ICIC Express Letters
Volume **16**, Number **5**, May **2022**

 ISSN 1881-803X
pp. 513–520

ROBUST PHONETIC FEATURES IN BONE-CONDUCTED SPEECH COMMUNICATION

Yasuki Murakami[1,*] and Hiroki Kurita[2]

[1]Faculty of Design
Kyushu University
4-9-1 Shiobaru, Minamiku, Fukuoka 815-8540, Japan
*Corresponding author: murakami@design.kyushu-u.ac.jp

[2]Electronic-Mechanical Engineering Department
National Institute of Technology, Oshima College
1091-1, Komatsu, Suo-Oshima, Oshima, Yamaguchi 742-2193, Japan

Received September 2021; accepted November 2021

Abstract. *Bone conduction (BC) has attracted attention as an alternative sound transmission method in loud environments, as well as in environments where air conduction (AC) cannot be used. This is because BC uses different sound transmission paths than air conduction. However, the speech produced by a sound transmitted through BC is not as intelligible as that transmitted through AC. In addition, the speech intelligibility of BC sounds as recorded by the BC microphone was lower than that of the AC sound. In this study, we investigated the possibility of communicating using both BC microphones and transducers in a loud environment. To investigate this, we recorded speech using the BC and AC microphones and measured the speech intelligibility using the BC transducer. The results show that the BC sound lost some phonetic features when compared to the AC sound; however, it conserved some phonetic features. Therefore, we conclude that improving the intelligibility of these lost phonetic features can enhance BC speech communication.*

Keywords: Bone conduction, Speech perception, Speech production, Phonetic features

1. **Introduction.** Speech communication is a necessary part of our lives, and hearing impairments such as age-related hearing loss, may severely hamper functioning. In addition, the difficulty of speech communication in occupational environments due to loud noises and the use of hearing protectors such as earplugs poses serious safety issues. This is true even among workers with normal hearing.

In this study, we focused on BC sounds that can be heard while wearing hearing protectors. BC sounds reach the inner ear through vibrations of the skull and skin [1]. In other words, BC sounds can vibrate and reach the inner ear directly without the use of air as a medium. Conversely, AC sounds reach the inner ear through the air, outer ear, and middle ear. When comparing the BC and AC pathways, we found that BC sound perception was less sensitive to the conditions of the ear and sound field than AC sound perception. As mentioned above, there are two main obstacles to verbal communication: hearing impairment and being in a loud environment. With regard to hearing loss, BC hearing aids have been commercialized for hearing assistance. Because a minimum level of speech understanding is guaranteed, speech communication is possible without an AC sound [2]. For this reason, the BC transducer may also be useful in loud environments for speech communication where AC sound cannot be used.

BC sounds are not only perceived but are also produced by the vocal cords and vocal tract as vibrations that propagate through the bone and skin. Normally, vibrations from

DOI: 10.24507/icicel.16.05.513

AC speech are emitted by the vocal cords and vocal tract and released through the lips into the air. When comparing the production of BC and AC speech, BC sounds are not emitted into the air like AC sounds. This principle has allowed the recording of BC sounds in very noisy environments using a BC microphone [3].

Combining and simultaneously using the BC transducer and the BC microphone is effective because both devices are robust in noisy environments, as shown in Figure 1. However, there has been little research about using BC microphones and BC transducers simultaneously because AC and BC microphones are developed independently [2, 3]. Therefore, it is unclear whether speech intelligibility is affected or possibly even enhanced when the sound recorded by the BC microphones is perceived by BC transducers.

