Evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of paralinguistic information: A comparison with cochlear implant simulator

Takayuki Kagomiya¹, Seiji Nakagawa¹

¹Health 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 been developing a bone-conducted ultrasonic hearing aid (BCUHA). The purpose of this study is to evaluate the usability of BCUHA regarding transmission of paralinguistic information. For this purpose, two series of listening experiments were conducted. One is a speaker’s intention identification experiment, the other is a speaker discrimination experiment. To compare performance of BCUHA to that of air-conduction (AC) and cochlear implant, both experiments were conducted under three conditions: BCUHA, AC, and cochlear implant simulator (CIsim). The results show that BCUHA can transmit intentions of speaker as well as CIsim. Also BCUHA can transmit speaker information better than CIsim.

Index Terms: ultrasound, bone-conduction, hearing aid, speaker discrimination, multi-dimensional scaling

1. Introduction

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 [1]. In addition, if BCU signals are amplitude-modulated by speech signals, listeners can perceive the original speech signals [1]. These voice-modulated BCU signals enable patients with acute sensorineural hearing loss to perceive speech signals [2]. Considering this fact, we have been developing a bone-conducted ultrasonic hearing aid (BCUHA).

The performance of the BCUHA was evaluated by using syllable articulation and word intelligibility. Syllable articulation scores were over 60% [3] and word intelligibility scores for words with high familiarity were over 85% [2]. The patterns of confusion in speech perception in the case of BCU have many points of similarity 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 textual messages. In other words, little attention has been paid to the transmission of paralinguistic or nonlinguistic information.

The main purpose of this study is to evaluate the usability of the BCUHA by considering paralinguistic information transmission, especially the transmission of speakers’ intention and the features of speakers. In daily oral communications, listeners perceive speakers’ intention as well as linguistic or textual information. For example, the English word “Really” can represent various intentions or attitudes such as “I don’t believe it” and “I’m surprised to know that.” Likewise, listeners are able to perceive features of speakers such as sexuality, age, body size, etc. This 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.

Another purpose of this study is to compare the usability of BCUHA with that of the cochlear implant (CI). CI is another accepted device to provide speech sounds for sensorineural hearing loss patients. CI gives partial hearing to patients by electric stimulation of the auditory nerve through an electrode surgically implanted in the cochlear. However, the quality of CI sounds is not as clear as normal hearing, thus CI users have difficulties perceiving voice gender, talker identity, and the speaker’s emotion [4]. Thus if BCUHA can transmit paralinguistic information well, it has a great advantage over CI. However, it is difficult to evaluate the usability of BCUHA and CI by the same listener. Therefore a CI simulator (CIsim) was adopted.

In this research, we conducted two series of listening experiments that involve a speaker’s intention identification task and a speaker discrimination task. Both tasks were conducted in AC, BCU, and CIsim conditions.

2. Method

2.1. Speech material

2.1.1. Speaker’s intention identification experiment

The speaker’s intention identification task was designed for identification of six intentions: admiration (A), suspicion (S), disappointment (D), indifference (I), focused (F), and neutral (N). 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].

With these intention, three sentences “So-desu-ka” (Is that so?) “Anata-desu-ka” (Is that you?) “I love you,” and “I’m surprised to know that.” Likewise, listeners are able to perceive features of speakers such as sexuality, age, body size, etc. This 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.

With these intention, three sentences “So-desu-ka” (Is that so?) “Anata-desu-ka” (Is that you?) “I love you,” and “I’m surprised to know that.” Likewise, listeners are able to perceive features of speakers such as sexuality, age, body size, etc. This 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.

The sentences were selected because they can convey all the intentions described above. 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 don’t believe it.” These textual messages were identical to those used in previous studies [6].
2.1.2. Speaker discrimination experiment

In the speaker discrimination experiment, whether or not the differences of F0 value and formant space size can be cues for speaker discrimination was examined. It is widely known that adult males have low F0, children have a high level, and females have a medium level. This tendency is derived from the difference in the larynx size and vocal fold length [8]. Likewise, format space size reflects the speaker’s vocal tract length. Adult males have a long vocal tract, children have a short one, and females have a medium one. Corresponding to this difference of vocal tract length, adult males have a narrow formant space, children have a wide one, and females have a medium one [9]. Moreover, listeners use these relationships between physical and acoustical characteristics as cues for speaker discrimination or identification [10].

In this research, whether or not F0 and formant space size can be cues to discriminate speakers was examined. For this purpose, a series of synthesized sounds which manipulated their F0 and formant space size was generated [11]. An original sound was selected from the Corpus of Spontaneous Japanese (CSJ) [12]. CSJ contains two short passage reading tasks. For this research, the first phrase “uchu:” (the universe) in the “Uchu:” dataset was used. An utterance whose F0 and speaking rate was nearest to the mean value among the corpus was selected. The F0 value of the selected utterance was 127.871 Hz, and the speaking rate value was 6.315 moras per second.

From the original sound, sound stimuli were generated by modifying the sound’s F0 value and formant space size in three levels. F0 values were modified to 20% lower, 20% higher and the original value. The lower and higher values were almost the mean value ±1SD in female voices of the “Uchu:” dataset [11], thus neither value was of an unusual level. Moreover, the lower value was lower than the mean +1SD in male voices in the equivalent part of CSJ (104.391 Hz), thus the value was not an unusual level even in a male voice.

