

RESEARCH ARTICLE

Transparency Mode of Hearable Reduces Your Spatial Hearing: Evaluation and Cancelling Method to Restore Spatial Hearing

HIROKI WATANABE¹ AND TSUTOMU TERADA², (Member, IEEE)

¹Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

²Graduate School of Engineering, Kobe University, Kobe 657-8501, Japan

Corresponding author: Hiroki Watanabe (hiroki.watanabe@ist.hokudai.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under Grant JP21K11973, in part by the Japan Science and Technology Agency (JST) Precursory Research for Embryonic Science and Technology (PRESTO) under Grant JPMJPR2138, and in part by the JST Core Research for Evolutional Science and Technology (CREST) under Grant JPMJCR18A3.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Review Committee on Research with Human Subjects, Hokkaido University Engineering under Application No. R4-1.

ABSTRACT Many earphone-type wearable devices (hearables) with noise-canceling features can capture external sound and present it to the user (transparency mode). This function helps the user avoid being cut off from the outside while wearing the earphones. However, because the built-in microphone/speaker has a frequency response, the presented sound differs from the sound acquired by the user's original auditory system. Since humans use a head-related transfer function (HRTF) to identify the direction of sound sources, changes in the user's HRTF caused by transparency mode may adversely affect their sound localization ability. Therefore, this study investigated the changes in sound localization ability when using the transparency mode of commercially available hearables. A sound localization experiment was conducted on 10 participants using three hearables to identify 12 sound sources separated by 30° around each participant. The results show that the accuracy of sound localization decreases from 91.5% to 68.9% when using transparency mode. Moreover, a method was proposed to cancel the device's microphone/speaker frequency characteristics. Evaluation results confirmed that the proposed method increased the accuracy of sound localization from 62.7% to 72.0%.

INDEX TERMS Hearable, sound localization, wearable computing.

I. INTRODUCTION

With recent technological developments, earphone-type wearable devices, referred to as “hearables” [1], [2], are rapidly becoming popular. In the fourth quarter of 2021, hearables accounted for almost two-thirds of the wearables market, according to market research firm International Data Corporation [3]. One of the main functions of a hearable is noise cancellation, which can reduce external noise. Many hearables with noise-canceling functions also provide a function that captures sounds from outside and presents them to

the user (transparency mode). This function solves the issue of the user being cut off from the outside world while wearing earphones. However, the sound presented using transparency mode differs from the sound acquired by the user's original auditory system because the built-in microphone/speaker has a frequency response. In addition, installing the hearable causes changes in the acoustic properties around the ear. Since humans use a head-related transfer function (HRTF) to identify the direction of sound sources [4], changes in the user's HRTF caused by the transparency mode of hearables may adversely affect sound localization. If this problem exists, it could lead to unexpected accidents while wearing a hearable (e.g., the user mistakes the position of a car/bicycle).

The associate editor coordinating the review of this manuscript and approving it for publication was Giuseppe Desolda¹.

Hearables have rapidly become popular in recent years and are available to everyone. However, the limitations of the transparency mode have yet to be well known. It is crucial to evaluate the performance of transparency mode for the recent commercially available hearables.

Therefore, this study first investigated the effect of the transparency mode of hearables on the user's sound localization ability. A sound localization experiment was conducted on 10 participants using three hearables to identify 12 sound sources separated by 30° around each participant. The result showed that the average accuracy of sound localization was 91.5% without and 68.9% with a hearable, respectively, indicating a decrease in sound localization ability.

In order to mitigate the degradation of sound localization, a method was proposed to cancel the microphone and speaker frequency characteristics, which is considered one of the causes of the degradation of sound localization. The frequency characteristics were measured in advance to determine how they change before and after passing through the hearable, and a method was implemented to cancel out these changes when the transparency mode is used. Evaluation results confirmed that the accuracy of sound localization improved from 62.7% to 72.0% by using the proposed method.

The contributions of this study are as follows:

- The sound localization ability was deteriorated by the transparency mode of commercially available hearables and alarmed possible threats when using the transparency mode.
- A method was proposed for improving the sound localization ability when using transparency mode by canceling the frequency characteristics of hearables.
- Evaluation results confirmed that the proposed method improved the accuracy of sound localization ability.

II. RELATED WORK

A. SOUND LOCALIZATION DEGRADATION

Studies have been conducted investigating the degradation of sound localization while wearing hearing aids and hear-through headphones or electronic hearing protectors [5], [6], [7], [8], [9]. Marentakis et al. evaluated the sound localization ability while wearing a hear-through system and found that it was compromised compared to listening without earphones [6]. Bogaert et al. investigated the effect of bilateral hearing aids on directional hearing in the frontal horizontal plane. The results showed bilateral hearing aids do not preserve localization cues [7]. Keidser et al. investigated the influence of hearing aids on horizontal-plane localization performance [8]. Due to inaudibility at a high frequency, front-back confusions occurred and remained prominent when aided with omnidirectional microphones. Denk et al. investigated the impact of the microphone location, signal bandwidth, different equalization approaches, and processing delays in superposition with direct sound leaking through a vent [9]. Albrecht et al. showed that users' auditory characteristics change when they wear earphones. However,

their study did not investigate changes in sound localization ability resulting from changes in auditory characteristics [10]. A method for manipulating the frequency of external sound has been proposed, and the changes in the direction of sound localization that occurs when this method is applied have been investigated [11]. However, because this study was conducted with the user free to move, the sound localization ability at rest must be clarified. It is known that the accuracy of sound localization is greatly improved in the head motion state compared with the headrest state [12]. Thus, it is essential to investigate the sound localization ability when using transparency mode, even in the headrest state. Altered Pinna is a system that uses a wearable device to physically deform the earlobe and manipulate the direction in which sound is perceived [13]. This study differs from the previous study in terms of investigating changes in sound localization using the transparency mode of hearables.

