Development of a novel hearing-aid for the profoundly deaf using bone-conducted ultrasonic perception: Assessments of the modulation type with regard to articulation, intelligibility, and sound quality

Seiji Nakagawa*, Chika Fujiyuki, Yuko Okubo, Takuya Hotehama and Takayuki Kagomiya

*Corresponding author’s address: Health Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-8-31 Midorigaoka, Ikeda, 563-8577, Osaka, Japan, s-nakagawa@aist.go.jp

Bone-conducted ultrasound (BCU) is perceived even by the profoundly sensorineural deaf. We have developed a novel hearing-aid using BCU perception (BCU hearing aid: BCUHA) for the profoundly deaf. In the BCUHA, ultrasonic sinusoids of about 30 kHz are amplitude-modulated by speech and presented to the mastoid. Generally, two sounds are perceived: one is a high-pitched tone due to the ultrasonic carrier, with a pitch corresponding to a 8-16 kHz air-conducted (AC) sinusoid, and the other is the envelope of the modulated signal. As a method of amplitude modulation (AM), double-sideband with transmitted carrier (DSB-TC) modulation had been used, however, the DSB-TC modulation is accompanied by a strong high-pitched tone. In this study, two new AM methods, double-sideband with suppressed carrier (DSB-SC) and transposed modulations, that can be expected to reduce the high-pitched tone were newly employed in the BCUHA, and their resulting articulations, intelligibilities and sound qualities were evaluated. The results showed that DSB-TC and transposed modulation had higher articulation and/or intelligibility scores than DSB-SC modulation. Further, in terms of sound quality, the transposed speech was closer than other types of BCU speech to AC speech. These results provide useful information for further development of the BCUHA.
INTRODUCTION

Although the upper frequency limit of human hearing is believed to be no higher than about 20,000 Hz, several studies have reported that high-frequency sound up to at least 100,000 Hz can be heard via bone conduction. (Gavreau, 1948; Plumphry, 1951; Bellucci, & Schneider, 1962). This “audible” ultrasound through bone conduction is referred to as bone-conducted ultrasound (BCU). Moreover, BCU hearing in humans has been found under various auditory pathological conditions, including sensorineural hearing loss and middle-ear disorders (Bellucci, & Schneider, 1962). In particular, BCU is even perceived by profoundly deaf subjects, who cannot obtain sufficient audition even with the use of a conventional hearing aid.

The mechanisms of BCU perception remain unclear; however, recent electrophysiological measurements in humans have revealed the neural pathway of BCU. The action potential (AP) of an electrocochleogram, which reflects the compound action potential of the auditory nerve, was observed clearly (S. Nakagawa & A. Nakagawa, 2006). Also, substantial auditory brainstem responses (ABRs) and middle latency responses (MLRs), which reflect the activation of the auditory pathway from the brainstem to the auditory cortex, are evoked by BCU (S. Nakagawa & A. Nakagawa, 2006; Nakagawa, 2009). These findings indicate that BCU evokes the same auditory pathway as audible-frequency sounds through the cochlear nerve, and there is no special organ for BCU perception. Additionally, recent studies have shown several characteristics of BCU perception that suggest unique perception mechanisms, although the cochlea contributes substantially to BCU perception. The fact that BCU strongly masks 10–15-kHz air-conducted (AC) sounds indicates the contribution of the cochlea to BCU perception (Nishimura et al., 2003). On the other hand, the pitch of sinusoidal BCU is approximately constant at around the pitch induced by AC pure tones at 8-16 kHz and independent of its frequency (Phumphry, 1951; Dierof & Ertel, 1975; Nakagawa & Tonoike, 2005; Ito & Nakagawa, 2008). Additionally, the dynamic range of loudness, i.e., the difference between the threshold and the uncomfortable level, is extremely narrow at less than 20 dB (Nishimura et al., 2003). These findings suggest that BCU is received in the cochlea through a special mechanism that is different from that of audible-frequency sound perception: the inadequate vibration of the basilar membrane and little contribution of the outer hair cells to BCU perception (Nakagawa, 2009). Meanwhile, some reports suggest the possibility of transforming BCU into low-frequency audible sound (Torndorf, 1966; Dobie, & Wiederhold, 1992). However, this hypothesis cannot explain the brain activities of the profoundly deaf subjects (Lenhardt, et al., 1991; Hosoi et al., 1988; Nakagawa et al., 2007), because they have little sensitivity to audible sound below 20 kHz. Furthermore, evidence for the generation of audible-frequency components has not been observed in physioacoustical measurements. No audible-frequency signals corresponding to the subjective pitch of a BCU tone have been found in the acoustic fields for the auditory meatus and vibrations of the tympanic membrane (Ito & Nakagawa, 2010), and the occurrence of nonlinear behavior in the transmission in the human head has not been observed (Ito & Nakagawa, 2011).