Similar to the F0 value, formant space size was stretched by 20%, shrank by 20%, and kept at the original size. These values were determined to represent the ratio of vocal tract length of males to that of females [8, 11].

All manipulation was conducted by using STRAIGHT [13]. Counting the results of each of the three levels of manipulation in F0 (Low, Original, High; L, O, H) and formant space (Narrow, Original, Wide; N, O, W), a total of nine stimuli were generated.

2.2. Participants

Nine native Japanese speakers (2 males and 7 females) with no reported hearing or speaking defects participated in the experiments. Their ages were in the range 19-26 years.

2.3. Cochlear implant simulator

For generating CI simulated sounds, Cochlear Implant Simulation (http://www.ugr.es/iat/web/ci_SIM/en/ci_sim_en.htm) developed by the University of Granada was adopted in this study. In this study, the software was configured to simulate the MEDEL COMBI 40+ and TEMPO+ system (see Table 1).

2.4. Presentation of the sounds

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

Table 1: 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) |

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

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 [2, 3]. With the DSB-TC method, the modulated speech signals U(t) are given by the following expression:

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

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 stimuli under BCU conditions were presented using a custom-made ceramic vibrator (Figure 1). Bone-conducted ultrasound can be perceived when it is applied to various parts of our 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 using a hair-band-like supporter (Figure 1).

2.5. Procedures

In the speaker’s intention identification task, each stimulus was presented at random, and participants were asked to identify the intention underlying the stimulus. Each stimulus was presented six times. In the speaker discrimination task, each stimulus was presented as a pair, and participants were requested to make a judgment about whether the speakers were the “same” or “different”. Each pair was presented ten times. The overall order of presentation of the pairs was also randomized.

All listeners participated in the BCU condition first, an then a few days later took part in the CI sim condition. The last session which used the AC condition was conducted a further few days later. All experiments were conducted in a soundproof chamber. The sound levels of the stimuli were adjusted to the most comfortable levels for each participant.

2.6. Analysis

The responses of all participants were pooled for each experiments. For the speaker’s intention identification task, the response rates were calculated for each intention and a confusion matrix was created. Table 2 shows the confusion matrix; the presented stimuli are the column headings and the participants’
choices are the row headings. According to Table 2, intentions “A”, “D”, and “S” have a higher correct perception rate (more than 0.850) under all conditions, in other words, these intentions were correctly perceived by the listeners under all conditions.

To obtain more information from Table 2, a series of multidimensional scaling (MDS) analyses were conducted. The response data shown in Table 2 were regarded as psychological distances between intentions, and Sammon’s nonlinear mapping method was used. Stress values were checked to decide appropriate numbers of dimensions. Each stress value revealed that three dimensions were sufficient for all the results. (AC: 0.001, CIsim: 0.001, BCU: 0.011).

3. Results and discussion

3.1. Speaker’s intention identification experiments

Figure 2 shows the distribution of the intentions; the distributions are obtained from the results of MDS. The result for AC conditions in Figure 2 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 a dimension of “salient” [7].

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 [14].

The same tendencies are found in the case of CIsim and 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 as well as those of the CIsim and BCU conditions. For the AC conditions, the differences between “N”, “F”, and “I” were small, but these intentions were still separated. In contrast, under the CIsim and BCU conditions, these intentions were distributed in quite a 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 the CIsim and BCU conditions.

3.2. Speaker discrimination experiments

Figure 3 shows the distribution of each stimulus according to the results of MDS. The shapes of the points indicate the F0 value (upward triangles: “H”, downward triangles: “L”, square: “O”), and the colors of the points represent the formant space size (black: “W”, white: “N”, grey: “O”). In the case of AC, three tight clusters were created in the 1-2 dimensions. Each cluster was grouped according to formant spaces. However, in the 2-3 dimensions, the stimuli were located by the order of the formant spaces size on dimension 2, and placed according to the F0 value on dimension 3. Although the distribution in dimen-
was distributed according to the F0 value. Moreover, in the 1-3 size of the formant spaces, while in dimension 2, every stimulus dimensions in BCU. In dimension 1, each sound was ordered by the BCU. However, there were clear tendencies in the 1-2 dimensions, the stimuli were loosely grouped by the F0 value. Moreover, in the 1-3 dimensions, the stimuli were loosely grouped by the F0 value. These results revealed that both the F0 value and formant space is a primal factor for speaker identification.

The usability of BCUHA regarding transmission of paralinguistic information was evaluated. The evaluation was conducted by comparing AC and CIsim. The results showed that the speaker’s intention was equally well transmitted when using CIsim, and comparing AC and CIsim. The results showed that the speaker’s intention identification experiments

4. Conclusions

The usability of BCUHA regarding transmission of paralinguistic information was evaluated. The evaluation was conducted by comparing AC and CIsim. The results showed that the speaker’s intention was equally well transmitted when using CIsim, and that information about formant space which plays an important role in speaker discrimination was better transmitted by BCUHA than by CIsim. This result indicates that BCUHA has an advantage in speaker identification over CI.

5. Acknowledgements

This research was supported by the 2006FY Industrial Technology Research Grant Program of the New Energy and Industrial Technology Development Organization (NEDO) Japan, 2009FY Grants-in-Aid for Scientific Research of Japan Society for the Promotion of Science, and the 2009FY Research Grant of the Sound Technology Promotion Foundation of Japan.

6. References