The above studies have investigated sound localization for hear-through systems and hearing aids; however, no recent commercially available hearables have been investigated to our knowledge. Hearables have rapidly become common in recent years. However, the limitations of the transparency mode need to be better recognized. It is crucial to evaluate the performance of transparency mode for the recent commercially available hearables because hearables are not devices targeted at a specific group of people, and their target users are significantly different from those in the past. Furthermore, the current study proposes a method to improve sound localization degradation by canceling the frequency characteristics of the microphone and speaker of the hearable.

B. PERSONALIZED HRTF

In order to improve the music experience and immersion in contents, there are methods to select/generate HRTFs that are suited to the individual based on the user's auricular shape. Iida et al. proposed a method to estimate the frequencies of the two lowest spectral notches, which play an essential role in vertical localization, in the HRTF of an individual listener by anthropometry of the listener's pinnae [14]. Spagnol et al. also developed a method to extract the frequencies of the main pinna notches in the frontal part [15]. Mokhtari et al. estimated the peak of HRTF with geometries of pinnae [16], [17]. In recent years, personalized HRTF has been available as a commercial service [18], [19]. They offer a service that can personalize HRTFs from an image captured by smartphone cameras. This technique could also be applied to the transparency mode to improve the sound localization capability of the transparency mode. Specifically, the sound obtained by the transparency mode is corrected by the user's personalized HRTF. However, since HRTFs vary depending on the direction of the sound source, predicting the location of the sound source is necessary for this method to be effective. It is considered difficult to accurately estimate the location of a sound source using only the two microphones in a typical hearable.

Since the proposed method cancels only device frequency characteristics, it does not require HRTF personalization or sound source estimation. Device characteristics are constant for each device; thus, once a measurement is made, it can be used by anyone. Additionally, since the proposed method uses acoustic signal input to a user-worn hearable, the input signal already contains HRTFs according to the sound source direction.

III. SOUND LOCALIZATION USING TRANSPARENCY MODE

A. SOUND LOCALIZATION PRINCIPLE

Sound waves are affected by the head, auricle, and torso before they reach the eardrum. Such changes in the physical properties of incident sound waves due to the head area are expressed in the frequency domain and are referred to as the HRTF [4]. The interaural time difference (ITD) and interaural level difference (ILD) generated by the HRTF are used as cues for human perception of sound's left and right directions. Specifically, the ITD is mainly used for low frequencies, and the ILD is mainly used for high frequencies [20]. Additionally, the amplitude spectrum of the HRTF (spectral cue) is used as a cue for the perception of the front-back vertical direction [21].

In the transparency mode of a hearable, the sound captured by a microphone is presented to the user through a built-in speaker. The frequency characteristics of the microphone/speaker are assumed to affect the above characteristics adversely.

B. SOUND LOCALIZATION

1) EXPERIMENTAL SETUP

A sound localization experiment was conducted to investigate the change in sound localization ability with/without a hearable. Figure 1 shows the experimental environment. The speakers (Fostex P650K) were placed at a height of 1.2 m in 12 directions at 30° intervals on a horizontal surface circumference of 1 m in radius, and the participant sat in the center. Participants underwent the sound localization experiment with their heads stationary and eyes closed. White noise was used as the stimulus sound with a sound pressure level of 70 dB for 3 seconds. Fade-in/out processing was applied at the beginning/end of the stimulus sounds. Participants responded with the perceived direction of the sound. When they could not sense the direction of the sound and perceived it as being in the center of the head, they were instructed to respond with "S0," shown at the center of Figure 1. Sound-absorbing materials were attached in four directions around the experimental environment to reduce the effect of reflected sound. The background noise level was approximately 39 dB. The 10 participants were 22 to 32-year-old males and females, and the hearables used were Anker Soundcore Liberty Air 2 Pro (Soundcore), Samsung Galaxy Buds Pro (Galaxy), and Apple AirPods Pro (AirPods). In order to investigate the relationship between

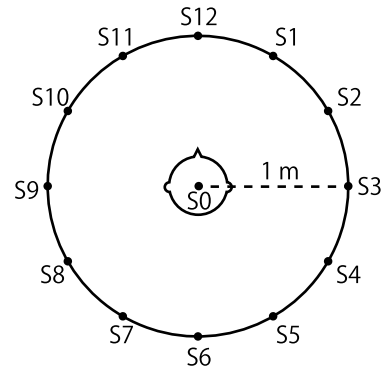


FIGURE 1. Experimental environment.

the price and performance of the hearables, three different types of devices were selected in different price ranges, that is, Soundcore, Galaxy, and AirPods, which were priced at approximately \$100, \$200, and \$250, respectively. Participants performed the sound localization experiment where they did not wear the devices (open-ear condition) and in one where they wore each hearable and used transparency mode (device-wearing condition). All participants first experienced the open-ear condition, and then the order of the hearables used in the experiment was changed for each participant. The experiments were approved by Ethics Review Committee on Research with Human Subjects, Hokkaido University Engineering (R4-1).