In 1991, Lenhardt et al. reported that BCU amplitude-modulated by speech sounds was intelligible to some extent and suggested the possibility of developing novel hearing aids based on BCU perception (Lenhardt, et al., 1991). We supported the arguments of Lenhardt et al. objectively with magnetoencephalographic findings (Hosoi et al., 1988; Nakagawa et al., 2000) and have developed a new hearing aid for the profoundly deaf—the bone-conducted ultrasonic hearing aid (BCUHA, Fig. 1) (Nakagawa et al., 2006). In the BCUHA, ultrasonic sinusoids with a frequency of about 30 kHz are amplitude-modulated by speech or environmental sounds detected by microphones and presented to the mastoid by a vibrator. Generally, two sounds are perceived by the BCUHA: one is a high-pitched tone due to the ultrasonic carrier, with a pitch corresponding to a 8–16 kHz AC sinusoid (Phumphry, 1951; Dierof & Ertel, 1975; Nakagawa & Tonoike, 2005; Ito & Nakagawa, 2008) and the other is the envelope of
the modulated signal (Fujimoto & Nakagawa, 2005; Okamoto et al., 2005). As a method of amplitude modulation (AM), double-sideband with transmitted carrier (DSB-TC) modulation was used.

The BCUHA is the first device to be developed that enables the profoundly deaf to sense sufficient audition without surgery, and it is far easier to attach than a cochlear implant. In hearing tests on profoundly deaf subjects using this prototype, more than 40% of profoundly deaf subjects were able to perceive some sounds and 17% were able to recognize words.18) The results already obtained have demonstrated the potential of the BCUHA; however, further development and improvements are needed to use the device in practice. In particular, there is room for improvement in terms of articulation and sound quality. The high-pitched tone due to the ultrasonic carrier is appears to be a key factor in the degradation of the articulation and sound quality of the BCUHA. It is thought that the high-pitched tone increases the discomfort of the speech. Furthermore, a previous report suggested the possibility that the high-pitched tone due to the carrier decreased articulation (Okamoto et al., 2005). Applications of new AM methods are thought to be effective for reducing the high-pitched tone. In this study, two new AM methods, double-sideband with suppressed carrier (DSB-SC) and transposed modulation (Bernstein & Trahiotis, 2002), that can be expected to reduce the high-pitched tone were newly employed in the BCUHA, and their resulting, intelligibilities and sound qualities were evaluated. Additionally, mono-syllable articulation test was conducted for DSB-TC and transposed modulations and their confusion matrices were compared.

METHODS

BCU speech using several AM methods and AC speech were presented to subjects, and word intelligibility, mono-syllable articulation, and the subjective impression of the sound was examined. Necessary information regarding the experiment was given to the subjects, and informed consent was obtained from each subject prior to the experiment. The experiment was approved by the Institutional Review Board on Ergonomic Research of AIST.

Methods of Amplitude Modulation

In this study, double-sideband with suppressed carrier (DSB-SC) modulation and transposed modulation were newly employed, in addition to the DSB-TC modulation. These AM methods are expressed in terms of A, s(t), f_c(t), and m, which represent a constant, the modulator signal (speech sound), a carrier signal, and the modulation depth, respectively. Examples of waveforms and spectra obtained by these AM methods are shown in Figs. 2.