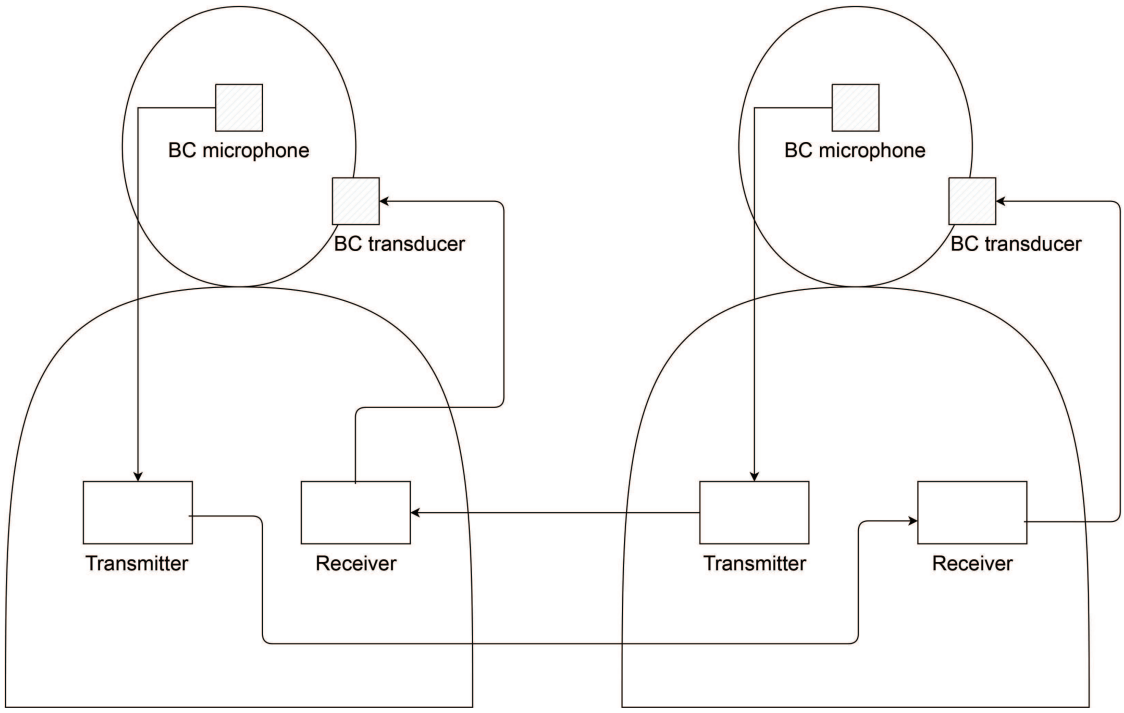


FIGURE 1. Conceptual diagram of bone conduction (BC) speech communication. The BC microphones and the BC transducers are attached to the head, which captures the speech signal, sends it to the transmitter, and vibrates the skin through the receiver to perceive the speech signal.

The goal of this study was to evaluate speech communication using both BC microphones and BC transducers in loud environments. However, the BC sounds recorded and presented are of low quality. This must be enhanced by pre-processing. However, it is unclear which phonetic feature should be enhanced and is more effective for BC communication. In this study, we investigated the phonetic features of BC-recorded speech transmitted via a BC transducer. To do this, we simultaneously recorded speech signals using both BC and AC microphones. The intelligibility of both BC and AC speech was then assessed using a two-way Japanese diagnostic rhythm test (JDRT) [5] transmitted via a BC transducer.

The remainder of this paper is organized as follows. The experimental method is described in Section 2. Section 3 evaluates speech intelligibility under BC communication conditions. Finally, the discussion and conclusions are presented in Sections 4 and 5, respectively.

2. Methods.

2.1. **Participants.** A 35-year-old male native Japanese speaker participated in the experiment. Six male native Japanese speakers aged 20 years old participated in the experiments. Before participating in the listening tests, they did not report any known hearing impairments. The experiments were conducted with the approval of the experimental ethics review committee of the National Institute of Technology, Oshima College.

2.2. **Equipment.** The speech uttered by the speaker was recorded simultaneously by a BC microphone (ACO TYPE7827) and an AC microphone (Audio-Technica AT4040) in a quiet soundproof room, as shown in Figure 2. An anaglog/digital converter (Roland Duo-capture EX) transforms the analog signal into a digital signal and inputs this into a personal computer with a sampling rate of 48,000 and a 16 bit quantifying bit rate.