As a metric, the accuracy of correct answers was defined as follows:

$$\text{Accuracy} = \frac{100N_{\text{cor}}}{N}, \quad (1)$$

where N_{cor} is the number of correct answers and N is the number of all stimulus.

2) RESULT

Figure 2 shows the result of each condition. The horizontal axis indicates the number of loudspeakers from which the stimulus sound was presented, and the vertical axis indicates the number of the sound sources perceived by the participant; that is, the more the circles are concentrated on the diagonal line from the lower left to the upper right, the more accurate the sound localization was. The size of the circle is proportional to the number of responses.

As shown in Figure 2a, participants could localize the sound accurately under the open-ear condition, which is their original hearing ability. Even when there was an error in the response, it was confirmed to be within 30°. In comparison, several cases were found where the presented sound source and the perceived sound source differed significantly for the device-wearing conditions. In particular, the front and back sound sources were confirmed to be mistaken (S1/S5, S2/S4, S6/S12, S7/S11, and S8/S10). Several cases were observed in which the front and back sound sources (S6 and S12) were perceived as being at the head center (S0).

Figure 3 shows the sound localization accuracy of each condition. In the open-ear condition, the accuracy was more

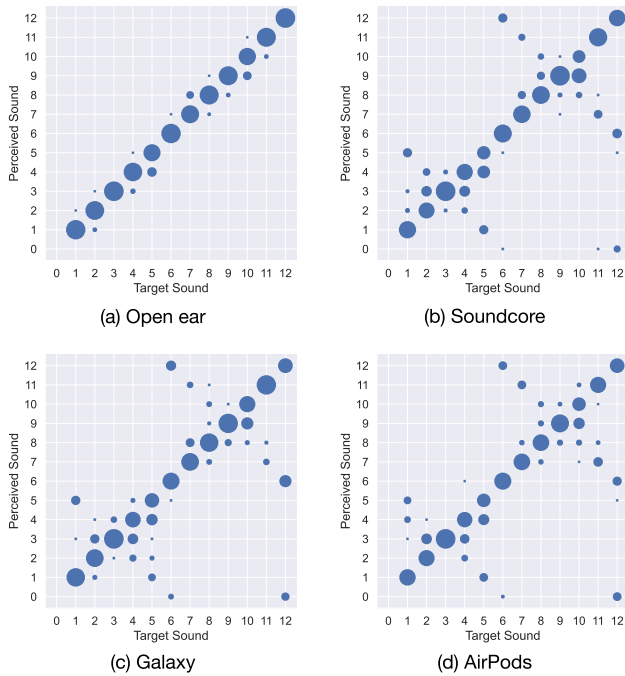


FIGURE 2. Result of sound localization experiment.

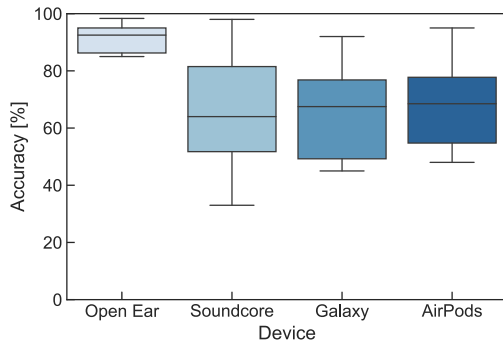


FIGURE 3. Accuracy of each condition.

than 90%, while in the device-wearing conditions, it was 68.0%, 70.0%, and 68.7% for Soundcore, Galaxy, and AirPods, respectively. The result of the analysis of variance confirmed that there was a significant effect of condition ($p < .01$). The result of multiple comparisons with Bonferroni corrections showed that there was a significant difference between open ear and Soundcore ($p < .05$); open ear and Galaxy ($p < .05$); and open ear and AirPods ($p < .05$).

Figure 4 shows the accuracy for each sound source. The accuracy from all sound sources decreased compared with the open-ear condition. The effect of sound localization degradation was limited for S3 and S9 compared with the other sound sources. This is because S3 and S9 are on the binaural axis, which makes both ILD and ITD distinctive. In comparison, as shown in Figure 2, S2/S4 and S8/S10, which shifted 30° from S3 and S9, were confused with each other, resulting in a significant decrease in accuracy.

