Evaluation of a Bone-Conducted Ultrasonic Hearing Aid in Vocal Emotion Transmission

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

[1]Health Research Institute,
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 developed a bone-conducted ultrasonic hearing aid (BCUHA). The purpose of this study was to assess the usefulness of the BCUHA in transmission of the emotional state of the speaker. The evaluation used emotion-identification experiments. Types of emotion included Ekman's basic 6 emotions ("anger," "disgust," "fear," "joy," "sadness," and "surprise") and "neutral." The experiments were also conducted under air-conduction (AC) and cochlear implant simulator (CIsim) conditions. The results showed that emotion was transmitted more effectively with the BCUHA than CIsim.
Index Terms: hearing aid, ultrasound, bone-conduction, emotion, multi-dimensional scaling, discriminant analysis

1. Introduction

We have developed a bone-conducted ultrasonic hearing aid (BCUHA) for sensorineural hearing-impaired patients [1]. A BCUHA consists of 2 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). Moreover, if the ultrasound is amplitude modulated by speech sounds, original speech sounds can be perceived besides the carrier sound, and this amplitude-modulated BCU can be perceived by not only normal-hearing (NH) listeners but also hearing-impaired patients. We based the development of our BCUHA on these observations.

On the other hand, the cochlear implant (CI) was developed for and has been widely adopted in sensorineural hearing-impaired patients. A CI also consists of 2 components: speech signal processors and electrodes mounted in the cochlea. Although CIs have been widely adopted in hearing-impaired patients, some problems have been reported. The biggest problem is that a CI requires surgical positioning, which causes irreversible damage to the cochlea. Another problem is that the CI cannot transmit the whole sound, but only partial or reduced information, because performance is limited by the number of electrodes. Nowadays, the number of CI electrodes is limited to 12 to 24, and frequency resolution is limited according to the number of electrodes. This number may be sufficient for transmission of linguistic messages; however, it is reported that CI users have great difficulty in perception of music, speaker identity, emotional state, and so on [2, 3].

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

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, BCUHA listeners perceive high-frequency sound due to the carrier signal in addition to speech signals [1]. Therefore, the carrier-originated sound might prevent clear perception of speech sounds. Consequently, the performance of the BCUHA has been assessed in various ways.

The performance of the BCUHA regarding speech signal transmission has mainly been evaluated using monosyllables [4] or word intelligibility scores [1]. These studies found that syllable articulation scores with BCU were over 60% [4] and that word intelligibility scores for words with high familiarity were over 85% [1]. The patterns of confusion in speech perception in BCU have many points in common with air-conduction (AC) [4]. However, speech sounds convey not only linguistic messages but also indexical information about talkers, such as gender, age, identity, and emotional state.

As mentioned above, CI users have difficulty in perception of such messages; thus, if the BCUHA performs better in this respect, it has a great advantage over the CI.

The purpose of this study was to assess the performance of the BCUHA regarding transmission of speaker emotions compared with the CI. However, evaluation of both a BCUHA and CI is difficult in the same listener. Therefore, a CI simulator (CIsim) was adopted.

In this research, we conducted a series of listening experiments with BCUHA and CIsim conditions. Besides these conditions, AC condition experiments were conducted as a baseline condition.

Table 1: *Selected words*

num. mora	initial acc.	N-1 acc.	no acc
3	midori (green)	naname (slant)	nagame (scenery)
4	arawani (revealed)	arayuru (every)	omonaga (long-faced)
5	naniyorimo (first of all)	yawarageru (soften)	amarimono (remains)

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
pulsae rate	1515 pps/ch (18180 pps)

2. Methods

2.1. Stimuli

The experiment was designed using emotion detection tasks. Stimuli were selected according to the following procedures.

2.1.1. Types of emotions

For this task, 7 emotion types were selected. Six types out of the 7 were Ekman's basic emotions [5]: "anger," "disgust," "fear," "joy," "sadness," and "surprise." In addition to these emotions, "neutral" was selected.

2.1.2. Selection of sounds

Half of the sounds were obtained from the Keio Emotional Speech Database (Keio-ESD). This corpus was developed by the Keio University, and contains 20 words or phrases spoken by a male in his 30s with 47 emotions [6].