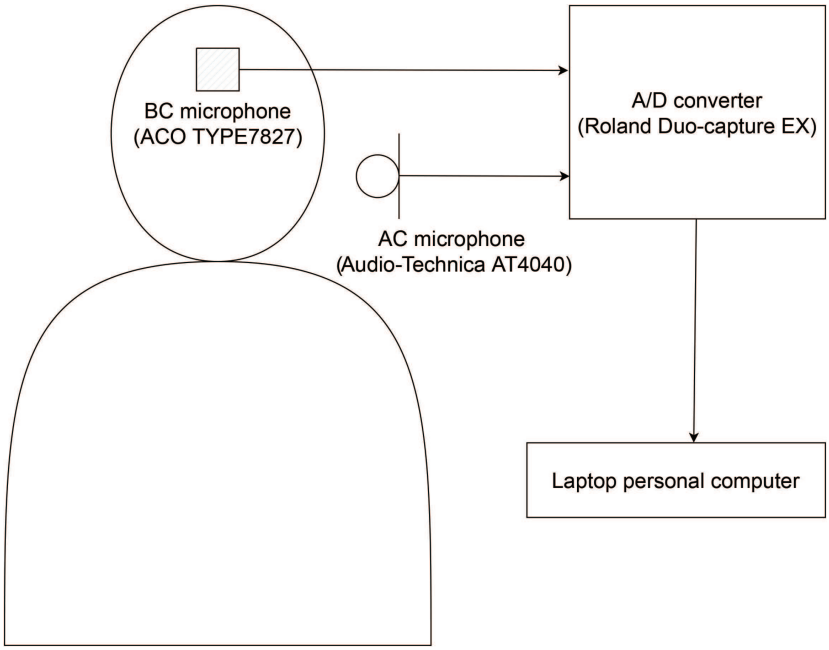


Figure 2. Wiring diagram in the recording

Listeners perceived speech sounds using a BC transducer (Radio Ear B71). A personal computer outputs a speech signal amplified by a digital/analog converter (Roland Duo-capture EX). A loudspeaker (GENELEC 8030C) was located in front of the listener to transmit the masking sound and connected to a personal computer via the converter. The distance between the loudspeaker and listener was 1.8 meters. Figure 3 shows a schematic of the experiment. In the experiment, an AC headphone was not used because we assumed that AC speech communication was impossible under loud conditions.

2.3. **Stimuli.** In this experiment, both the BC and AC speech sources and white noise were presented as stimuli.

AC speech is formed by a series of speech-producing organs, such as the lungs, vocal cords, jaw, tongue, and lips. AC speech primarily consists of sound waves emitted from the mouth into the space. BC speech uses the same series of speech-producing organs. However, BC speech vibrates through the skull and skin. Speech waves are generated by vocal fold vibration and are characterized by the resonant feature of the vocal tract, where turbulent noise changes relatively slowly [4].

Japanese speech is composed of vowels and consonants. Consonants are further subdivided into phonemes that are classified into seven phonetic features: voicing, nasality, sustention, sibilation, graveness, compactness, and vowel-like. Table 1 lists Japanese

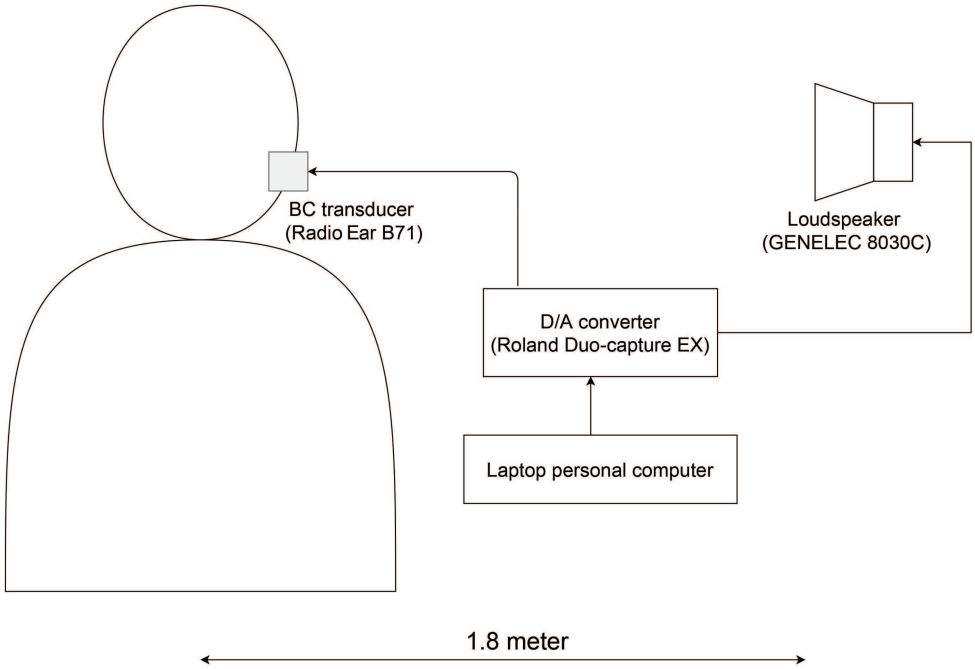


FIGURE 3. Wiring diagram in the listening experiment

TABLE 1. Phonetic features of consonants. For each phoneme, the presence of a specific phonetic feature is indicated by "+", its absence by "−", and neither by "0".