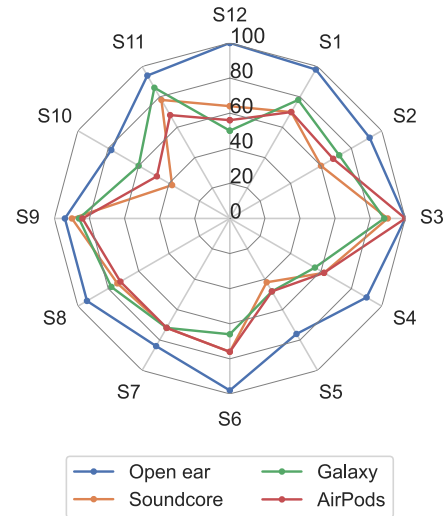


FIGURE 4. Accuracy of each sound source.

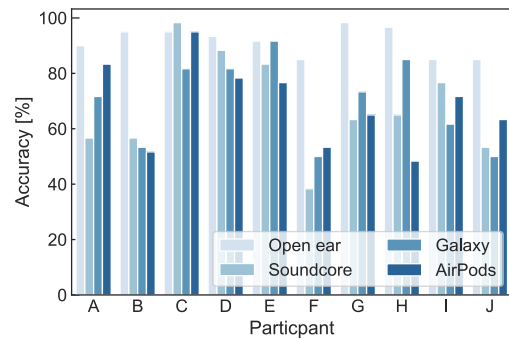


FIGURE 5. Accuracy of each participant.

Figure 5 shows accuracy for each participant. As we can see from the figure, the results vary depending on the person, e.g., some participants, such as participants B, F, and G, showed a significant decrease in accuracy in the device-wearing condition. In contrast, others, such as participants C, D, and E, showed little change in accuracy in the device-wearing condition.

3) DISCUSSION

As shown in Figure 2, front-back confusions occurred with transparency mode. The danger of over-reliance on transparency mode is being highlighted based on the results of this study. While the user can hear external sounds using transparency mode, it should be known that the sound localization performance deteriorates. In particular, front-back confusion can cause significant problems, e.g., wearing hearables in transparency mode while riding a bicycle or driving a car can be dangerous even if the user is not listening to audio content.

This study hypothesized that there would be performance differences among devices in different price ranges, resulting in differences in sound localization ability, and three different types of devices in different price ranges were used. However, as shown in Figure 3, device price did not correspond to sound

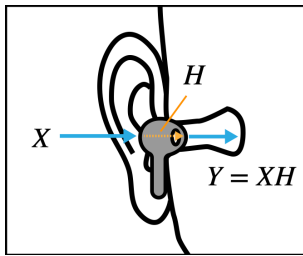


FIGURE 6. Overview of the system.

localization accuracy. These results confirm that sound localization degradation occurs not only in inexpensive models but also in high-end models.

IV. IMPROVEMENT OF SOUND LOCALIZATION

As confirmed in the previous section, the transparency mode of the hearable causes degradation of sound localization. This is because the changes in HRTF cause the degradation of sound localization. Based on the previous study [9], two points are considered to be major factors in the degradation of sound localization:

- Non-optimum microphone location disrupts sound localization in the vertical domain and reduces the accuracy in lateral localization.
- Frequency response characteristics of the hearable.

The first factor is difficult to improve, given the general shape of the hearables. Therefore, in this study, the second factor is given importance; that is, it is hypothesized that canceling acoustic characteristics caused by passing through the hearable would improve the performance of sound localization. One advantage of considering device characteristics is that it is unnecessary to consider each user’s personalized HRTF because device characteristics are constant from device to device. In addition, the proposed method does not require sound source localization to apply the personalized HRTF.

Figure 6 shows the overview of the system. The signal obtained by the user via the hearable is shown below;

$$Y(\omega) = X(\omega)H(\omega) \tag{2}$$

where X , Y , H , and ω denote the input signal to the outside microphone, output from the inside speaker, acoustic characteristics of the hearable, and frequency, respectively. Therefore, the characteristics of a hearable can be derived from the following equation;

$$H(\omega) = \frac{Y(\omega)}{X(\omega)}. \tag{3}$$

Since the signal the user initially expects to obtain is $X(\omega)$, dividing the input signal by the device characteristics $H(\omega)$ yields $X(\omega)$ as the output signal.

V. IMPLEMENTATION

We measured the hearable’s frequency characteristics based on the abovementioned idea. As shown in Figure 7, microphones with flat frequency response were placed (Earthworks

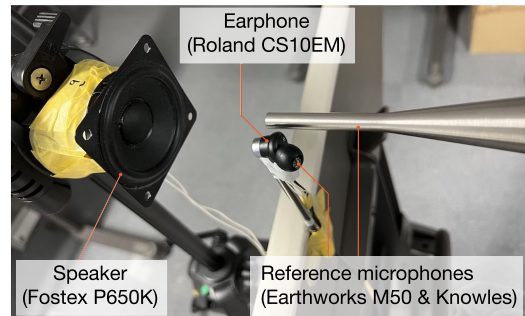


FIGURE 7. Measurement of device frequency characteristics.

M50 and Knowles SPU0410LR5H-QB) near the outside microphone and inside the earphone speaker for reference and investigated how the frequency response changes via the device.

Because the proposed method requires processing the sound acquired by the device’s outside microphone and presenting it to the user, applying it to existing true wireless hearables was difficult. Therefore, Roland’s CS10EM was used as an earphone to obtain the output of the outside microphone.