From the Keio-ESD corpus, 9 accentual phrases each spoken with the 7 emotions described above were selected according to accent types. The Japanese language has tonal or pitch accent. Manner of accent is one of the major cues used to convey emotional state or other paralinguistic messages [7]. Accent types selected were "initial accent," "N-1 accent," and "no accent." "Initial accent" words are enunciated with an accent on the first mora. "N-1 accent" words have a penultimate accent. In this study, "no accent" words included 2 types: words with no lexical accent and words with a final-mora accent; both these types of words have no accent in isolated speech. The number of moras was also considered in the selection of words, and all words had 3–5 moras. Following these procedures, the 9 words listed in Table 1 were selected.

As mentioned above, the speech sounds contained in the Keio-ESD were spoken by one male speaker. Thus, a female voice was recorded to balance the gender of speakers. A female speaker in her 20s participated in recording. She was an amateur actress and majored in opera singing in a university of music. The words and emotions were identical to those selected from the Keio-ESD.

As a result of these processes, 126 stimuli were generated (7 emotions $\times$ 9 words $\times$ 2 genders).

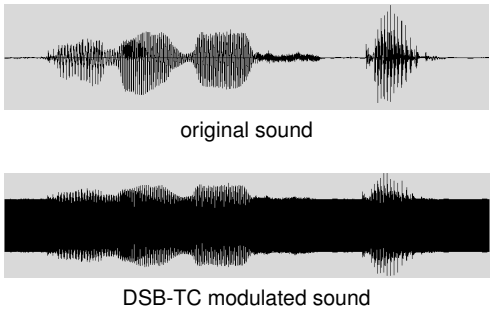

Figure 2: *DSB-TC amplitude-modulation*

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 the DSB-TC is the best amplitude-modulation method for the BCUHA [1, 4]. With the DSB-TC method, the modulated speech signals $U(t)$ are represented by the following expression:

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

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 left or right mastoid of the subject by using a hair band-like device (Figure 1).

2.2.2. Cochlear implant simulator

For generating 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 [8] was adopted in this study. In this study, the software was configured to simulate the MEDEL COMBI 40+ and TEMPO+ systems (see Table 2).

2.2.3. Participants

Nine native Japanese speakers (7 males and 2 females) with no reported hearing or speaking defects participated in the experiments. Their ages were in the range 19-42 years.

2.2.4. Procedures

For the AC and CIsim conditions, the sound stimuli were presented through a headphone (Sennheiser HD200). Participants were presented with the stimuli, and they were requested to identify the emotion from 7 choices: the 7 emotion types.

The order of the stimuli was counterbalanced. 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.

Table 3: *Confusion matrices of each emotions*

AC	anger	disgust	fear	joy	neutral	sad	surprise
anger	0.735	0.154	0.012	0.012	0.019	0.006	0.062
disgust	0.110	0.457	0.006	0.165	0.220	0.000	0.043
fear	0.000	0.006	0.411	0.000	0.000	0.454	0.129
joy	0.006	0.006	0.000	0.865	0.086	0.000	0.037
neutral	0.012	0.086	0.006	0.000	0.883	0.000	0.012
sad	0.006	0.000	0.165	0.000	0.000	0.823	0.006
surprise	0.074	0.012	0.123	0.049	0.019	0.031	0.691
correct perceived ratio: 0.695							
BCU	anger	disgust	fear	joy	neutral	sad	surprise
anger	0.500	0.173	0.025	0.062	0.123	0.025	0.093
disgust	0.123	0.370	0.043	0.105	0.327	0.012	0.019
fear	0.031	0.123	0.340	0.012	0.043	0.364	0.086
joy	0.105	0.043	0.012	0.660	0.136	0.000	0.043
neutral	0.056	0.099	0.000	0.031	0.784	0.012	0.019
sad	0.043	0.080	0.327	0.000	0.006	0.506	0.037
surprise	0.173	0.031	0.056	0.056	0.068	0.031	0.586
correct perceived ratio: 0.535							
CIsim	anger	disgust	fear	joy	neutral	sad	surprise
anger	0.506	0.128	0.056	0.039	0.111	0.089	0.072
disgust	0.167	0.222	0.111	0.056	0.272	0.094	0.078
fear	0.166	0.133	0.160	0.077	0.066	0.304	0.094
joy	0.178	0.100	0.022	0.544	0.083	0.011	0.061
neutral	0.056	0.156	0.022	0.056	0.622	0.072	0.017
sad	0.071	0.110	0.236	0.143	0.110	0.291	0.038
surprise	0.239	0.056	0.044	0.211	0.061	0.033	0.356
correct perceived ratio: 0.386							

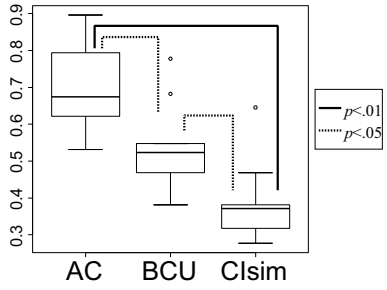

Figure 3: *Correct perceived ratio in each condition*

3. Results and Analysis

3.1. Confusion Pattern of the Emotions

Table 3 shows the confusion patterns of each emotion. Responses of all participants were pooled. Rows represent stimuli and columns show responses. Correct perception ratios of each emotion are shown as diagonal elements.

