**4aSCb16. Evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of phonetic features**

Takayuki Kagomiya* and Seiji Nakagawa

*Corresponding author's address: Health Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan, 1-8-31, Midorigaoka, Ikeda, 563-8577, Osaka, Japan, t-kagomiya@aist.go.jp

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 ability of BCUHA to transmit Japanese distinctive features to the recipient. For this purpose, a series of mono-syllable intelligibility experiments was conducted. A series of sequential information transfer analyses (SINFA) were carried out to analyze what kind of articulatory features were well transmitted. Results of the SINFA showed that in vowel perception, ‘openness’ and ‘frontness’ were well transmitted, while in consonant perception, Japanese ‘you-on’ (palatalized sound) feature was well transmitted, however, transmission of other features like articulatory position or manner was limited. These results indicated that the BCUHA has sufficient frequency resolution to transmit vowel information, while some signals are masked by carrier sound. To improve this problem, further investigation and development is required.

Published by the Acoustical Society of America through the American Institute of Physics
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.

Performance of BCUHA regarding transmission of speech sounds was assessed by using mono-syllable articulation tests or word intelligibility tests so far. [1, 3]. Although scores of a mono-syllable articulation test or a word intelligibility test can represent broad aspects of performance of speech sound transmission, however, it is difficult to understand what types of features of speech sounds are well transmitted or not.

In this study, we focused on transmission of phonetic distinctive features. Distinctive features are basic features of phonological analysis which are expressed in terms of the phonetic characteristics of each phoneme. Interpreting the performance of BCUHA using these features makes the results easier for speech and hearing therapists to understand.

In addition, studies in acoustic phonetics have shown the relationship between acoustic features and the perception of speech[4]. Using phonetic features leads to understanding and improvement of the disadvantage of BCUHA regarding transmission of acoustic signals.

For these purposes, a series of hearing experiments and analysis was conducted.

EXPERIMENT

A monosyllable perception test was conducted to determine the confusion patterns in BCUHA hearing.

Stimuli

The “100 Japanese mono-syllable articulation test” contained in the “Familiarity-controlled word lists 2003 (FW03)” [5] was utilized for the hearing experiments. FW03 was developed for assessment of Japanese spoken word intelligibility and articulation score, and widely adopted for assessment of hearing assistance devices in Japan. FW03 contains 4 speakers utterances (2 males and 2 females). Considering deference of intelligibility among the speakers [6], we selected utterances of a female (Speaker FTO) whose utterence has the best intelligibility.

Each mono-syllable contained in the test had a V, CV or CCV (“you-on”) structure. Most of Japanese mora are formed by a single vowel or CV structure. In addition to them, Japanese language has “you-on” (contracted sounds or palatised sounds). A you-on is a single mora, however, in which /j/ is added between consonant and vowel, for example, /ka/ v.s. /kja/. In this research, we treat a “you-on” as a single mora formed by combined consonants and vowel: analyzed into Cj and V.

Participants

Nine native Japanese speakers (6 males and 3 females) with no reported speech or hearing defects participated in the experiments. Their ages were ranged from 22 to 40 years old.
Presentation of the sounds

The stimuli presented 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:

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

where \( S(t) \) is the speech signal, \( S_{\text{min}} \) is the minimum amplitude of \( S(t) \), and \( f_c \) is the carrier frequency (30 kHz).

Generally, two types of sound are perceived by the DSB-TC modulated bone-conducted ultrasound: one is a high-pitched tone due to the ultrasonic carrier, with a pitch approximately corresponding to a 10 kHz AC sinusoid [6, 7], and the other is the envelope of the modulated signal (Figure 1) [3, 8].

The stimuli 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 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 2).

Procedures

The mono-syllable articulation test was carried out in two sessions with different amplitude modulation methods. The order of each modulation types of stimulus was counterbalanced. In each session, the stimulus was presented in random order with 5.0 second inter-stimulus intervals.

Participants were requested to listen to the presented stimuli and during each inter-stimulus interval, to write down on an answer form the mono-syllable that they perceived. The sound levels of the stimuli were adjusted to the most comfortable and clear levels for each participant.

All experiments were conducted in a soundproof chamber. The stimuli were played using a FireWire-based audio interface (Echo Audiofire12) attached to a personal computer.
**Table 1:** Confusion matrix of vowels

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>e</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.991</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>i</td>
<td>0.000</td>
<td>0.879</td>
<td>0.111</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>u</td>
<td>0.005</td>
<td>0.072</td>
<td>0.894</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>e</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>0.926</td>
<td>0.065</td>
</tr>
<tr>
<td>o</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Correct perceived ratio: 0.949

**Table 2:** Distinctive features for analysis of vowels

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>e</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>openness</td>
<td>open</td>
<td>close</td>
<td>close</td>
<td>mid</td>
<td>mid</td>
</tr>
<tr>
<td>frontness</td>
<td>center</td>
<td>front</td>
<td>back</td>
<td>front</td>
<td>back</td>
</tr>
</tbody>
</table>

