Acoust. Sci. & Tech. 36, 6 (2015)

PAPER

Evaluation of tooth-conduction microphone for communication under noisy environment

Yusuke Torikai[1,*], Dai Kuze[1,†], Junko Kurosawa[2,‡], Yasuhiro Oikawa[1,§] and Yoshio Yamasaki[2,¶]

[1]*Department of Intermedia Art and Science, Waseda University, 3–4–1 Ohkubo, Shinjuku-ku, Tokyo, 169–8555 Japan*
[2]*Comprehensive Research Organization, Waseda University, 3–4–1 Ohkubo, Shinjuku-ku, Tokyo, 169–8555 Japan*

(*Received 30 January 2015, Accepted for publication 21 July 2015*)

Abstract: We investigated a new communication-aid system focused on bone-conduction through a tooth, for listening to and recording voices. In this paper, we developed a tooth-conduction microphone (TCM) and evaluate the articulation of tooth-conducted voice (TCV). Because the TCM has the shape of one's dental mold, it is wearable like a mouthpiece. Moreover, it can extract tooth vibration during phonation as TCV. To evaluate articulation of TCV, we adopted monosyllable articulation for subjective assessment and linear predictive coding cepstral distance for objective assessment. The results of articulation show that TCV is not sufficiently clear compared to air-conducted. However, it is confirmed that TCV is robust to environmental noise because the accuracy rate is not decreased when the TCV is recorded under high ambient noise.

Keywords: Voice acquisition, Bone-conducted voice, Tooth-conducted voice, Communication-aid, Piezoelectric device

PACS number: 43.48.Kb, 43.38.Si [doi:10.1250/ast.36.500]

1. INTRODUCTION

Today, there are many applications using voice such as telephone, voice and video chat, and voice recognition. In these applications, it is necessary to acquire the talker's voice clearly, and therefore, researchers and engineers have focused on improving the clarity of voice.

In general, the use of a microphone is the most popular method used to acquire voice. Microphones acquire air-conducted voice (ACV), which refers to voice propagating through air. However, voices recorded using microphones are easily affected by environmental noise. To reduce the effect of environmental noise, several methods have been investigated, such as the use of close-talking microphones [1,2].

In contrast, methods for acquiring voice that does not interfere with the sound fields, such as using high-speed camera [3], avoids the influence of environmental noise. However, because the method needs machinery on a large scale and complex image processing, it can not be said an easy method.

On the other hand, there is an another simple method focusing on bone conduction. Bone-conducted voice (BCV) can be clearly extracted from voice vibration by only using a bone-conduction microphone (BCM) even if the environment is noisy, because BCV is voice which propagates through the skull during phonation. To record voice with bone conduction, several researches has been investigated. Some researches were focusing on an ear-insert type microphone and found that it can record voice in a noisy environment [4,5]. Tran *et al.* investigate the effective location of a BCM and reported that the forehead seems to be the best location for the recording of BCV [6]. However, since the BCV's intelligibility and quality depends on the difference among individuals, such as the shape of skull and thickness of skin, the method needs solutions which do not needs to think about the difference.

At the same time, because tooth can be regarded as bare bone, it was considered that bone conduction through tooth provides a better opportunity for efficient listening and

*e-mail: via.marnero.2@ruri.waseda.jp
†e-mail: dai-kuze-kjkw98@akane.waseda.jp
‡e-mail: junkokurosawa@gmail.com
§e-mail: yoikawa@waseda.jp
¶e-mail: y.yamasaki@gmail.com

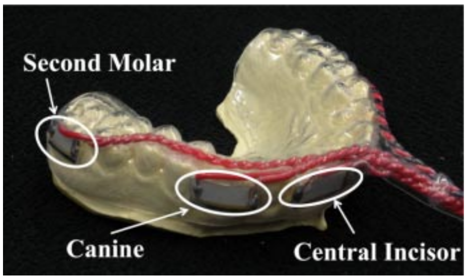

Fig. 1 Experimental TCM. Piezoelectric devices were put on Central Incisor, Canine, and Second Molar.