I) Double-sideband with Transmitted Carrier (DSB-TC)

\[ f(t) = A(1 + m \times s(t)) \times f_c(t) \]  

In the DSB-TC modulation, the envelope of the modulated signal corresponds to the modulator, i.e., speech sounds. On the other hand, since the DSB-TC modulation procedure has a substantial peak of power at the carrier frequency in the frequency domain, it is accompanied by a strong high-pitched tone, especially when the modulation depth is low.

II) Double-sideband with Suppressed Carrier (DSB-SC)

\[ f(t) = A(m \times s(t)) \times f_c(t) \]  

In the DSB-SC modulation, the peak of power at the carrier frequency is suppressed. Therefore, it is thought that this method has the advantages of not only power saving, but also the reduction of the high-pitched tone. However, the envelope of the modulated signal does not correspond to the modulator signal. The pitch accompanied by the envelope is almost twice as high as that of the modulator signal19) and is thought to contain some distortion.

III) Transposed Modulation

\[ f(t) = A(m \times s_{tp}(t)) \times f_c(t) \]  

Here \(s_{tp}(t)\) represents a half-wave-rectified and low-pass-filtered modulator signal. In this study, the modulator signal was low-pass filtered at 8 kHz, to avoid impairing the speech information. In this method, since the peak of power at the carrier frequency is suppressed, it is anticipated that the high-pitched tone due to the carrier is reduced.
Furthermore, the intervals between peaks of the envelope of the modulated signal are the same as the intervals between the peaks of the modulator signal, thus; the pitch due to the envelope is expected to be similar to that of the modulator signal. On the other hand, it is thought that transposed BCU speech essentially contains some distortion because of the rectification process.

**Apparatus**

All types of BCU speech were presented to one of the two mastoid portions of the subject’s temporal bone by a piezoelectric ceramic vibrator (Murata Manufacturing MA40E7S) with a newly devised plastic housing. The vibrator was fixed by a headset with a clamping pressure of approximately 5 N (see Fig. 1). The stimulation side was decided by subject preference. BCU speech was generated according to a personal computer at a sampling frequency of 96 kHz and fed to the vibrator via a 24-bit digital-to-analog converter (Echo Digital Audio AudioFire 12), an amplifier (Mess-Tek M-2629B), and a programmable attenuator (Tucker-Davis Technologies PA-5). Before the session, each subject was requested to confirm that no AC sounds were sensed via the vibrator.

Additionally, AC speech, i.e., the modulator signal itself, was presented to the same side as the BCU stimuli for comparison. AC speech was also generated by a personal computer at a sampling frequency of 48 kHz, and fed to headphones (Koss R/80) via the same digital-to-analog converter and programmable attenuator as those used for BCU speech. All experiments were carried out in a soundproof room.

**WORD INTELLIGIBILITY TEST**

**Methods**

31 normal-hearing Japanese subjects (21–41 years old, mean 24.2 ± 4.9 years old) participated. Four-mora Japanese words recorded with a female voice and a male voice were taken from a commercially available database.
(NTT-AT FW03) in which Japanese words are classified into four levels of ‘familiarity’ to control the degree of word difficulty: familiarity 7.0–5.5, familiarity 5.5–4.0, familiarity 4.0–2.5, and familiarity 2.5–1.0. Here, ‘familiarity’ represents the degree of people’s familiarity with each word. In this study, words were selected from mid-range familiarity groups (5.5–4.0 and 4.0–2.5) to avoid ceiling and floor effects.

Three types of BCU stimuli were produced using the DSB-TC, the DSB-SC, and the transposed modulation. Air-conducted stimulus was also used. The test was carried out in four sessions with different stimulus types. The order of each type of stimulus was counterbalanced. In each session, 25 words were selected from each group of familiarity (5.5–4.0 and 4.0–2.5) for both female and male speakers; thus, 100 words were used in a session.

Before the intelligibility test, the threshold of each subject was determined by the method of limits with a 1-kHz sinusoid. The average sound level of AC speech was set at 20 dB greater than the threshold. The intensities of BCU speech were set at the most clearly perceived level for each subject by the method of adjustment. The interstimulus interval was 7.0 s. Subjects were instructed to write down what they had heard on a prepared answer sheet during the interval. Before the session, a training session was conducted using 50 words, that were not used in the experiment.