	m	n	z	j̆	b	d	g	w	r	j	ɸ	s	š	č	p	t	k	h	ŋ	ts	ȼ
Voicing	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	+	−	−
Nasality	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
Sustention	−	−	+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	−	−	+
Sibilation	−	−	+	+	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	+	−
Graveness	+	−	−	0	+	−	0	+	−	0	+	−	0	0	+	−	0	0	0	−	0
Compactness	−	−	−	+	−	−	+	−	−	+	−	−	+	+	−	−	+	+	−	−	+
Vowel-like	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−

phonetic symbols. The seven phonetic features were taken from the phonetic feature classification by Jacobson, Fant, and Halle (JFH), and are listed as follows [5, 6].

1) Voicing is a phonemic feature which corresponds to the vocalic-non-vocalic spectrum, which is a relatively clear feature classification.
2) Nasality corresponds to the nasality-orality spectrum, which is a relatively clear feature classification.
3) Sustention corresponds to the continuous-interrupted spectrum, which is a clear classification of sustained sounds and other sounds such as bursts and breaks.
4) Sibilation corresponds to the strident-mellow spectrum. It is a classification of waveform irregularity, where irregular phonemes are classified as strident and regular phonemes are classified as mellows.
5) Graveness corresponds to the grave-acute spectrum. It is classified as sulcus and acute, and is considered to be the former if the energy in the spectrum is concentrated at low frequencies and the latter if it is concentrated at high frequencies. If the volume of the entire oral cavity is large during vocalization, it is considered to be the former, and if

the oral cavity is divided into smaller parts by the tongue, it is considered to be the latter.

6) Compactness corresponds to the compact-diffuse spectrum. If the energy in the spectrum is concentrated in one formant frequency, it is considered as compact, and if it is dispersed, it is considered as diffuse. If the volume of the anterior part of the oral cavity is larger than that of the posterior part across the narrow chain, it is considered as compact, and if it is smaller, it is considered as diffuse.
7) Vowel-like is not an actual classification; however, it is used to distinguish between glide and other consonants.

The loudspeaker emitted white noise at a sound pressure level of 90 dBA as a masking sound at the listening site. The listeners wore earplugs for safety reasons and practiced how to wear them before conducting the experiment.

Before conducting the experiment, the loudness of the BC sound was contoured to the AC sound at a comfortable level to avoid exposure to loud sounds. Using the adjusted vibration level as a reference, the vibration levels varied from 0 dB, −5 dB, to −10 dB.

2.4. **Procedures.** This study employed JDRT [5] to evaluate speech intelligibility. The JDRT is a comprehension test method in which subjects listen to one of a set of candidate word pairs that differ only by one initial phoneme and are asked to choose one of the word pairs. JDRT assumes the following.

1) Disturbances such as additive noise and transmission distortions affect consonants, which have relatively low energy. The vowels were almost unaffected. In addition, consonants play a more important role in speech communication because they carry most of the linguistic information.
2) Even if intelligibility is evaluated using only one phoneme at the beginning of a word, it is possible to measure intelligibility in other positions. In other words, the intelligibility of the initial phoneme and intelligibility of the other positions are similar in response to interference.
3) By limiting the number of candidates selected by the subject to a very small number, the effects of word familiarity, phonemes, and context can be eliminated.

The word list contained 96 sets, with each set containing two minimal pairs of words that differed only at the beginning of the word. In the DRT, the correct response rate of the subjects was evaluated using the six phonetic features or the average of all responses. In addition, the correct response rate was adjusted as follows to eliminate chance:

$$S = \frac{100(R - W)}{T}, \tag{1}$$

where S is the adjusted percentage of correct answers (%), R is the percentage of correct answers, W is the number of incorrect answers, and T is the total number of trials. Thus, because there are two choices even if the answers are completely wrong, $R \simeq W$, and the origin is adjusted so that $S \simeq 0$ at this time.

Subjects were presented with word pairs and asked to choose the correct response. Each test contained 96 words, which were graded six times with two different sources and three different signal levels. The experiment was conducted at the subject's own pace.