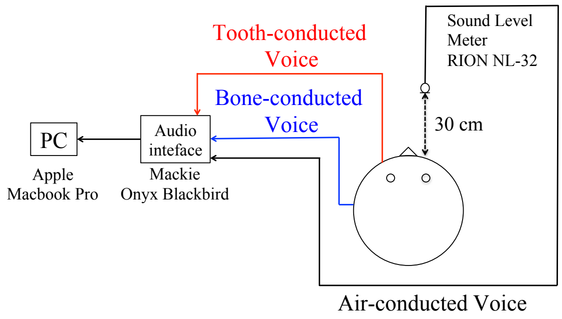

Fig. 2 Experimental Image of settings.

voice recording than bone conduction, which is influenced by the skin.

From above consideration, communication device using bone conduction through tooth has been developed and investigated. Microphones which use the method has been developed in several ways, one using MEMS accelerometer and the other one use magnet and coil [7,8]. These microphones pick-up the BCV through tooth and are not affected by environmental noise.

In contrast, we propose the device of tooth-conducted voice (TCV) for communication-aid which use piezoelectric device. Muramatsu *et al.* proposed the mouthpiece-form actuator that can be tightly worn in the mouth for listening and recording TCV [9,10]. In addition, we investigate the difference in the listening and recording of voices between types of teeth [11–14]. Thus, a tooth-conducted communication system has the possibility to become a widely used technology.

The purpose of this study was to incorporate tooth conduction into a communication-aid system. In the present study, we propose a tooth-conduction microphone (TCM) and evaluated its articulation of TCV. We confirmed that TCV is robust to environmental noise because the accuracy rate does not decrease even if it is recorded under high ambient noise.

2. STRUCTURE OF THE TOOTH-CONDUCTION MICROPHONE

In this study, we developed a tooth-conduction microphone (TCM). We term the voice extracted using the TCM as TCV.

The TCM consists of a piezoelectric device and dental material. In this study, we use the piezoelectric device (K2512U1) manufactured by THRIVE. The dental material is made from ethylene vinyl acetate. The piezoelectric device is laminated using dental material and formed to be a dental mold. Because the TCM is designed with an individual's dental mold, it can be tightly worn in the mouth.

The piezoelectric device located on the mouthpiece generates electricity corresponding to the vibration propagating through tooth during phonation. The electricity is recorded as signal as TCV through an audio interface.

An experimental TCM was developed for the experiment below and is shown in Fig. 1. Three locations were selected for TCV pickup; central incisor, canine, and second molar on the upper jaw. The pickup we used in the TCM is piezoelectric device, which is 12.5-mm long, 5-mm wide, and 0.1-mm thick.

3. EVALUATION OF TOOTH-CONDUCTED VOICE

3.1. Experiment 1: Recording Voice through Tooth

In this experiment, we extracted TCV, ACV, and BCV simultaneously under quiet condition.

The voices extracted in this study were Japanese vowels (/a/, /i/, /u/, /e/, /o/). The talker's voice was simultaneously recorded with the experimental TCM, a commercial bone-conduction microphone (BCM) and a sound-level meter, which is an air conduction microphone (ACM), for reference.

The recordings were made in a recording studio at the Waseda Univ. The noise floor level of the studio was 25.6 dBA. The recording equipments include the TCM, the BCM G-450 manufactured by NHC, the sound-level meter NL-32 manufactured by RION, and sound recording system consisting of a laptop computer manufactured by Apple, Onyx Blackbird audio interface manufactured by Mackie, and the DAW. A driving method of the BCM is a piezoelectric device. A schematic of the experimental settings is shown in Fig. 2.

The sound-level meter was placed in front of the talker. The distance between the talker and the sound-level meter was almost 30 cm. The talker wore the TCM and the BCM during phonation. The talker was instructed to utter all monosyllables at approximately 80 dB sound pressure level (SPL). The talker monitored the SPL of his/her voice on the display of the sound-level meter.

