



Evaluation of Bone-conducted Ultrasonic Hearing-aid Regarding Transmission of Speaker Discrimination Information

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 speaker discrimination information. For this purpose, a prototype of speaker discrimination test was developed. The test consists of 120 pairs of 10 words spoken by 10 speakers, and examinee is requested to judge the speakers of each pair are “same” or “different”. The usability of BCUHA was assessed by using the speaker discrimination test. The test was also conducted to air-conduction (AC) and cochlear implant simulator (CIsim) condition. The results show that BCUHA can transmit speaker information speaker as well as CIsim.

Index Terms: ultrasound, bone-conduction, hearing aid, speaker discrimination.

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 is defined as sound waves which travel at such a high frequency that they cannot be heard by humans. However, if the ultrasound is presented through a bone-conducted stimulator (bone-conducted ultrasound, BCU) the ultrasound 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.

On the other hand, cochlear implant (CI) is also a 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 implanted in the cochlea. However, a surgery operation is required to mount electrode in cochlear. In contrast, the BCUHA does not need any surgery operation, but simply attached to mastoid. Thus this device is more user-friendly than CI.

Usability of the BCUHA have been assessed 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% [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].

Also the patterns of these scores have similarities of those of CI. This result suggested that the performance of BCUHA regarding transmission of linguistics messages reaches the level of CI.

As described above, performance of BCUHA was assessed mainly focused on transmission of linguistic message. However, we are receiving more information from speech sounds besides linguistic messages, for example, we can perceive speakers' gender, age, and emotion etc. from speech sounds. These extralinguistic messages or paralinguistic messages enrich oral communication in comparison with written language, thus hearing assistance devices must transmit such messages.

Recently, we have been focusing on evaluation of BCUHA regarding transmission of paralinguistic messages. Speakers' intention [4] and speaker information [5] can be transmitted as compared with CI simulator (CIsim) [6]. Also performance of CI regarding speaker discrimination was evaluated recently [7]. If BCUHA has better performance at speaker discrimination than CI, it has a great advantage over CI.

In this research, we assessed usability of BCUHA regarding speaker discrimination. First, we developed a prototype of speaker discrimination test for hearing assistance devices. Second, a series of experiment using the speaker discrimination test was conducted. The experiment was conducted in BCUHA, AC and CIsim conditions.

2. Development of Speaker Discrimination Test

We developed a speaker discrimination test based on German version of test [7], however, we arranged some attributes to adopt Japanese language. The test was designed as speaker discrimination task. Examinees are requested to judge the speakers of two sounds are ‘same’ or ‘different’.

2.1. Speech Material

The speech material was extracted from the “Japanese phonetically-balanced word speech database” which was developed by Electrotechnical Laboratory, Japan (ETL-WD corpus). The corpus consists of 1542 phonetically balanced real words read by 10 male and 10 female native Japanese speakers. The speech sounds are recorded at 16 bit / 16 kHz sampling quality. Also F0 and F1-F4 formant information data are attached to each speech file.

2.2. Selection of Speakers

Five male and five female speakers were selected, considering their F0 and formant space size.

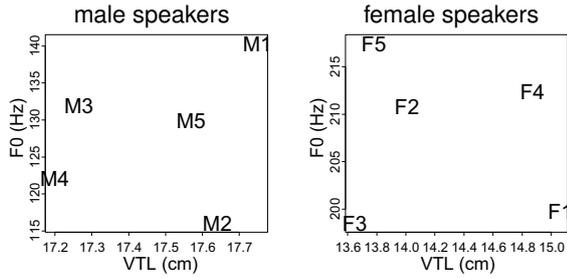


Figure 1: F0 and estimated VTL of the selected speakers

F0 is one of the most salient speaker-specific features. As is well known, the F0 for an adult male voice is low, for a child it is high, and the adult female’s is medium. These differences of F0 are derived from differences in the size of the speaker’s vocal cords. The F0 value of each speaker was estimated from the F0 data attached to the corpus by calculating the average value of each speaker.

Also, the formant space size reflects the speakers’ vocal tract length (VTL). If VTL is long, the formant space size is narrow. Thus, the adult male has a narrow formant space, children have a wide one, and females have a medium-sized one. According to source-filter theory, each formant (F_n) can be predicted using the following equation:

$$F_n = \frac{(2n-1)c}{4l} \quad (1)$$

where c is the speed of sound (34400 cm/sec.), and l is vocal tract length. Vocal tract length can be estimated from formant frequencies using inverse operations of the equation. Thus the vocal tract length was estimated from measured formant frequencies using the following procedure:

1. Vocal tract lengths l were estimated from each F_n using the following formula (reverse operation of equation (1).):

$$l = \frac{(2k+1)c}{4F_{k+1}} \quad (2)$$

where $k = (0, 1, 2)$, F_{k+1} is the formant frequency of interest.