**Table 3:** Summary of the SINFA for response to vowels

<table>
<thead>
<tr>
<th>features</th>
<th>Trans. info</th>
<th>% Trans. info</th>
</tr>
</thead>
<tbody>
<tr>
<td>openness</td>
<td>1.466</td>
<td>0.935</td>
</tr>
<tr>
<td>frontness</td>
<td>0.470</td>
<td>0.702</td>
</tr>
<tr>
<td>Total bits sent</td>
<td>1.944</td>
<td></td>
</tr>
<tr>
<td>Total bits received</td>
<td>1.936</td>
<td></td>
</tr>
<tr>
<td>% info received</td>
<td></td>
<td>0.996</td>
</tr>
</tbody>
</table>

**Analyses and Results**

All participants data were pooled. Overall correct perceived ratio was 0.434. To analyze transmission and perception of each phonetic feature, analyses were conducted separately for the vowel and consonant in each syllable.

**Transmission of Vowel Features**

Table 1 shows confusion matrix of vowels; the presented stimuli are the column headings and the participants’ choices are the row headings. Correct perceived ratio of vowels was 0.949. These results indicate that most vowel information was correctly transmitted. However, a small number of confusions between /i/ and /u/ were also observed.

To analyze what types of phonetic features were transmitted or not, we conducted a sequential information analysis (SINFA) [9]. In this analysis, the sequence in which the information transmission per phonetic feature was analyzed influences the outcome. Phonetic features applied in the analysis were “openness” and “frontness” (Table 2).

Table 3 shows summary of the SINFA for vowels. “% Trans. info” represents the ratio of per phonetic feature were transferred. Also “% info received” represents the proportion of the stimulus-response patterns that is accounted for by the chosen set of features. Thus a “% info received” provide a measure of goodness of the model.

From the results of SINFA, 99.6% of information transfer was accounted for by “openness” and “frontness.” Considering “% Trans. info,” “openness” was transferred at 93.5% level, and “frontness” was transferred at 70.2% level.
T. Kagomiya and S. Nakagawa

Transmission of Consonant Features

Table 4 shows confusion matrix of consonant perception. “φ” represents the proportion of listeners who responded that particular stimuli were vowel-only mora and no consonants were perceived. Overall correct ratio for consonant perception was 0.443.

Also a SINFA was conducted for analysis of consonant feature transmission. Table 5 shows feature matrix applied in this analysis. Features applied in the SINFA were “voicing,” “nasality,” “articulatory manner,” “articulatory position” and “you-on.” The result of the SINFA are shown in Table 6.

According to the results of the SINFA, 76.8% of information transfer was accounted by the five features. “You-on” information was well transferred at 71.2% level. On the other hand, the listeners had difficulty in perception of the other features. “Nasality” was transmitted only 48.2% level, “voicing” and “manner” were limited to 31.9% and 31.4% respectively, and “position” was lower than 27% level.

DISCUSSION

The results showed that the BCUHA can transfer vowel information. The “openess” feature was almost perfectly transmitted at 93.5 % level and the “frontness” was at 70.2 % level. As is well known,
“openess” and “frontness” of vowels were related to F1 and F2 values [4] respectively. This result indicates the BCUHA has sufficient frequency resolution to transmit vowel formant information.

On the other hand, transmission of consonant features was limited. Transfer ratio of voicing information was limited at 31.9 % level. Also transmission of articulatory position information was limited at 26.8 % and articulatory manner was 31.4 % level. It is known that major acoustic cue for voice/unvoice is voice onset time (VOT), cues for nasal/stop is nasal murmur and articulatory position is formant transition in a succeeding vowel [4]. The results of the experiment indicate that the listeners had difficulty in perceiving these short-duration or low-intensity sound events.

This reduction of accuracy in consonant feature perception is able to accounted for by the existence of a carrier sound accompanying the DSB-TC amplitude modulation. With the DSB-TC modulation, the speech sounds are accompanied by a relatively strong high-pitched tone due to the carrier sound. It is possible that the carrier sounds masked speech signals in part [6] and the listeners had difficulty to perceive the acoustic cues of consonant discrimination.

On the contrary, “you-on” is realised as short vowel /i/ inserted between consonant and following vowel, which is relatively clear and long duration event. moreover, vowels have high intelligibility as mentioned above. High transfer ratio in “you-on” feature was attributable to these factors.

## Conclusion

The usability of BCUHA regarding transfer of phonetic features was evaluated. For this purpose a mono-syllable articulation test was conducted. Results were that: vowel features including “openess” and “frontness” were well transmited, while transmission of consonant features were limited.

These results indicated that the BCUHA has sufficient frequency resolution to transmit vowel information, while some signals are masked by carrier sound. To improve this problem, further investigation and development is needed.

## Acknowledgments

This study was supported by the Grants-in-Aid for Scientific Research of Japan Society for the Promotion of Science (24500675) and the Funding Program for Next-Generation World-Leading Researchers provided by the Cabinet Office, Government of Japan.

## References