The waveforms of voice recorded with the TCM, BCM, and ACM are shown in Fig. 3. The spectrograms of the voice are shown in Fig. 4. From Fig. 3, we can confirm that

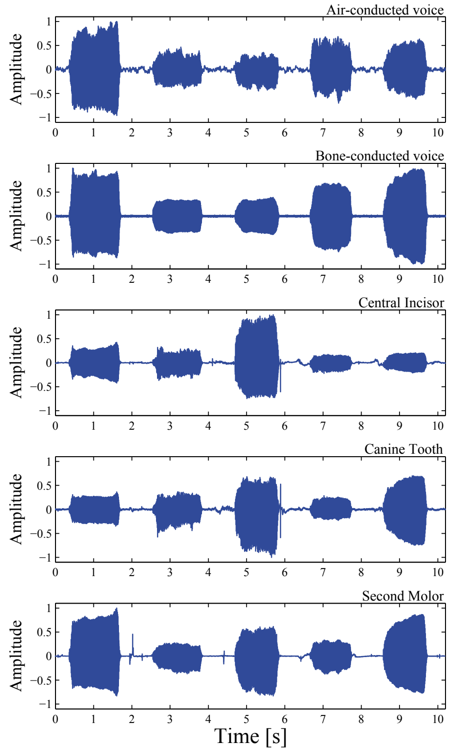

Fig. 3 Waveforms of voice recorded in Experiment 1.

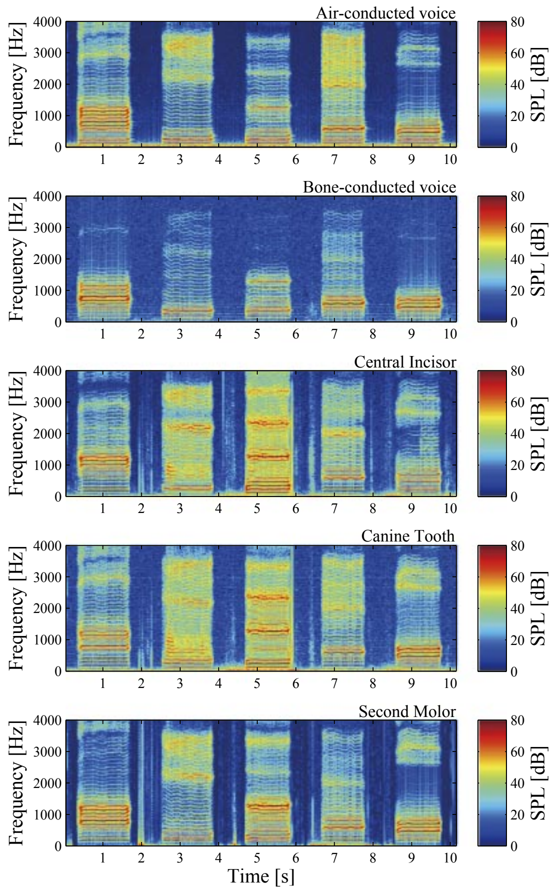

Fig. 4 Spectrogram of voice recorded in Experiment 1.

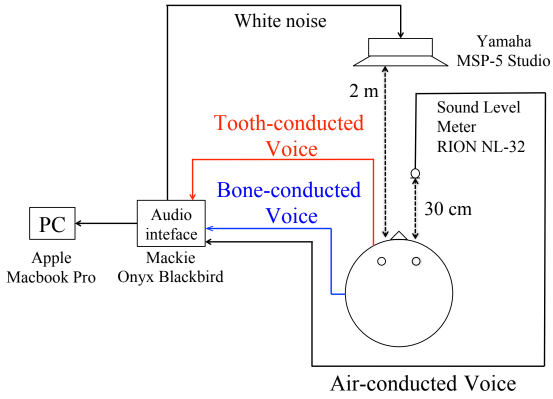

Fig. 5 Experimental Image of settings.

TCV can be extracted from all of the locations where piezoelectric devices were placed, as can be seen from each monosyllable's waveform. It can also be seen from Fig. 4 that each TCV contains frequency characteristics similar to that of ACV. In addition, we can confirm that BCV's contents decrease in the high-frequency band above 2,000 Hz.

3.2. Experiment 2: Under Noisy Environment

In this experiment, we extracted TCV by using the experimental TCM under noisy conditions to confirm the usability of the TCM under high ambient noise.

The voices extracted in this study formed the word "waseda." The voices produced by the talker were simultaneously recorded through the experimental TCM, the BCM, and sound-level meter.

The recording was held in the same studio as Experiment 1. White noise was played from a speaker (MSP-5 Studio, manufactured by YAMAHA) at a level of approximately 90 dB SPL, as measured by the sound-level meter in front of the subject. A schematic of the experimental setting shown in Fig. 5.