2. The mean value of l calculated from the three formants of each speech file was regarded as the estimated vocal tract length (VTL).

Considering these F0 and VTL values, 1) high F0 and short VTL speaker, 2) high F0 and long VTL, 3) low F0 and short VTL, 4) low F0 and long VTL, 5) middle F0 and middle VTL were selected respectively from males and females (Fig. 1).

2.3. Selection of words

From the 1542 words stored in the corpus, 10 words were selected (Table 1) under the scheme described below.

2.3.1. Number of moras

The length of words was fixed. Japanese word length is usually counted in moras. In this research, word length was set at four moras. This number of moras is also applied in “Familiarity-Controlled Word Lists” [8], because four-mora words are the most frequently occurring type in Japanese [8].

Table 1: List of the selected words

word ID	transcription	translation	familiarity
W0041	bon-pyaku	various	1.438
W0231	kokubyaku	black and white	1.812
W1533	yuiwata	a traditional woman’s hair style	1.938
W0109	gen-un	vertigo	2.125
W1405	ron-kitsu	confute	2.281
W1069	bin-patsu	hairs at the sideburns	2.344
W1484	ten-ita	top board	2.344
W0005	an-utsu	melancholy	2.625
W1518	yaseyama	barren mountain	2.750
W0402	soeuma	carriage horses	2.844

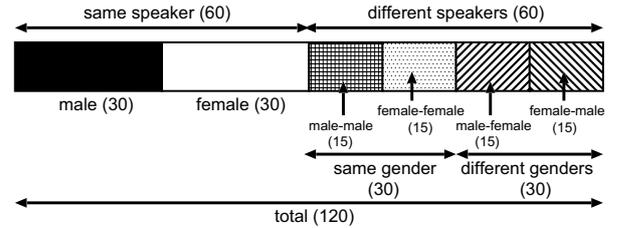


Figure 2: Component ratio of speaker-combination

2.3.2. Lexical accent

Japanese is a pitch-accented language. In this research, accent type was also fixed. The L-H-H-H type (unaccented type) was selected because this type is most common in Tokyo Japanese [8].

2.3.3. Familiarity

Word familiarity has a correlation with the speaker identification score [9] therefore, to avoid this effect, low familiarity score words were selected. Familiarity scores were extracted from “Lexical Properties of Japanese” [10].

2.4. Sound Pressure Level

The sound files were normalized for overall amplitude.

2.5. Stimuli Lists

To control the difficulty and procedure of the test, randomized stimuli lists were generated (Fig. 2). Each list consisted of 120 word pairs. In each list, the number of pairs in the “same-speaker condition” and “different-speaker condition” were balanced (each had 60 pairs).

In the different-speaker condition, 30 pairs represented a “both speakers are different-gender condition”, and the other 30 pairs represented a “both speakers are same-gender condition”.

The same-gender condition was divided into two parts; 15 pairs with a “both speakers are male condition” and 15 pairs with a “both speakers are female condition”.

Also the different-gender condition was divided into two parts; 15 pairs of “male-female order” and 15 pairs of “female-male order”.

If the stimuli sentence was the same, the speaker discrimination score increased compared with the different sentence condition [11]. To avoid this effect, we ensured that each word pair in the lists contained different words.

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)



Figure 3: Ceramic vibrator of the BCUHA attached to the mastoid with a hair-band-like device

3. Evaluation of BCUHA

A series of experiments was conducted to assess the performance of BCUHA using the speaker discrimination test described above. 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).

3.1. Participants

14 native Japanese speakers (7 males and 7 females) with no reported hearing or speaking defects participated in the experiments. Their ages were in the range 18-49 years.

3.2. Cochlear implant simulator

For generating CI simulated sounds, 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. In this study, the software was configured to simulate the MEDEL COMBI 40+ and TEMPO+ system (see Table 2).

