1	Func.2
0.962	0.250

Table 4: Canonical discriminant coefficients

	Func.1	Func.2
Intensity	2.873	0.316
F_0 mean	-0.633	0.562
F_0 range	-1.316	0.228
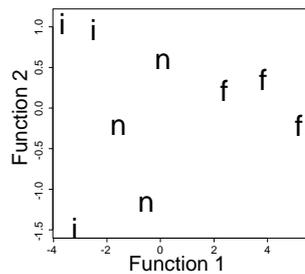
Figure 4: Distribution of the stimuli, obtained using the discriminant functions 1 and 2

by canonical discriminant analysis using the predictor subset. All the nine stimuli were correctly identified by using the functions, and Wilks' Λ was 0.070.

Figure 4 shows the distribution of the stimuli; the distributions are determined by using discriminant functions 1 and 2. First, as evident from Figure 4 and Table 3, each stimulus is well separated by function 1. Table 4 indicates that "Intensity" is the dominant variable of function 1. From these observations, it is inferred that the main factor that helps in discriminating between "N," "F," and "I" is the "Intensity." In other words, under BCU conditions, participants had difficulties in perceiving differences in the sound pressure level between the stimuli.

4. General Discussion

As described in section 3.1, *admiration*, *suspicion*, and *disappointment* are well perceived even under BCU conditions. These intentions can be distinguished by using the F_0 pattern or duration [4]. In our previous study, we showed that sensitivity to F_0 and duration is stable even in the case of BCU hearing [9]. The current results also indicate that the transmission of F_0 and duration information are robust in the case of BCU hearing.

On the other hand, intention types which contrasted by its intensities were not well discriminated. However, the psychological distances between these intentions were small under AC conditions; thus, it is understandable that the discrimination of these intentions is difficult under BCU conditions.

Although there was confusion when distinguishing between intensions on the basis of the F_0 pattern or duration, the perception pattern in the case of BCU hearing closely resembled that for AC hearing. This result indicates that users of the BCUHA can perceive speakers' intentions without much effort.

In addition, it is worth pointing out that the results of this study show the effectiveness of using paralinguistic information transmission experiments for the evaluation of hearing aids. The effectiveness results from two points: First, from the results of this study, we found that there were problems when using BCUHAs in communicational scene. These problems were not encountered in experiments using word intelligibility or syllable articulation. Second, we can now describe the advantages and disadvantages of BCUHAs in plain words, for example, "You can understand speakers' intention, but you may confuse focused and neutral voices." Such an easy description will help potential users of BCUHAs to understand the performance of BCUHAs.

5. Conclusion

The performance of the BCUHA was examined by considering the transmission of paralinguistic information. The results indicated that the BCUHA can transmit not only linguistic information but also paralinguistic information. Thus, it was shown that the BCUHA can help patients with sensorineural hearing loss to participate in expressive oral communications.

6. Acknowledgments

7. References

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

[2] M. L. Lenhardt, R. Skellett *et al.*, "Human ultrasonic speech perception," *Science*, vol. 253, pp. 82–85, 1991.

[3] Y. Okamoto, S. Nakagawa *et al.*, "Intelligibility of bone-conducted ultrasonic speech," *Hearing Research*, vol. 208, pp. 107–113, 2005.

[4] K. Maekawa, "Production and perception of 'paralinguistic' information," in *Proceedings of Speech Prosody 2004*, 2004, pp. 367–374.

[5] O. Mizutani and N. Mizutani, *Aural Comprehension Practice in Japanese*. Tokyo: The Japan Times, 1979.

[6] K. Maekawa, "Phonetic and phonological characteristics of paralinguistic information in spoken Japanese," in *Proceedings of Fifth International Conference on Spoken Language Processing (ICSLP 1998)*, 1998, pp. 635–638.

[7] K. Maekawa and T. Kagomiya, "Influence of paralinguistic information on segmental articulation," in *Proceedings of Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, vol. 2, 2000, pp. 349–352.

[8] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[9] T. Kagomiya and S. Nakagawa, "Perception of suprasegmental phoneme of Japanese through bone-conducted ultrasonic hearing," in *Proceedings of 2008 Autumn Meeting of Acoustical Society of Japan*, 2008, pp. 459–460.