The waveforms of voice (a word, "waseda") recorded with the TCM, the BCM, and sound level meter are shown in Fig. 6. The spectrograms of the voice is shown in Fig. 7. From Fig. 6, becuase the voice is masked by white noise, the waveform of /waseda/ is hardly seen in ACV. However, the extracted TCV is clearer than ACV. In Fig. 7, it is also indicated that the frequency band of ACV containing the word /waseda/ cannot be identified because it is masked by white noise. And the TCV extracted from canine is less affected by white noise at frequencies above 3,000 Hz than those extracted from other locations.

4. ARTICULATION OF TOOTH-CONDUCTED VOICE

In the previous section, we simultaneously recorded the ACV, BCV, and TCV and numerically analyzed the TCV

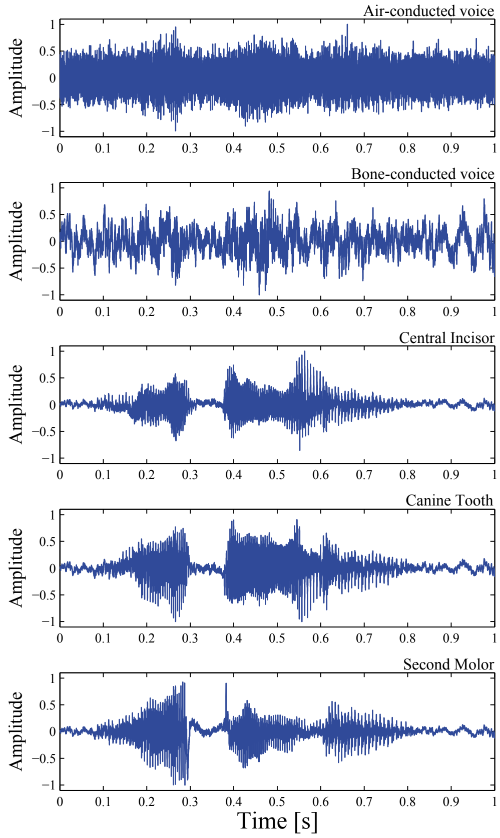

Fig. 6 Waveforms of voice recorded in Experiment 2.

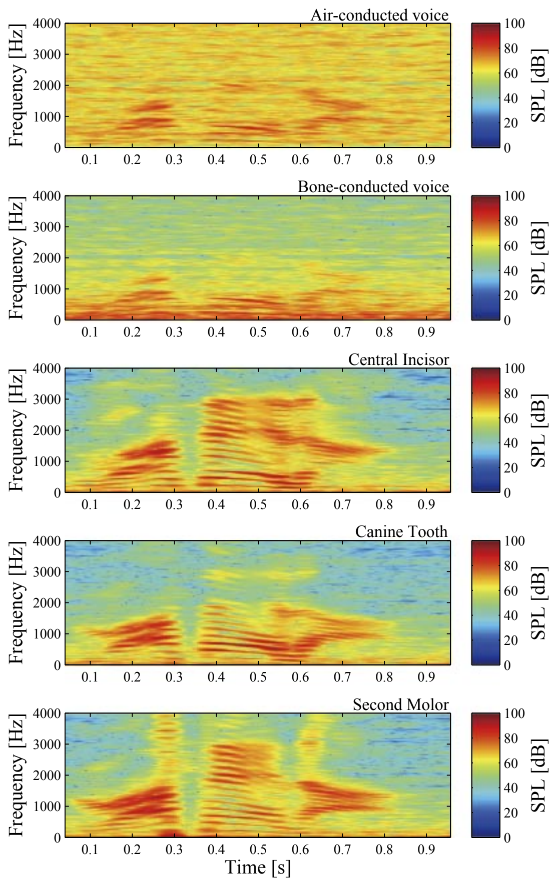

Fig. 7 Spectrogram of voice recorded in Experiment 2.

Table 1 Experiment condition for the test voices, condition means noise floor level of each conditions.