Figure 8 shows the device configuration of the proposed method. The sound captured by the outside microphone is sent to the PC (MacBook Pro) via the audio interface (Tascam DR-07X). The PC performs the modification described in Section IV, and the modified sound is played back from the earphone speaker via the audio interface. Python was used for signal processing on the PC. The system calculates fast Fourier transform (FFT) and obtains frequency characteristics of the input signal. Then, the frequency response of the input signal is divided by the frequency response of the hearable. Finally, inverse FFT transforms the divided frequency response into the time domain and outputs from the earphone. The number of FFT samples is 1024. Note that the system needs at least 23 ms to capture 1024 samples, which produces an echo-like effect because the user hears the sound leaking from outside and delayed sound from the earphones. The effect of the delay is discussed in Section VII.

Out of the obtained device characteristics, only the 2 kHz–20 kHz range was used. This is because to mitigate a drastic change in the user’s usual listening experience and to modify above 5 kHz, which is necessary for sound localization [21]. Additionally, the obtained device characteristics were unstable above 20 kHz.

Figure 9 shows the frequency response when the device characteristics are canceled by the measured $H(\omega)$. Roland CS10EM does not have a transparency mode; therefore, transparency mode is reproduced by playing back the sound captured by the outside microphone from the inside speaker. As the figure shows, there is no significant change up to 5 kHz with modification; however, the proposed method reduced the notch around 7 kHz and enhanced the higher frequencies afterward.

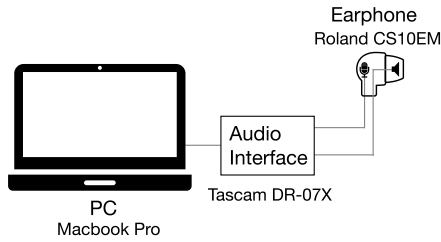


FIGURE 8. Device configuration.

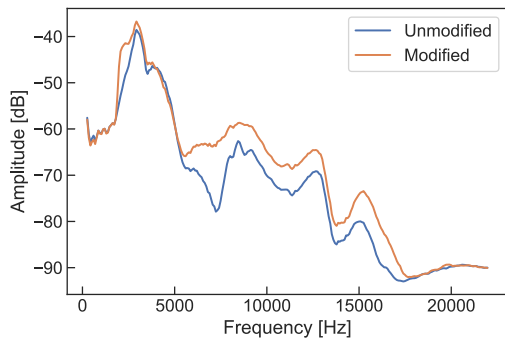


FIGURE 9. Comparison of spectrum.

VI. EVALUATION

A. EXPERIMENTAL SETUP

The effectiveness of the improvement method was evaluated. Sound localization experiments were conducted, and the procedure was the same as in Section III. Two experimental conditions were performed:

- Unmodified condition: conventional transparency mode, i.e., capturing the sound by the outside microphone and playing it from the inside speaker.
- Modified condition: the proposed method was applied to the sound captured from the outside microphone.

The 13 participants were 22 to 33-year-old males and females. Five out of 13 participants were the same as in Section III evaluation. Note that the participants' IDs in Section III and this section differ for the same participant. To investigate the participants' inherent sound localization ability, participants who had not experienced the previous experiment were pre-screened in the open-ear condition. As a result, it was confirmed that participant L had front-back confusion even in the open-ear condition. Therefore, the following discussion is based on the data of 12 participants, excluding participant L. The order of experimental conditions was adjusted to avoid bias, e.g., half of the participants performed the unmodified condition first, followed by the modified condition. The remaining participants performed in reverse order.

In addition to accuracy, front-back confusions and lateral errors [9] were used as evaluation metrics. Front-back confusion is defined as the error rate between presentation and response beyond the binaural axis, e.g., the presentation was S1, and the response was S5. Lateral error, which is a

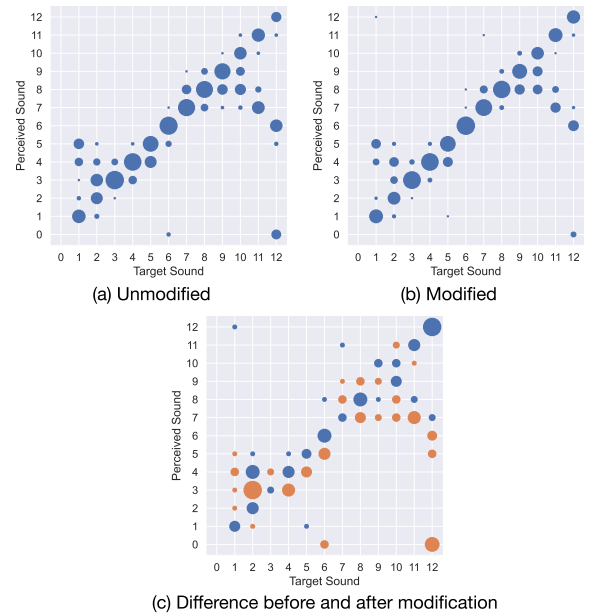


FIGURE 10. Results of improved method: (a) unmodified, (b) modified, and (c) difference before and after modification. Blue indicates the number of responses increased by the modification, and orange indicates the number of responses decreased by the modification.

small localization error, is defined as the root-mean-square error between presentation and response lateral angles, with possible front-back errors disregarded. Questionnaires were also administered.