3.3. Presentation of the sounds

The sound stimuli were presented through a headphone (Sennheiser HD200) under AC and CIsim 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:

$$U(t) = (S(t) - S_{\min}) \times \sin(2\pi f_c t) \quad (3)$$

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 3). 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

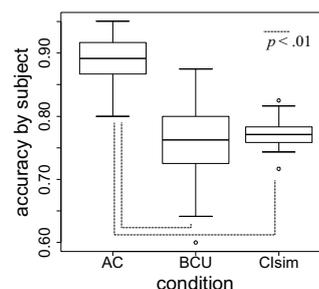


Figure 4: Correct response ratio by subject

perception is high. Therefore, we applied the vibrator to the left or right mastoid of the subject using a hair-band-like supporter (Figure 3).

All listeners participated in the BCU condition first, and then a few days later took part in the CIsim 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.

4. Analysis and Results

4.1. Correct response ratio of each condition

Correct response ratios of each condition of each participant were estimated. Figure 4 shows distribution of correct response ratio of each participant for each condition.

AC condition has the highest average score at 0.871 (SD=0.046), while BCU condition has the lowest score at 0.750 (SD=0.071). Even in AC condition, average score +1SD was lower than 1.0, while average score -1SD was higher than chance level (0.5) in BCU condition, thus it was regarded that floor effect and ceiling effect were not observed in this experiment.

A repeated one-way ANOVA and post-hoc test (Tukey's HSD) revealed that difference of correct response ratio between AC and BCU, AC and CIsim were significant ($p < 0.001$), while between BCU and CIsim was not significant. This result indicates that BCUHA users can distinguish speakers as accurate as CI users can do.

4.2. Confusion ratios of each speaker

Responses of all participants were pooled for each condition and confusion ratios of each speaker were estimated. Table 3 shows confusion ratios between each speaker. Also correct response ratio for "same-speaker condition", "both speakers are male condition", "both speakers are female condition" and "both speakers are different gender condition" were estimated for each hearing condition.

Diagonal elements of Table 3 which represent same-speaker condition have high confusion scores. On the other hand, below-left elements which represent different-gender condition have low confusion scores in each condition.

A series of binomial test revealed that the correct response ratio was over chance level in all conditions except for female-female pairs of CIsim and BCU conditions at 5% significant level. At the female-female pairs of CIsim and BCU condition, the correct response ratio was significantly lower than chance level.

Table 3: Confusion ratios between each speaker

AC	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
M1	0.979									
M2	0.000	0.667								
M3	0.000	0.083	0.968							
M4	0.000	0.368	0.350	0.859						
M5	0.038	0.067	0.000	0.321	0.533					
F1	0.000	0.000	0.050	0.000	0.000	0.987				
F2	0.000	0.056	0.000	0.000	0.000	0.125	0.928			
F3	0.000	0.000	0.067	0.000	0.000	0.043	0.267	0.941		
F4	0.000	0.000	0.000	0.000	0.000	0.250	0.318	0.111	0.960	
F5	0.000	0.000	0.000	0.000	0.000	0.300	0.360	0.357	0.368	0.863
	accuracy overall:0.884, same speakers:0.901, male-male:0.871, female-female:0.749, different genders:0.993									
CIsim	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
M1	0.787									
M2	0.045	0.662								
M3	0.667	0.238	0.878							
M4	0.250	0.550	0.571	0.750						
M5	0.500	0.286	0.316	0.586	0.632					
F1	0.143	0.048	0.053	0.000	0.000	0.948				
F2	0.000	0.000	0.056	0.000	0.000	0.550	0.930			
F3	0.053	0.000	0.000	0.000	0.000	0.913	0.600	0.913		
F4	0.000	0.000	0.000	0.000	0.000	0.643	0.739	1.000	0.900	
F5	0.000	0.000	0.000	0.000	0.000	0.667	0.696	0.733	0.789	0.942
	accuracy overall:0.772, same speakers:0.771, male-male:0.590, female-female:0.276, different genders:0.988									
BCU	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
M1	0.747									
M2	0.143	0.615								
M3	0.368	0.211	0.772							
M4	0.364	0.619	0.429	0.629						
M5	0.393	0.333	0.294	0.355	0.493					
F1	0.000	0.000	0.000	0.000	0.000	0.778				
F2	0.000	0.062	0.053	0.000	0.000	0.278	0.900			
F3	0.048	0.000	0.000	0.000	0.000	0.630	0.593	0.786		
F4	0.000	0.050	0.062	0.000	0.071	0.448	0.550	0.333	0.878	
F5	0.000	0.000	0.000	0.000	0.000	0.647	0.700	0.769	0.905	0.803
	accuracy overall:0.750, same speakers:0.757, male-male:0.638, female-female:0.419, different genders:0.986									

These results indicate that; BCUHA users can detect the same speakers as “the same” to a certain degree; also BCUHA users can distinguish speakers if genders of them are different; however BCUHA users have difficulty to distinguish speakers if gender of them is female; and these tendencies are also observed in CIsim condition.