<table>
<tr><th>talker</th><th>talker A</th><th>talker B</th></tr>
<tr><td>listener (male:female)</td><td>5 persons (5:0)</td><td>6 persons (4:2)</td></tr>
<tr><td>under quiet condition</td><td colspan="2">27 dB</td></tr>
<tr><td>under noisy condition</td><td>80 dB</td><td>90 dB</td></tr>
</table>

signals. The analysis showed that TCV contains resonance frequencies similar to those of ACV, and it is robust to noise. However, the numerical analysis could not evaluate the practical usability considering the auditory perception of humans. Therefore, we evaluated the voices through subjective and objective assessments.

In typical subjective assessment, there are an intelligibility test and an articulation test. An intelligibility test evaluates the accuracy rate of a hearing test using the meaningful word and sentences. Therefore, it can be evaluated as good rate than expected because it is possible to presume even if there is a syllable uncertain in the words and sentences. In contrast, because an articulation test using the meaningless monosyllables and word, it can evaluate how the extracted voice signals can be transmitted to the listener exactly.

In objective assessment, it is evaluated by comparing the original voice signal (ACV) and the acquired voice signal (TCV, BCV). The sound-to-noise ratio (SNR) is one of the most basic parameters used for the assessment of voice quality and it is focused on the difference between data in the time domain. On the other hand, since it is known that the auditory perception of humans is not affected by the phase distortion of voice signals, it is important to focus on the difference between voice signals in the frequency domain. To evaluate the signal in frequency domain, linear predictive coding (LPC) cepstral distance (LCD) was introduced. It has been reported that LCD values show good agreement with subjective assessments because it is based on the characteristics of human hearing [15].

From the above reasons, a monosyllable articulation test was adopted for subjective assessment, and LCD was adopted for objective assessment.

4.1. Monosyllable Articulation

In this experiment, we evaluate TCV through the articulation test.

Table 1 lists the experiment conditions for the test voice. In the test, ACV and TCV were recorded simultaneously as test sounds. Test voices were recorded under two conditions, which are the same as in Experiment 1 and 2. Test sounds contain 20 monosyllables from the Japanese monosyllable word list 67-S, developed by the Japan Audiological Society.

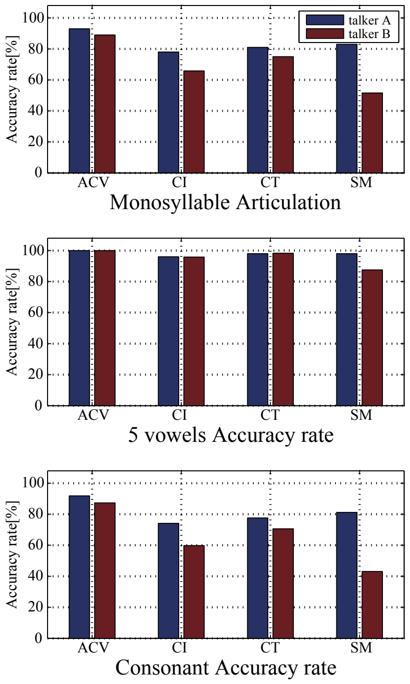

Fig. 8 Accuracy rate of Monosyllable articulation, 5 vowels and consonant (under quiet condition).

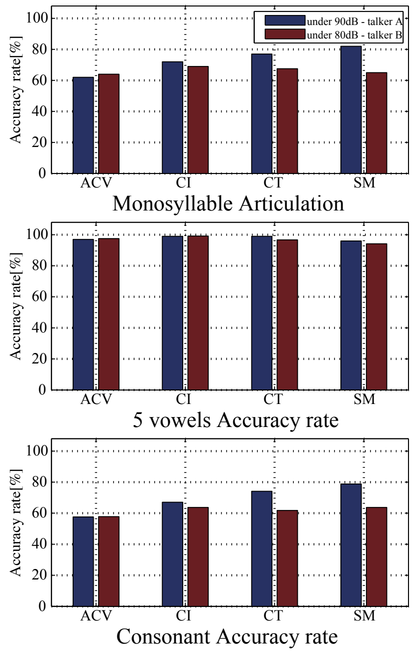

Fig. 9 Accuracy rate of Monosyllable articulation, 5 vowels and consonant (under noisy condition).