Results

Figure 3 shows the results of the word intelligibility test for all subjects. The mean intelligibility scores were 45.1% for DSB-TC speech, 18.9% for DSB-SC speech, 48.2% for transposed speech, and 93.4% for AC speech. Two-way analysis of variance (ANOVA) showed the effects of the stimulus type (p < 0.001) and speaker’s gender (p < 0.001). All types of BCU speech have lower intelligibility than AC speech (p < 0.001), and DSB-TC speech and transposed speech have higher intelligibility than DSB-SC speech (p < 0.001). For each type of BCU speech, the intelligibility for the female voice is higher than that for the male voice (DSB-TC, transposed: p < 0.001; DSB-SC: p < 0.05). For all types of BCU speech, the intelligibility for the experienced group was significantly higher than that for the inexperienced group (DSB-TC: p < 0.05, DSB-SC: p < 0.05, transposed p < 0.01).

EVALUATION OF SOUND QUALITY

Methods

After each session of the word intelligibility test, the subjective impression of the sound was recorded in a questionnaire. Appropriate questions were selected from a questionnaire for listening by hearing aids (Okamoto et al., 2002), which is designed to assess the subjective impression of listening with hearing aids. Subjects were asked to rate the following on three-point scales.

1) Loudness: loud—not loud
2) Distortion: distorted—not distorted
3) Indistinctness: indistinct—not indistinct
4) Shrillness: shrill—not shrill

FIGURE 3. Scores of intelligibility for each stimulus type (mean + standard deviation).
Results

Figure 4 shows the result of the questionnaire on listening by the hearing aid. One-way ANOVA showed the effect of the stimulus type on the answers to all questions (p < 0.001). A post-hoc test (Tukey’s HSD) revealed that the DSB-SC stimulus was less clear than the other BCU stimuli and more distorted than the DSB-TC stimulus (p < 0.001). The transposed stimulus was more pleasant than the other BCU stimuli (p < 0.005) and less shrill than the DSB-TC stimulus (p < 0.05). For all types of BCU speech, no significant differences were observed between the experienced and inexperienced groups in the answers to all questions. In terms of the differences between BCU and AC speech, all BCU stimuli were louder, more distorted, shriller, less pleasant, and less clear than the AC stimulus (p < 0.001).

MONOSYLLABLE ARTICULATION TEST

Methods

To investigate perception characteristics of the newly employed transposed modulation in more detail, monosyllable articulation for was investigated and compared with that of DSB-TC modulation.

11 normal-hearing Japanese subjects (22–40 years old, mean 28.1 ± 6.4 years old) participated. 100 Japanese monosyllable recorded with a female voice were taken from a commercially available database (NTT-AT FW03). Before the test, the threshold of each subject was determined by the method of limits with a 1-kHz sinusoid. The intensities of both BCU speech were set at the most clearly perceived level for each subject by the method of adjustment. The interstimulus interval was 5.0 s. Subjects were instructed to write down what they had heard on a prepared answer sheet during the interval. Confusion matrices were obtained using the results of the articulation tests.

Results

The scores of the articulation of DSB-TC and transposed modulation were 43.5% and 35.2%, respectively. The transposed modulation tended to show lower articulation, however, significant difference was observed between the both scores.

Figure 5 shows the respective confusion matrices for DSB-TC and transposed modulations. Some differences between DSB-TC and transposed modulations; two-way ANOVA regarding the number of misidentified phonemes with spoken phonemes and sound types showed that the number perceived as /j/ was significantly larger for the transposed modulation than for DSB-TC modulation when voiced consonants were presented (p<0.05). Additionally,
the number perceived as vowels was significantly larger for the transposed modulation than for DSB-TC modulation when consonants were presented (p<0.05).

**DISCUSSION**

**Monosyllable articulation test and word intelligibility**

**I) Differences between BCUs and AC**

There are clear differences between intelligibility of BCU and AC speech. It is thought that speech perception via amplitude-modulated BCU depends on the demodulation due to the non-linearity of the inner ear mechanisms or conduction pathway. Therefore, not only the shape of the envelope but also the precision of demodulation affects speech perception. Since the shape of the envelope is identical to the original speech signal in the DSB-TC speech, the precision of the demodulation can be considered as dominant factor causing the difference between AC and BCU speech.