5. Discussions

The results of the experiments revealed advantages and disadvantages of BCUHA. As an advantage of BCUHA, BCUHA listeners can speaker discrimination. Especially, if gender of speaker is different, BCUHA listeners can discriminate speakers at 98% accuracy. This result indicates that the BCUHA can transmit speakers’ gender information. The reason for high accuracy in gender discrimination can be explained as difference of F0 value between male and female. BCUHA listeners can perceive Japanese pitch accent [12] or prosodically-salient paralinguistic information [4]. The result of this study also shows BCUHA can transmit F0 information.

On the other hand, BCUHA listeners have difficulty to discriminate same gender speakers. This result indicates some kinds of timbre information of voice is lost in BCU listening. However, this weak point was also observed in CIsim, and significant difference of the correct discrimination ratio between BCU and CIsim was not observed. Thus the results of this study indicates that the usability of the BCUHA regarding speaker discrimination reaches the level of CI.

Another contribution of this study is to develop an assessment tool for evaluation of hearing assistance devices. The

speaker discrimination test which developed in this study can represent usability of hearing assistance devices regarding transmission of speaker information. Thus this test is able to contribute assesment and development of various hearing assistance devices.

6. Summary and Conclusions

The usability of BCUHA regarding transmission of speaker information was evaluated. For this purpose a speaker discrimination test was developed. The evaluation was conducted by comparing AC and CIsim. The results showed; the usability of BCUHA regarding speaker discrimination reaches the level of CI. BCUHA users can discriminate speakers if gender of speakers are different, however BCUHA users have difficulty to discriminate speakers if gender of speakers are the same. To solve this problem, further investigations are required.

7. Acknowledgements

This research was supported by Grants-in-Aid for Scientific Research of Japan Society for the Promotion of Science (21700592, 22680038) and the 2009FY Research Grant of the Sound Technology Promotion Foundation of Japan.

8. References

- [1] S. Nakagawa, Y. Okamoto, and Y. Fujisaka, “Development of a bone-conducted ultrasonic hearing aid for the profoundly sensorineural deaf,” *Trans. of the Japanese Soc. 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] Y. Okamoto, S. Nakagawa, K. Fujimoto, and M. Tonoike, “Intelligibility of bone-conducted ultrasonic speech,” *Hearing Research*, vol. 208, pp. 107–113, 2005.
- [4] T. Kagomiya and S. Nakagawa, “An evaluation of bone-conducted ultrasonic hearing aid regarding perception of paralinguistic information,” in *Proc. Speech Prosody 2010*, 2010, pp. 100 867:1–4.
- [5] —, “The effects of fundamental frequency and formant space on speaker discrimination through bone-conducted ultrasonic hearing,” in *Proc. Interspeech 2009*, 2009, pp. 2915–2918.
- [6] —, “Evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of paralinguistic information: A comparison with cochlear implant simulator,” in *Proc. Interspeech 2010*, 2010, pp. 2474–2477.
- [7] R. Müller, M. Ziese, and D. Rostalski, “Development of a speaker discrimination test for cochlear implant users based on the oldenburg logatome corpus,” *J. of Oto-Rhino-Laryngology, Head and Neck Surgery*, vol. 71, no. 1, pp. 14–20, 2009.
- [8] S. Amano, S. Sakamoto, T. Kondo, and Y. Suzuki, “Development of familiarity-controlled word lists 2003 (fw03) to assess spoken-word intelligibility in japanese,” *Speech Communication*, vol. 51, pp. 76–82, 2009.
- [9] K. Amino and T. Arai, “Effects of linguistic contents on perceptual speaker identification: Comparison of familiar and unknown speaker identifications,” *Acoust. Sci. Tech.*, vol. 30, pp. 89–99, 2009.
- [10] S. Amano and T. Kondo, *Lexical Properties of Japanese*. Tokyo: Sanseido, 1999.
- [11] J. Gonzalez and J. C. Oliver, “Gender and speaker identification as a function of the number of channels in spectrally reduced speech,” *J. Acoust. Soc. Am.*, vol. 118, pp. 461–470, 2005.
- [12] T. Kagomiya and S. Nakagawa, “An evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of japanese prosodic phonemes,” in *Proc. ICA 2010*, 2010, pp. 972:1–5.