In the articulation test, the recorded ACV and TCV were played from the computer while listeners responding to the voice by using only a pencil. The listeners were presented with a set of words from the Japanese monosyllables word list 67-S recorded in this experiment; the word list consisted of 17 consonant-vowel (CV) syllables and 3 vowel syllables, and each set consisted of 4 presentations (ACV + 3 locations of TCV) and 2 situations (TCV recorded under quiet condition and under noisy condition).

Eleven listeners with normal hearing in the age range of 20 to 23 years participated in this experiment. A participant wearing a headphone was seated in front of the computer that played test sounds. The participant was instructed to write the words as it heard from the test sounds.

The result of the articulation test are shown in Figs. 8 and 9. In the figures of this paper, "CI," "CT," and "SM" are abbreviations for "central incisor," "canine tooth," and "second molor." The figures contains three result; "Monosyllable Articulation" focusing on the accuracy rate of all monosyllable, "5 vowels Accuracy rate" focusing on the accuracy rate of the monosyllables which contain 5 vowels, and "Consonant Accuracy rate" focusing on the accuracy rate of the monosyllables which contains consonants.

From Fig. 8, the accuracy rate of the monosyllable articulation with the TCV recorded under quiet conditions is nearly 70% (mean value calculated between 2 talkers and 3 locations on teeth), which it is a little inferior compared to that of ACV. However, from Fig. 9, the accuracy rate of the articulation with the ACV and TCV recorded under noisy conditions indicated that the accuracy rate of ACV decreases almost 30%, while the accuracy rate of TCV maintains the value obtained under quiet conditions.

4.2. LPC Cepstral Distance

In addition to the previous subjective assessment, we evaluated the LPC cepstral distance of the TCV for objective assessment.

LCD values indicate the vector distance between an original voice signal (ACV in this paper) and a comparison voice signal (BCV and TCV) and is defined as

$$\mathrm{LCD} = \frac{10}{\log 10}\sqrt{2\sum_{i=1}^{M}[c_x(i) - c'_x(i)]^2} \tag{1}$$

where $c_x(i)$ is the LPC cepstral coefficient of the original voice signal, $c'_x(i)$ is the LPC cepstral coefficient of the comparison voice signal, and M is the order of the coefficient. A smaller LCD indicates that the comparison signal is more similar to the original voice signal.

In this experiment, we record the ACV, BCV, and TCV simultaneously. The same equipments as in Experiment 1 were used. The contents of voice were Japanese vowels (/a/, /i/, /u/, /e/, /o/). To value the LCD, it was calculated for 0.5 s when phonation was stable and the order of LPC cepstral coefficient was 16.

The result of LCD is shown in Fig. 10. To understand the result of the LCD more simple, we calculated time mean value of the LCD. From Fig. 10, we placed sets of the LCD which contains BCV, CI, CT, and SM, for each 5

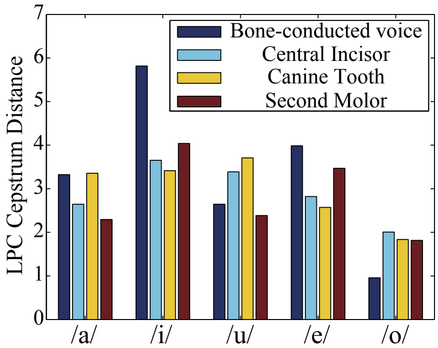

Fig. 10 Average values of LPC cepstral distance, sets of the LCD for each vowels contains BCV, CI, CT, and SM (from left to right).

vowels in X-axis. Y-axis shows the time mean values of the LCD.

From Fig. 10, the LCD of TCV is smaller than the LCD of BCV at /a/, /i/, and /e/, indicating that TCV is similar to ACV for these three vowels. In contrast, the LCD of BCV is much smaller than that of TCV at /o/.

5. DISCUSSION

We investigated the data presented in this study to reveal the articulation of TCV and effective location to extract TCV.

5.1. Recording Voice with Tooth-conduction Microphone

Through the Experiment 1 and 2, we evaluated the TCV through numerical analysis.

In Experiment 1, the ACV, BCV, and TCV which recorded simultaneously under quiet condition were evaluated. The result of this experiment shows the TCV from each tooth contains similar frequency characteristic to that of the ACV. At the same time, the frequency characteristic of the BCV was decreased above 2,000 Hz. This result shows the characteristic of BCV, the result of interference of skin.

