Title	子音強調による骨導デバイスの音声了解度の改善
Author(s)	朱, 雯雨
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17634
Rights	
Description	Supervisor:鵜木　祐史, 先端科学技術研究科, 修士（情報科学）

Improvement of speech intelligibility of bone conduction devices by consonant emphasis

2010087 ZHU WENYU

In recent years, computers have become an integral part of human life. They have evolved from the first large machines to the current cell phones and even smart watches. The size of computers is getting smaller and smaller, making them more convenient to be carried around. There are small computers such as smart watches and other wearable computers, sensors, and information presentation devices, and various efforts are being made. From these efforts, it can be seen that wearable computers (wearable devices) are becoming smaller than conventional computers. Therefore, wearable devices are considered to be an inevitable step in the development of computers.

It has been pointed out that there are two types of wearable devices: information wearable devices and environmental wearable devices. Among them, information wearable devices correspond to the five senses such as ears, eyes, etc. Bone conducted device is an information wearable device for the ears. Bone-conductive device is a device that can transmit sound by vibration of the skull or skin without blocking the ear. Since the bone-conductive device transmits sound by bone vibration, it has an advantage. That is users use bone-conducted device can hear the sound presented by bone vibration while air-conducted sound is heard. It is thought bone-conducted device can be useful for safe and secure communication. In addition to providing convenience and safety to people's lives, bone guidance devices can be effective in supporting the physically challenged and expanding the range of activities in their lives. Bone conduction devices can be used to assist people with conductive hearing loss who have problems in the ear canal or middle ear and have difficulty in hearing sounds by air-conducted headphone, or people with hearing impairment who have hematoma in ears, a condition in which the ear becomes blocked.

However, there is one problem when using bone conduction device. When speech is presented with a bone-conduction device, the sound quality and speech intelligibility are reduced compared to air-conduction speech. In particular, it was pointed out that the intelligibility of bone-conducted speech was reduced in noisy environment. There are two possible causes that may affect the intelligibility of bone-conducted speech. The first is that when users use a bone-conducted device, users can hear bone-conducted without covering their ears, then they can hear background sounds in the surrounding environment at the same time. In an environment with high levels of noise, the heard noise may affect the intelligibility of the bone-conducted speech.

Second, the high-frequency component of bone-conducted speech is attenuated by the characteristic of bone-conducted transmission. The characteristic of bone-conducted transmission is that the high-frequency component of bone-conducted speech is attenuated compared to air-conducted speech. If the speech intelligibility is reduced, there is a possibility that a discrepancy will occur when using bone-conducted device. This may cause a serious problem when communicating important instructions in the wearable information sharing space.

It is thought that the high-frequency attenuated of bone conduction affects the intelligibility of bone-conducted speech. Fujita focused on the high-frequency attenuated characteristic in the bone-conducted transmission and proposed a method to improve the intelligibility of bone-conducted speech.The method using high-frequency enhancement to compensate for the bone conduction transmission characteristic. This method improved the intelligibility of bone-conducted speech in noisy environment. On the other hand, consonants are concentrated at higher frequencies than vowels. So the high-frequency attenuated characteristic of bone-conduction may also affect the intelligibility of consonants. If we add processing to the consonants, we can expect further improvement in intelligibility.

The purpose of this paper is to improve the intelligibility of bone-conducted speech in noisy environment.

There are two important domains for speech: frequency domain and time domain. Since the method proposed by Fujita focused only on the frequency domain, it is expected to improve the intelligibility of bone-guided speech by focusing on the time domain as well. Focusing on the overall temporal structure of speech signals, Drullman pointed out that the temporal spectrum is important for speech intelligibility in the time domain. It was found that the intelligibility of bone-guided speech can be restored based on the modulation transfer function focusing on the time spectrum. The temporal structure of the speech signal leads to the phonological levels of consonants and vowels. The acoustic features of consonants are more complex than those of vowels in speech. Many consonants are concentrated in the high frequency range. In addition, the low power of consonants compared to vowels suggests that consonants are easily masked in noisy environments. Furthermore, the maintenance time of consonants is shorter than that of vowels. An important component to the perception of consonants is the formant transition from consonant to vowel. Due to the acoustic properties of consonants, it is thought that consonants are harder to identify than vowels in airborne speech. Hoshino proposed a method of consonant enhancement by emphasizing the consonant and formant transitions. Yasutake and Nakajima showed that consonant stress improved the intelligibility and comprehensibility of speech presented by loudspeakers in noisy environment.

Bone-conduction transmits sound by vibration of the skin and skull. The vibration of the skin and skull also transmits to the outer and middle ear. And the outer ear vibrates, the process of sound transmission from the outer ear to the middle ear and the characteristic of air-conducted sound are similar. Therefore, it is thought that consonant enhancement is effective in air-conducted sounds and will also effective in the bone conduction pathway. Therefore, the aim of this study is to improve the intelligibility of speech presented with a bone guide in a noisy environment by effectively emphasizing consonants on the time and frequency axes.

In this study, two methods were proposed that focus on the time domain. Consonant emphasize and modulation transfer function improvement methods were proposed. Among them, consonant emphasize was found to be able to improve speech intelligibility in noisy environments.

The proposed method in this study is to emphasize the consonants concentrated in the high frequency region which are easily affected by the bone conduction high-frequency attenuated characteristic. Other than that, the maintenance time of consonants was short, and it was difficult to obtain sufficient improvement effect only by emphasizing consonants. The formant transitions, which are important for the perception of consonants, should be emphasized together.

For improving the bone-conducted speech intelligibility well, the combination of the high-frequency emphasis proposed by Fujita and the consonant emphasis(CE) is proposed. Therefore,the first-order high-frequency emphasis (FOE) compensates for the transfer characteristic of regio temporalis(RT) vibration,which improved intelligibility well has been combined with consonant enhancement(RT-FOE+CE).

In order to confirm the improvement effect of the proposed method on bone-conducted speech intelligibility, speech intelligibility tests were conducted in noisy environment(55 dB, 75 dB).According the results of tests, in noisy environment CE has significant difference in word correctness with No emphasis. And there was no significant difference between CE and RT-FOE. This indicates that CE has the same level of comprehension improvement as RT-FOE in Fujita. In addition, RT-FOE+CE significantly increased the correct response rate compared to CE and RT-FOE alone. This indicates that the RT-FOE+CE is most effective in improving the intelligibility of bone-conducted speech in noisy environment.