**II) Differences among the modulation types of BCU**

In terms of the differences among the modulation types of BCU, DSB-SC speech was less intelligible than other BCU speech. In the DSB-SC speech, the high-pitched tone due to the carrier was reduced, however, the pitch of the sound due to the envelope was twice as high as that of the modulator signal. Additionally, the sound due to the envelope contained some distortion. These changes in the shape of the envelope can be considered as the causes of the lower intelligibility of the DSB-SC speech.

On the other hand, no significant difference was observed between intelligibility and articulation of DSB-TC and transposed speech, whereas the transposed modulation tended to show lower articulation. In the DSB-TC and the transposed speech, the pitch of the sound due to the envelope is almost identical to that of the modulator signal, whereas the DSB-TC speech is accompanied by a relatively strong high-pitched tone due to the carrier. Nishimura et al. reported that BCU masked 10–14 kHz AC sounds significantly but did not mask AC sound below 8 kHz. Therefore, it is reasonable to consider that the high-pitched tone due to the carrier did not disturb the perception of speech, which consists of only lower-frequency components below 8 kHz in the DSB-TC speech. It is also indicated that BCU speech is intelligible to some extent if the pitch of the sound due to the envelope is maintained.

Some significant differences were observed in confusion matrices for DSB-TC and transposed modulations.

**FIGURE 3.** Confusion matrices based on the results of monosyllable-articulation tests with DSB-TC and the transposed CBU. The phonemes were classified into four groups: (1) vowels, (2) unvoiced consonants, (3) voiced plosive and fricative consonants, and (4) other voiced consonants. Blocks with larger grey values indicate higher appearance frequencies for those pairs.
Generally, the results indicated that transposed sound contains more distortions than DSB-TC sound. It is considered that such distortions in transposed sounds result from the rectification process.

**Sound quality**

Differences among the AM methods for BCU speech were observed. Generally, the sound quality of the transposed speech is closer than that of other AM methods to the sound quality of AC speech. In particular, the transposed speech was not only less shrill but also more pleasant than DSB-TC speech. Since the score for pleasantness involves an integrated evaluation of speech perception, this result suggested a comprehensive advantage of the transposed modulation.

On the other hand, the DSB-TC speech has a significantly higher score for shrillness than the other stimulus types, indicating the effects of the strong high-pitched tone accompanying the DSB-TC speech. Furthermore, it is thought that the higher score for distortion of the DSB-SC speech and the transposed speech reflect distortions in the shape of the envelope resulting from these AM methods. Furthermore, the DSB-SC speech generally has a worse score than the speech for the other AM methods except for shrillness. In particular, the DSB-SC speech was significantly less clear than the speech for the other AM methods. Although it is difficult to conclude the mechanism of this lower evaluation of clarity, it is thought that the score for clarity is strongly affected by the distortion of the envelope.

The differences between BCU and AC speech were obvious. The high-pitched tone due to the carrier of BCU speech appeared to increase the scores of shrillness and loudness. Furthermore, distortions contained in the BCU speech itself and produced in the demodulation process appeared to result in higher scores for distortion and indistinctness and a lower score for clarity. Since the scores for pleasantness involves an integrated evaluations of speech perception, it is considered that both the high-pitched tone and the distortions of BCU speech reduced it.

**CONCLUSION**

In this study, two new AM methods, double-sideband with suppressed carrier (DSB-SC) and transposed modulation, that can be expected to reduce the high-pitched tone were newly employed in the BCUHA, and their resulting intelligibility, sound quality, and articulation were evaluated. The results showed that DSB-TC and TM had higher intelligibility scores than DSB-SC. Transposed modulation tends to contain more distortion than DSB-TC modulation. Further, in terms of sound quality, the transposed speech was closer than other types of BCU speech to AC speech. These results provide useful information for further development of the BCUHA.

**ACKNOWLEDGMENTS**

This research was partially supported by the Funding Program for Next-Generation World-Leading Researchers provided by the Cabinet Office, Government of Japan, and a Grant-in-Aid for Scientific Research (22680038) from the Japan Society for the Promotion of Science for SN.

**REFERENCES**