In Experiment 2, we confirmed the usability of TCV under noisy conditions. From the result of each numerical analysis, the ACV was masked by the white noise and couldn't find out in details. In contrast, the TCV from each locations could find in detail.

From above result, it was found that the TCV have some similarities with the ACV and is robust to environmental noise. However, it couldn't be found a clear difference between the TCV from each locations.

5.2. Articulation of Tooth-conducted Voice

To reveal the characteristic of TCV in detail, we carried out the articulation tests.

In the evaluation of monosyllable articulation, it is assumed that the TCV is robust to noise. To focus on the recorded voices of talker A from Figs. 8 and 9, both of the TCV, recorded under quiet and noisy condition, have a characteristic that the inner tooth scored high accuracy rate than in front of that. It is also confirmed that the accuracy rate of monosyllable articulation rated 20% at most, comparing the ACV and TCV at second molor of talker A from Fig. 9. In contrast, the result of talker B showed completely different. This may caused by the difference of individuals, such as the shape of the tooth and the way of speaking. Therefore, we need to develop the device which is not influenced by such difference of individuals.

From the evaluation, we also confirmed that the accuracy rate of 5 vowels of TCV under each condition scored high. The result above indicated that TCV can record the formant of 5 vowels, which is similar to ACV and it is also indicated that the TCV is almost equal to ACV. In contrast, to focus on the accuracy rate of consonant, it scored low articulation. It is assumed that the low accuracy rate of monosyllable caused by the low accuracy rate of consonant. And the reason why accuracy rate of consonant is low came from the difficulty of speaking with TCM, because the wire which send the voice signal is appeared from mouth.

In the evaluation of LPC cepstral distance, we confirmed that the vowel, /a/, /i/, and /e/ of the TCV have advantages to BCV, and other have not. The reason why the LCD of BCV is small is thought to be the formant that phonation /o/ contains the low frequency band; BCM could record the formant similar to ACV.

From all of the result, we confirmed that the TCV is robust to the environment noise. However, we could not make the difference between tooth clear because the result might be depended on the difference of individuals. Therefore, we will perform the experiment with a number of subjects.

5.3. Future Work

To develop a mutual communication system that use tooth-conduction, as shown in the Fig. 11, wireless devices are required. Until now, wireless tooth-conduction hearing-aid using telecoil has been developed in our laboratory [10]. However, this device needs to stay inside of a magnetic loop to listen to a voice. Accordingly, we developed the TCM using the Bluetooth, one of the wireless technology standard for exchanging data over short distances. The Bluetooth enable to exchange the data between the device which has the Bluetooth module, therefore, it is suitable for the device which can listening and speaking.

Figure 12 shows the wireless TCM using the Bluetooth module. The Bluetooth module that used for the wireless

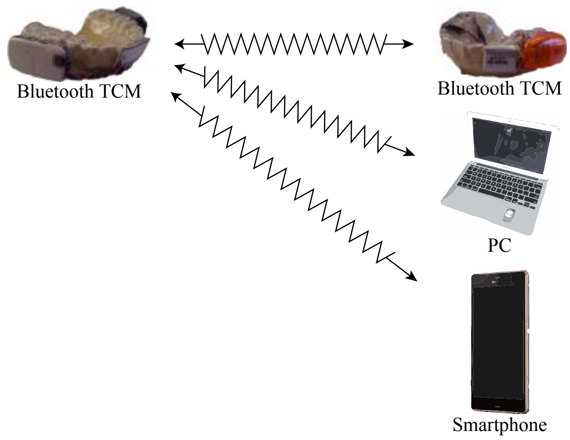

Fig. 11 Mutual communication system through tooth-conduction.

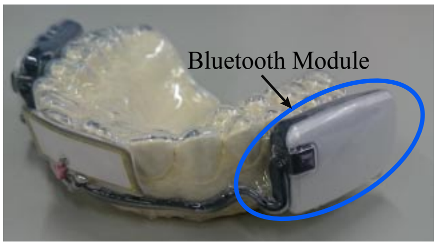

Fig. 12 Mouthpiece-form wireless TCM, for practical use as communication-aid device.

TCM is ver.4.0. The wireless TCM enables not only acquire the TCV in wireless but also removable of the wire which is appeared from mouth.