B. RESULT

1) ACCURACY

Figure 10a and Figure 10b show the results for the two conditions. Although all front-back errors could not be corrected, the decrease in front-back errors and response variability can be confirmed. Figure 10c shows the result obtained by subtracting the number of responses before and after the modification. Blue indicates positive values, and orange indicates negative values. In other words, the modification was effective if the number of blue responses increased along the diagonal line and the number of orange responses increased in other areas. This figure shows the above trend and confirms the effect of the modification.

Figure 11 shows the comparison of accuracy. As shown in this figure, the mean accuracy is improved from 62.7% to 72.0% by applying the improved method. A paired t-test was conducted to confirm that the mean accuracy between conditions was significantly different ($p < .05$). Figure 12 shows accuracy at each sound source. The accuracy improves in 11 of 12 locations by the improved method. The improvement is confirmed, especially in S12. This is because the number of responses mistaking the S12 for S0 decreased, as shown in Figure 10. In contrast, the improvement effect on S3 and S9 was limited. The reason for this is the same in Section III, that is, S3 and S9 are on the binaural axis, which makes both ILD and ITD distinctive, and the degradation was limited.

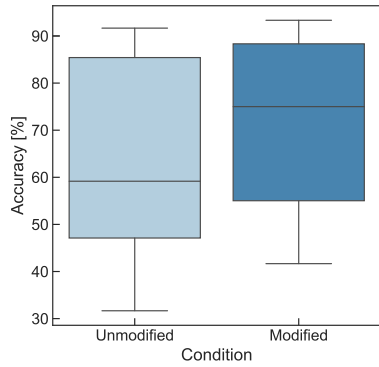


FIGURE 11. Comparison of accuracy.

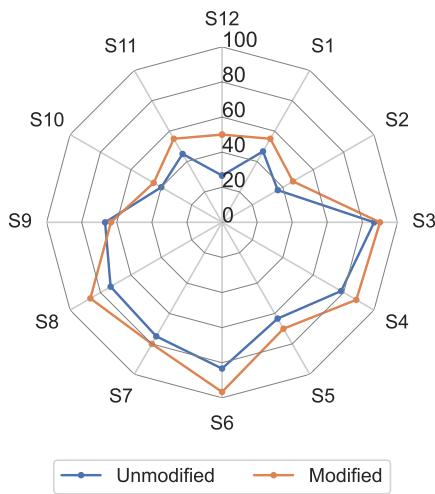


FIGURE 12. Accuracy of each sound source.

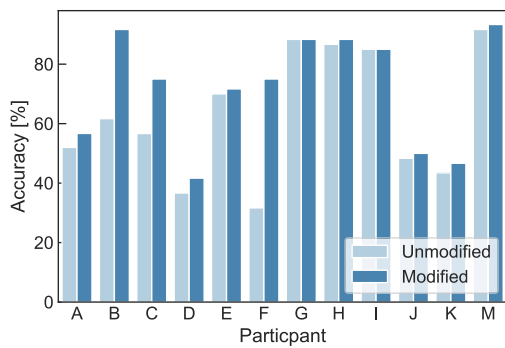


FIGURE 13. Accuracy of each participant.

Figure 13 shows the accuracy of each participant. As shown in the figure, the effectiveness of the proposed method varies depending on the participant. Participants B, C, and F showed significant improvement compared to the other participants; however, participants G, H, I, and M, who have high sound localization ability, did not significantly improve. In all participants, applying the proposed method did not reduce the accuracy.

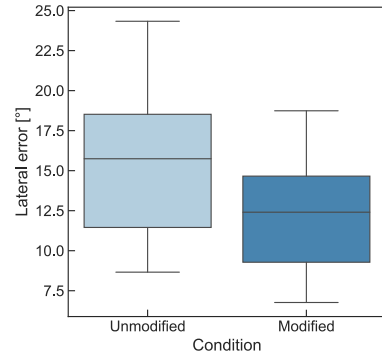


FIGURE 14. Lateral error.

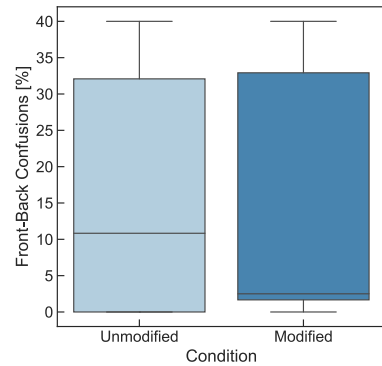


FIGURE 15. Front-back confusions.

2) LATERAL ERROR

Figure 14 shows a lateral error with/without modification. The mean lateral error decreased from 15.6° to 12.0° using the proposed method. A paired t-test confirmed that the mean lateral error significantly differed between conditions ($p < .01$).

3) FRONT-BACK CONFUSION

Figure 15 shows the front-back confusion rate. As shown in the figure, the mean front-back confusion decreases from 16.4% to 13.5%; however, there was no statistically significant difference.

4) QUESTIONNAIRE

Two questionnaire surveys were conducted: the impression of earphone sound under each condition and which condition was easier to localize the sound.