Further, as shown in Table 3, correct perceived ratio in the AC condition was 0.695, BCUHA was 0.535, and CIsim was 0.386. An ANOVA and a post hoc test (multiple comparison test with Holm's adjusting method) revealed that these differences were significant (1% level between AC and CIsim, 5% levels between AC and BCU and between BCU and CIsim; Figure 3).

3.2. Multi-Dimensional Scaling Analysis

To obtain more information from the confusion matrices, a series of Kruscal's multi-dimensional scaling (MDS) analyses were conducted. MDS analyses were applied for each confusion matrix shown in Table 3. Stress scores were checked and all scores were below 0.1 in the 2-dimensional condition (AC, 0.057; BCU, 0.076; CIsim, 0.107). These scores indicated that the fitness of the MDS models to the confusion matrices was "good." Figure 4 shows distribution of each emotion according to the results of MDS.

In the BCUHA condition diagram, "fear" and "sadness" had relatively small areas of distribution, which nearly overlapped. A similar tendency was also observed in the AC condition; "fear" and "sadness" were also closely distributed. Thus, it is inferred that these 2 types of emotions tended to be confused, even in the AC condition, and might understandably have been confused in the BCUHA condition.

On the other hand, in the CIsim condition, "surprise," "joy," and "anger" had small areas of distribution. These emotions were separate in the AC and BCUHA conditions. From this result, it is assumed that CI performance in transmission of acoustic cues that discriminate these emotions was inferior compared with the AC and BCUHA.

3.3. Factors to Discriminate "Surprise," "Anger," and "Joy"

As described above, the results of MDS suggest that CI listeners have difficulty in perceiving acoustic cues that help in discriminating among "surprise," "joy," and "anger." In this respect, the BCUHA is superior to the CI. Therefore, determining which cues contribute to discrimination of these emotions facilitates understanding of the characteristics of BCUHA perception and the differences between BCUHA and CI performances.

To determine the parameters used to discriminate "surprise," "anger," and "joy," a series of discriminant analyses and variable selection were conducted. Acoustic parameters listed in Table 4 were used as predictor variables, and all parameters were estimated for each stimulus. Wilks' Λ was adopted for choosing the best predictor subsets.

As a result of these procedures, a predictor subset consisting of "F0 avg" and "RMS avg" was selected. Table 5 shows the canonical discriminant functions and Table 6 presents the canonical discriminant coefficients obtained by canonical discriminant analysis using the predictor subset (Wilks' *Lambda* was 0.159). Discrimination of the stimuli applying the functions is shown in Table 7.

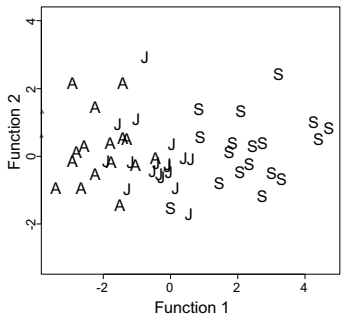

Figure 5: *Distribution of the stimuli, obtained using the discriminant functions 1 and 2. "A," "J," and "S" represent "anger," "joy" and "surprise" respectively.*

Figure 5 shows the distribution of the stimuli; the distributions were determined using discriminant functions 1 and 2. First, as evident from Figure 5 and Table 5, each stimulus was mainly separated by function 1. Table 6 indicates that "F0 avg" is the dominant variable of function 1. From these observations, it is inferred that the main factor that helps in discriminating