6. CONCLUSION

In this study, we investigated the articulation of the tooth-conducted voice and effective tooth locations for tooth conduction microphone. In the experiment of extracting TCV, tooth vibration during phonation could be converted into a sound signal by using the TCM. It was confirmed through numerical analysis that TCM could record voices even if under high ambient noise. In addition, the articulation test of TCV shows that TCV can record the formant similar to ACV and is robust to environmental noise. Moreover, it is indicated that TCV is disadvantageous to extract consonants clearly.

For a more effective extraction of voice, it is necessary to improve the accuracy rate of consonants and to remove of the influence of difference of individuals. As this disadvantage is thought to be caused by the structure of the TCM, we will redesign the structure to remove the difficulties in phonation when wearing the TCM.

In addition, we introduced a wireless TCM, however, it can only record TCV. To put the tooth-conduction communication-aid system into practical use, it is necessary to enable simultaneous listening of TCV when using the TCM.

Through these improvements, we aim to achieve the practical use of this system with development of the mouthpiece-form TCM's structure in the future.

REFERENCES

[1] Y. Haneda, K. Furuya, S. Koyama and K. Niwa, "Close-talking microphone array based on spherical harmonic expansion," *IEICE Trans. Fundam.*, **J97-A**, 264–273 (2014) (in Japanese).
[2] Y. Haneda, K. Furuya, S. Koyama and K. Niwa, "Close-talking spherical microphone array using sound pressure interpolation based on spherical harmonic expansion," *Proc. ICASSP 2014*, pp. 604–608 (2014).
[3] M. Akutsu, Y. Oikawa and Y. Yamasaki, "Extract voice information using high-speed camera," *Proc. ICA* (*POMA*, Vol. 19), pp. 055019 (2013).
[4] R. Black, "Ear-insert microphone," *J. Acoust. Soc. Am.*, **39**, 260–264 (1957).
[5] H. Ono, "Improvement and evaluation of the vibration pick-up-type ear microphone and two-way communication device," *J. Acoust. Soc. Am.*, **62**, 760–768 (1977).
[6] P. Tran, T. Letowski and M. McBride, "Bone conduction microphone: Head sensitivity mapping for speech articulation and sound quality," *ICALIP*, pp. 107–111 (2008).
[7] C. Wieland and M. Yager, "Tooth-magnet microphone for high noise environments," U.S. Patent, US20090022351, A1 (22, January, 2009).
[8] M. Anjanappa, X. Chen, R. Bogacki and A. Babik, "Method and apparatus for tooth bone conduction microphone," U.S. Patent, 7,486,798, B2 (22, January, 2009).
[9] M. Muramatsu, Y. Oikawa, Y. Yamasaki and J. Kurosawa, "Communication aid using the actuator for bone conduction," *IEICE Trans. Fundam.*, **J95-A**, 623–630 (2012) (in Japanese).
[10] M. Muramatsu, J. Kurosawa, Y. Oikawa and Y. Yamasaki, "Communication aid utilizing bone-conducted sound via teeth by means of mouthpiece form actuator," *Proc. ICA* (*POMA*, Vol. 19), pp. 050090 (2013).
[11] Y. Torikai, D. Kuze, M. Muramatsu, J. Kurosawa, Y. Oikawa and Y. Yamasaki, "Speech acquisition via teeth by means of mouthpiece form actuator," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 1337–1378 (2013) (in Japanese).
[12] D. Kuze, Y. Torikai, J. Kurosawa, Y. Oikawa and Y. Yamasaki, "Legibility of hearing by bone conducting with a mouthpiece form actuator," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 505–506 (2013) (in Japanese).
[13] Y. Torikai, D. Kuze, J. Kurosawa, Y. Oikawa and Y. Yamasaki, "Articulation evaluation of Tooth-conducted speech," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 1479–1480 (2014) (in Japanese).
[14] D. Kuze, Y. Torikai, J. Kurosawa, Y. Oikawa and Y. Yamasaki, "Hearing via teeth: Loudness level and intelligibility through mouthpiece formed actuator," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 1481–1482 (2014) (in Japanese).
[15] N. Kitawaki, H. Nagabuchi and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," *IEEE J. Sel. Areas Commun.*, **6**, 242–248 (1998).