In the first questionnaire, participants were asked how they felt the naturalness of the earphone sound compared to the sound heard by the open-ear condition. Participants were asked for their impressions with a Likert scale (1: unnatural – 5: natural). Figure 16 shows the result of the questionnaire. From the result, it was found that there was no significant difference in the naturalness of sound between conditions, that is, the proposed method did not degrade the impression of the original sound.

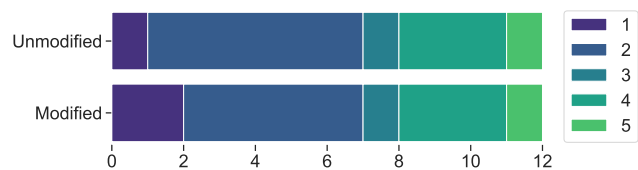


FIGURE 16. Result of questionnaire 1.

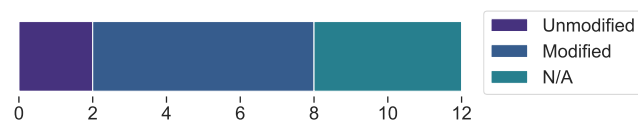


FIGURE 17. Result of questionnaire 2.

In the second questionnaire, participants were asked which condition was easier to localize the sound. Figure 17 shows the result of questionnaire 2. It was found that half of the participants explicitly felt that sound localization became easier. The rest of the participants felt that there were no differences between conditions or that the unmodified condition was better than the modified condition; however, it was confirmed that there was no decrease in accuracy due to the proposed method in all participants, as shown in Figure 13. Especially, although participant F felt that the unmodified condition was better than the modified condition, the most significant improvement (43.3 points) was observed among the participants.

VII. DISCUSSION

A. EFFECT OF IMPROVED METHOD

As described in Section VI, it was confirmed that the improved method could reduce the sound localization degradation caused by the hearable's transparency mode. Lateral errors were also improved by the proposed method. However, as for front-back confusion, the results of this study did not indicate statistically significant improvement, although there was an improvement in the mean rate of front-back confusion. This is because, as described in Section III-A, the cause of the degradation of sound localization is composed of multiple factors, and the proposed method is a countermeasure against one of them, i.e., the frequency response change via the device. To correct other factors, i.e., changes in acoustic characteristics around the auricle and ear canal due to wearing the hearable, it is necessary to measure the user's personalized HRTF when the device is not worn, which may involve measurement costs. Moreover, sound source localization is required to leverage personalized HRTF. The proposed method does not require any special measurements on the user side because the characteristics of each device are considered to be constant. Furthermore, no localization of sound sources is required to utilize personalized HRTF.

The participant's subjective evaluation based on the questionnaire survey confirmed that the proposed method improves the sound localization capability without significant

changes in sound quality compared to the conventional transparency mode.

B. APPLICABILITY OF PROPOSED METHOD

This study aimed to improve the degradation of the transparency mode; however, the improvement method can be applied to other applications as well. For example, in virtual reality, audio augmented reality, and spatial audio, the sound direction is reproduced by considering the HRTF for the sound source to increase immersion. However, the actual sound reproduced may be affected by the frequency characteristics of the speaker on the device, and the reproduced sound may be different from what was initially expected. The introduction of cancellation of frequency characteristics by the equipment may improve the sound localization ability in other applications.

C. INDIVIDUAL DIFFERENCES

Although it was confirmed that the transparency mode of hearables degraded sound localization ability, there were significant individual differences in the results. For example, under the device-wearing conditions in Section III, the average accuracy of participant F was 42.7%, while that of participant C was 88.8%. This may be because there are individual differences in the HRTF, and the adverse effect of transparency mode differs depending on the user.

Also, in Section VI, the effectiveness of the proposed method varied depending on the person. However, all participants showed no effect or improvement; thus, the proposed method was effective.

D. LIMITATION

1) EFFECT OF DELAY

The current implementation introduces a delay, which may lead to degrading sound localization. The previous study pointed out that the delay causes the degradation of sound localization [9]. The accuracy under the unmodified condition in Section VI (62.7%) was lower than the accuracy of the device-wearing condition in Section III (68.9%). Therefore, despite the improvement, there was limited improvement from the existing hearables. In this study, the FFT size was set to 1024 based on previous studies; however, there is a trade-off between delay and sound quality due to the size of FFT [11]. Considering delay and sound quality, the optimal FFT size for the proposed method is a topic for further investigation.

2) ADAPTATION TO TRANSPARENCY MODE

Hofman et al. reported that HRTFs can be relearned in three to six weeks [22]. Thus, if the user always wears a hearable with transparency mode, the problem of sound localization degradation can be solved. However, a hearable is not always in transparency mode but is expected to be used in multiple ways, such as for noise-canceling functions, navigation, and listening to music. Therefore, further investigation is needed

to determine whether the relearning of HRTFs is possible for hearables.

VIII. CONCLUSION

This study investigated the sound localization ability using the transparency mode of commercially available hearables. The results showed that the average accuracy was 91.5% in the open-ear condition, while that was 68.9% in the device-wearing conditions. In order to mitigate the degradation of sound localization, a method was proposed to cancel device characteristics. Evaluation results confirmed that the accuracy of sound localization improved from 62.7% to 72.0%.