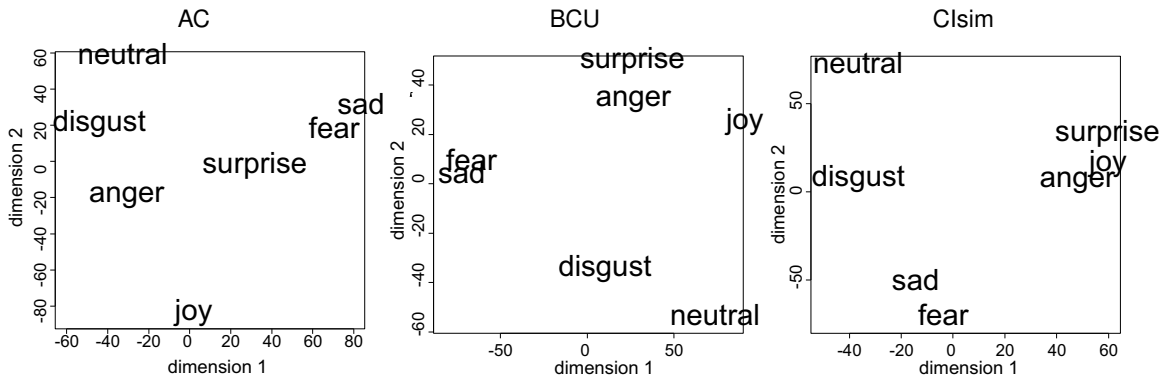

Figure 4: *Distribution of each emotion by the results of MDS*

Table 4: *Acoustic parameters applied for discriminant analysis*

F0 avg	mean value of F0 (standardized within speakers)
F0 SD	standard deviation of F0 (standardized within speakers)
RMS avg	mean value of RMS power
RMS SD	standard deviation of RMS power
Duration	duration of utterance
Speed	mora/second

Table 5: *Canonical discriminant functions*

Func.1	Func.2
0.9935	0.0065

Table 6: *Canonical discriminant coefficients*

	Func.1	Func.2
F0 avg	2.105	0.780
RMS avg	0.002	-0.003

Table 7: *Result of canonical discriminant analysis*

	anger	joy	surprise
anger	15	3	0
joy	3	15	0
surprise	0	1	17
accuracy: 0.870			

among "anger," "joy," and "surprise" is the average F0 value. In other words, under the CIsim condition, listeners may have had difficulties in perceiving differences in average F0 value level between the stimuli.

4. General Discussion

First, the results of this study reveal that emotional state is transmitted more efficiently with the BCUHA than CI. CI users are well known to have difficulty in perceiving speaker emotion [3]. From the results of MDS and discriminant analysis, this advantage of the BCUHA is mainly due to differences in transmission of F0 information. Further, CI users have difficulty in perceiving pitch information, for example, Chinese tone [9]. On the other hand, BCUHA listeners can detect Japanese pitch accent correctly [10]. The results of this study also indicate that the BCUHA performs better in F0 transmission than the CI.

In addition, the results of this study demonstrate the effectiveness of applying emotion transmission experiments for the evaluation of hearing aids. First, we found that there were problems when using BCUHAs in a 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 speaker emotion, but you may confuse sad and fearful voices." Such user-friendly descriptions will help manage the expectations of potential BCUHA users.

5. Conclusion

The performance of the BCUHA was evaluated by considering the transmission of speaker emotions. The evaluation was conducted by comparison with the AC and CIsim conditions. The results showed that the BCUHA can transmit speaker emotion better than the CIsim. This result indicates that the BCUHA is superior to the CI in this respect.

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] Nakagawa, S. Okamoto, Y. and Fujisaka, Y., "Development of a bone-conducted ultrasonic hearing aid for the profoundly sensorineural deaf," *Trans. of the Japanese Soc. for Medical and Biological Engineering : BME*, 44(1):184–189, 2006.

[2] Fu, Q.-J., Chinchilla, S. and Galvin, J. J., "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," *J. of the Association for Research in Otolaryngology*, 5(3):253–260, 2004.

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

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

[5] Ekman, P. and Friesen, W. V., *Unmasking the face: a guide to recognizing emotions from facial clues*. Prentice-Hall, 1975.

[6] Moriyama, T., Mori, S. and Ozawa, S.,"A synthesis method of emotional speech using subspace constraints in prosody," *J. of Information Processing Soc. of Jpn.*, 50(3):1181–1191, 2009, (in Japanese).

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

[8] De La Torre Vega, Á., Martí, M. B., De La Torre Vega, R. and Quevedo, M. S., *Cochlear Implant Simulation version 2.0*, http://www.ugr.es/˜atv/web_ci_SIM/en/ci_sim_en.htm

[9] Wei, C.-G., Cao, K., and Zeng, F.-G., "Mandarin tone recognition in cochlear-implant subjects," *Hearing Research*, 197:87–95, 2004.

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