3. **Results.** Figure 4 shows the average percentages of correct responses for BC and AC sound sources transmitted via the BC transducer. The result shows that the average percentages of correct responses are approximately 40% and 70% for the BC and the AC sound sources, respectively. No change in intelligibility was obtained when the stimulus level presented by the BC transducer was reduced from a comfortably audible level to −10 dB with a −5 dB step.

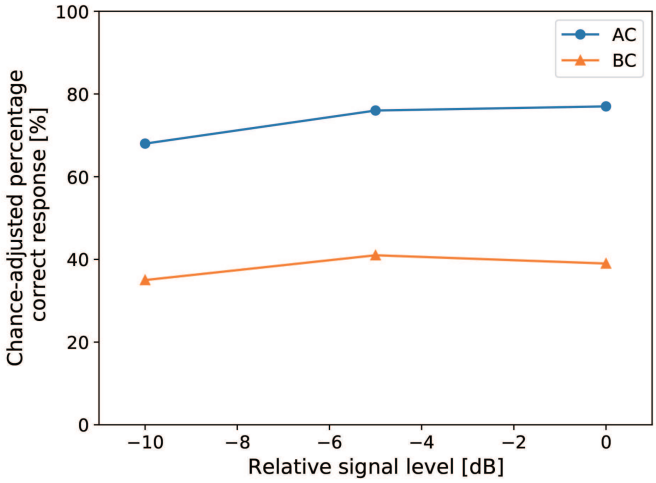


FIGURE 4. Chance-adjusted percentage correct answer rate for words under conditions of AC and BC sources

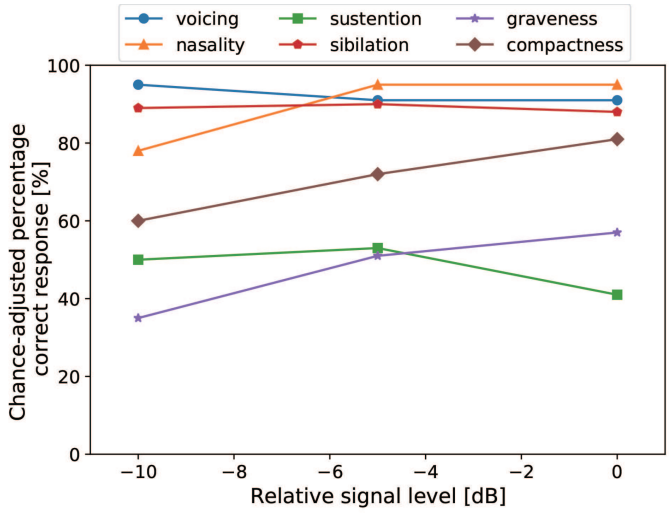
Figure 5(a) shows the average percentages of correct responses for each phoneme feature recorded by the AC microphone. This result shows that the average percentages of correct responses depend on the phonetic feature. Figure 5(b) shows the average percentage of correct responses for each phonetic feature in the BC speech. The results show that the average percentages of correct responses also depend on the phonetic feature for BC speech source.

Comparing the speech intelligibility of AC and BC speech, the average percentages of correct responses for voicing and nasality were high under both conditions. This implies that voicing and nasality are robust phonetic features. The four phonetic features with lower response rates for BC sound than for AC sound were sustainability, sibilation, graveness, and compactness. This means that these features lost phonetic features when using the BC microphone.

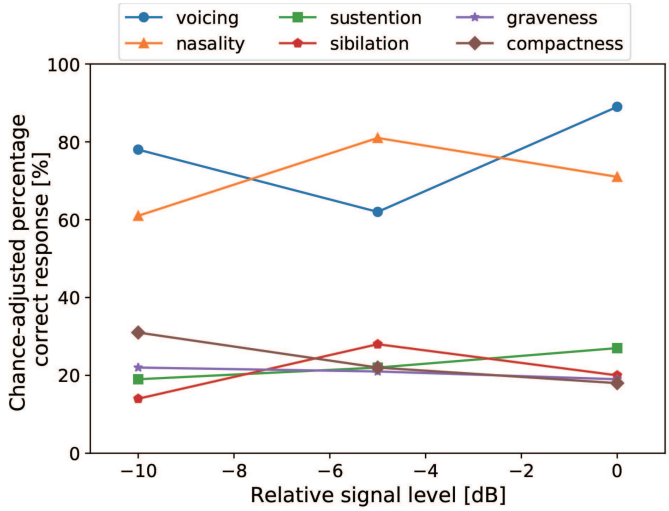
4. **Discussion.** In this study, we investigated the possibility of speech communication via BC sound. Speech intelligibility was measured for the BC-recorded sound and presented by the BC transducer. The results of this study demonstrate the potential of BC speech communication, as shown in Figure 4. However, Figure 5 shows that both the AC and BC transducers have both robust and weak phonemic features. There are a high percentage of correct answers for voicing and nasality and phonetic information is retained. However, other phonetic features, such as sustainability, bibliation, graveness, and compactness have low correct response rates and there is lost phonetic information. This difference was also noticeable.