REFERENCES

- [1] J. Plazak and M. Kersten-Oertel, "A survey on the affordances of 'hearables,'" *Inventions*, vol. 3, no. 3, p. 48, Jul. 2018.
- [2] P. Crum, "Hearables: Here come the: Technology tucked inside your ears will augment your daily life," *IEEE Spectr.*, vol. 56, no. 5, pp. 38–43, May 2019.
- [3] IDC. (2022). *Wearable Devices Market Share*. [Online]. Available: <https://www.idc.com/promo/wearablevendor>
- [4] J. Blauert, *The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.
- [5] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *J. Audio Eng. Soc.*, vol. 52, no. 6, pp. 618–639, 2004.
- [6] G. Marentakis and R. Liepins, "Evaluation of hear-through sound localization," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, Apr. 2014, pp. 267–270.
- [7] T. Van den Bogaert, T. J. Klases, M. Moonen, L. Van Deun, and J. Wouters, "Horizontal localization with bilateral hearing aids: Without is better than with," *J. Acoust. Soc. Amer.*, vol. 119, no. 1, pp. 515–526, Jan. 2006.
- [8] G. Keidser, A. O'Brien, J.-U. Hain, M. McLelland, and I. Yeend, "The effect of frequency-dependent microphone directionality on horizontal localization performance in hearing-aid users," *Int. J. Audiol.*, vol. 48, no. 11, pp. 789–803, Jan. 2009.
- [9] F. Denk, S. D. Ewert, and B. Kollmeier, "On the limitations of sound localization with hearing devices," *J. Acoust. Soc. Amer.*, vol. 146, no. 3, pp. 1732–1744, Sep. 2019.
- [10] R. Albrecht, T. Lokki, and L. Savioja, "A mobile augmented reality audio system with binaural microphones," in *Proc. Interacting Sound Workshop, Exploring Context-Aware, Local Social Audio Appl.*, New York, NY, USA, Aug. 2011, pp. 7–11.
- [11] H. Watanabe and T. Terada, "Manipulatable auditory perception in wearable computing," in *Proc. Augmented Hum. Int. Conf.*, New York, NY, USA, Mar. 2020, pp. 1–12.
- [12] I. Toshima, S. Aoki, and T. Hirahara, "Sound localization using an acoustical telepresence robot: TeleHead II," *Presence*, vol. 17, no. 4, pp. 392–404, Aug. 2008.
- [13] K. Shirota, R. L. Peiris, and K. Minamizawa, "Altered pinna: Exploring shape change of pinna for perception and illusion of sound direction change," in *Proc. 23rd Int. Symp. Wearable Comput.*, New York, NY, USA, Sep. 2019, pp. 220–224.
- [14] K. Iida, Y. Ishii, and S. Nishioka, "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *J. Acoust. Soc. Amer.*, vol. 136, no. 1, pp. 317–333, Jul. 2014.
- [15] S. Spagnol and F. Avanzini, "Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model," in *Proc. Int. Conf. Digit. Audio Effects*, 2015, pp. 1–12.
- [16] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry," *J. Acoust. Soc. Amer.*, vol. 137, no. 2, pp. 690–701, Feb. 2015.
- [17] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry," *J. Acoust. Soc. Amer.*, vol. 140, no. 2, pp. 814–831, Aug. 2016.
- [18] Sony. (2023). *Sony 360 Reality Audio*. [Online]. Available: <https://electronics.sony.com/360-reality-audio>
- [19] Apple. (2023). *Listen With Personalized Spatial Audio for AirPods and Beats*. [Online]. Available: <https://support.apple.com/en-us/HT213318>
- [20] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. Leiden, The Netherlands: BRILL, 2012.
- [21] R. A. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane," *J. Acoust. Soc. Amer.*, vol. 61, no. 5, pp. 1264–1269, May 1977.
- [22] P. M. Hofman, J. G. A. Van Riswick, and A. J. Van Opstal, "Relearning sound localization with new ears," *Nature Neurosci.*, vol. 1, no. 5, pp. 417–421, Sep. 1998.



HIROKI WATANABE received the B.E., M.E., and Ph.D. degrees in engineering from Kobe University, Hyogo, Japan, in 2012, 2014, and 2017, respectively.

From 2015 to 2017, he was a Research Fellow of the Japan Society for the Promotion of Science (DC2), Kobe University. Since 2017, he has been an Assistant Professor with the Graduate School of Information Science and Technology, Hokkaido University, Hokkaido, Japan. His research interests include wearable computing and ubiquitous computing.



TSUTOMU TERADA (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in engineering from Osaka University, Osaka, Japan, in 1997, 1999, and 2003, respectively.

From 2000 to 2004, he was an Assistant Professor with the Cybermedia Center, Osaka University, where he was a Lecturer, from 2005 to 2007. From 2007 to 2018, he was an Associate Professor with the Graduate School of Engineering, Kobe University, Hyogo, Japan, where he has been a Professor, since 2018. His research interests include wearable computing, ubiquitous computing, and entertainment computing.

• • •