Previous studies [7] used a speech intelligibility test called the modified rhythm test (MRT) [8] to measure the intelligibility of bone guidance. The MRT is based on a set of 50 words consisting of a minimum of six pairs of words that differ only by one phoneme at the beginning or end of the word. In the MRT, only one phoneme was used. However, the MRT cannot measure speech intelligibility based on phonetic features. Therefore, this study is the first to measure the influence of phonetic features on BC speech communication using both AC and BC transducers simultaneously.

Voicing was categorized as either vocalic or non-vocalic and classified according to whether it was accompanied by vocal fold tremor or not. From the phoneme classification in Table 1, phoneme features other than voicing and nasality are predominantly categorized as non-vocalic sounds. Therefore, BC microphones, which pick up tremors and convert them into sound, were not able to completely record non-vocalic sounds without



(a) AC source



(b) BC source

FIGURE 5. Chance-adjusted percentage correct answer rate depending on phonetic features under conditions of AC and BC sources

vocal fold tremors. Therefore, BC microphones may not be able to completely record non-vocalic sounds without vocal fold tremors. For this reason, the percentages of correct answers for the four phonetic features of sustainability, sibilation, graveness, and compactness were low.

5. **Conclusions.** The aim of this study was to investigate the possibility of communication using both BC microphones and BC headphones in a loud environment. However, it was unclear which factors affect speech intelligibility in the process of BC speech production and perception. In this study, speech recorded with BC and AC microphones was processed by a BC transducer, and speech intelligibility was evaluated using the JDRT.

The results of this study are as follows: the percentages of correct answers were high for voicing and nasality, meaning that information was retained. The percentages of correct answers were low for other phonetic features, meaning that the information was lost. This may have been due to a problem in recording voiceless sounds with a BC microphone.

Based on these results, we conclude that improving the performance of the BC microphone in order for it to pick up non-vocalic sounds is necessary for adequate BC speech communication.

Acknowledgment. This work was supported by JSPS KAKENHI (Grant Number 18K 18081).

REFERENCES

[1] S. Stenfelt and R. L. Goode, Bone-conducted sound: Physiological and clinical aspects, *Otology and Neurotology*, vol.26, no.6, pp.1245-1261, 2005.

[2] A. Hagr, BAHA: Bone-Anchored Hearing Aid, *International Journal of Health Sciences*, vol.1, no.2, pp.265-276, 2007.

[3] S. Ishimitsu, M. Nakayama and Y. Murakami, Study of body-conducted speech recognition for support of maritime engine operation, *Journal of the Japan Institution of Marine Engineering*, vol.39, no.4, pp.263-268, 2004.

[4] K. Kakehi, Speech perception in prelexical processing, *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, vol.22, no.4, pp.659-669, 2015.

[5] K. Kondo, R. Izumi, M. Fujimori, R. Kaga and K. Nakagawa, Two-to-one selection-based Japanese speech intelligibility test, *Journal of the Acoustical Society of Japan*, vol.63, no.4, pp.196-205, 2007.

[6] R. Jacobson, C. G. M. Fant and M. Halle, *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*, MIT Press, Cambridge, 1963.

[7] K. A. Pollard, R. K. Tran and T. Letowski, Effect of vocal and demographic traits on speech intelligibility over bone conduction, *Journal of the Acoustical Society of America*, vol.137, no.4, pp.2060-2069, 2015.

[8] A. House, C. Williams, M. Hecker and K. Kryter, Articulation testing methods: Consonantal differentiation with a closed-response set, *Journal of the Acoustical Society of America*, vol.37, no.1, pp.158-166, 1965.