Exploring Visual-Auditory Redirected Walking Using Auditory Cues in Reality

Kumpei Ogawa, Kazuyuki Fujita, Shuichi Sakamoto, Kazuki Takashima, and Yoshifumi Kitamura

***Abstract*—We examine the effect of auditory cues occurring in reality on redirection. Specifically, we set two hypotheses: the auditory cues emanating from fixed positions in reality (Fixed sound, FS) increase the noticeability of redirection, while the auditory cues whose positions are manipulated consistently with the visual manipulation (Redirected sound, RDS) decrease the noticeability of redirection. To verify these hypotheses, we implemented an experimental environment that virtually reproduced FS and RDS conditions using binaural recording, and then we conducted a user study ($N = 18$) to investigate the detection thresholds (DTs) for rotational manipulation and the sound localization accuracy of the auditory cues under FS and RDS, as well as the baseline condition without auditory cues (No sound, NS). The results show, against the hypotheses, FS gave a wider range of DTs than NS, while RDS gave a similar range of DTs to NS. Combining these results with those of sound localization accuracy reveals that, rather than the auditory cues affecting the participants' spatial perception in VR, the visual manipulation made their sound localization less accurate, which would be a reason for the increased range of DTs under FS. Furthermore, we conducted a follow-up user study ($N = 11$) to measure the sound localization accuracy of FS where the auditory cues were actually placed in a real setting, and we found that the accuracy tended to be similar to that of virtually reproduced FS, suggesting the validity of the auditory cues used in this study. Given these findings, we also discuss potential applications.**

***Index Terms*—Redirected walking, room-scale VR, visual-auditory redirection.**

I. Introduction

ROOM-SCALE virtual reality (VR) enables users wearing a head-mounted display (HMD) to freely explore a virtual environment (VE) with their own physical movement (e.g., walking), but the fundamental limitation is that it requires a large physical space. Redirected Walking (RDW) [1] is a promising methodology for overcoming this limitation. The basic idea of RDW is to imperceptibly manipulate the user's actual walking direction by slightly differentiating the mapping of their vision between reality and VR, which exploits the phenomenon that vision often dominates proprioception and vestibular sensation when they disagree [2]. One common approach to RDW is to apply a gain to the walking distance, curvature, or rotation angle in VR and explore the gain range within which it is imperceptible to the user (i.e., detection thresholds: DTs). Yet most of the current RDW techniques are not sufficient for use in a physical space of practical size (e.g., Steinicke reported that a space with a 22-m radius is required [3]), and researchers have been actively working on improving RDW by compressing the physical space required for the targeted experience.

Along with vision, audition also plays an important role in the human capability of spatial perception. Accordingly, to enhance the effect of RDW, researchers have explored visual-auditory redirection techniques in which auditory feedback is combined with vision. A straightforward approach is to provide the user with spatialized auditory feedback from virtual object(s) that is consistent with visual manipulation [4], [5], [6]. This approach is based on a phenomenon in which the consistency of the auditory and visual cues makes it more likely to perceive the rotation of the entire VE as self-motion [7]. However, the results of those studies show a limited effect of auditory cues; while one study showed an expanded range of DTs [4], two others showed no differences in DTs between usage with and without auditory feedback [5], [6]. Nilsson et al. [5] suggested that the reason for such a limited effect is that vision is generally superior to audition when estimating spatial locations of objects [8]. In response, Gao et al. [9] attempted to lessen this effect by manipulating the reliability of vision and audition, and they reported that curvature manipulation was less noticeable with incongruent visual-auditory cues. Accordingly, previous efforts have revealed both the effectiveness and limitations of visual-auditory RDW techniques.

In contrast to the above approaches to using audio *in VR*, we focus on the effect of auditory cues *in reality* on the user's perception of RDW. This idea is motivated by our assumption that VR users may infer their actual position and/or direction from external auditory noises emanating from a fixed location in reality around the users (e.g., TV sounds or operating noises of a washing machine in a home's room) (Fig. 1). We believe that the use of real-space auditory cues would offer the following two potential benefits. First, real-space auditory cues (Fig. 1(b)) are different from those of VR (Fig. 1(a)) in that the user cannot visually recognize the position of the sound source, which does not apply to the above-mentioned effect that vision superiors to

Manuscript received 31 March 2023; revised 24 July 2023; accepted 16 August 2023. Date of publication 28 August 2023; date of current version 1 July 2024. This work was supported by JSPS Grant KAKENHI under Grant 22H00523. Recommended for acceptance by F. Steinicke. *(Corresponding author: Kazuyuki Fujita.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Review Committee for Human-Subject Research at Research Institute of Electrical Communication, Tohoku University.

The authors are with the Research Institute of Electrical Communication, Tohoku University, Sendai, Miyagi 980-8577, Japan (e-mail: kumpei.ogawa.r5@dc.tohoku.ac.jp; k-fujita@riec.tohoku.ac.jp; saka@ais.riec.tohoku.ac.jp; takashima@riec.tohoku.ac.jp; kitamura@riec.tohoku.ac.jp).

This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2023.3309267, provided by the authors.

Digital Object Identifier 10.1109/TVCG.2023.3309267

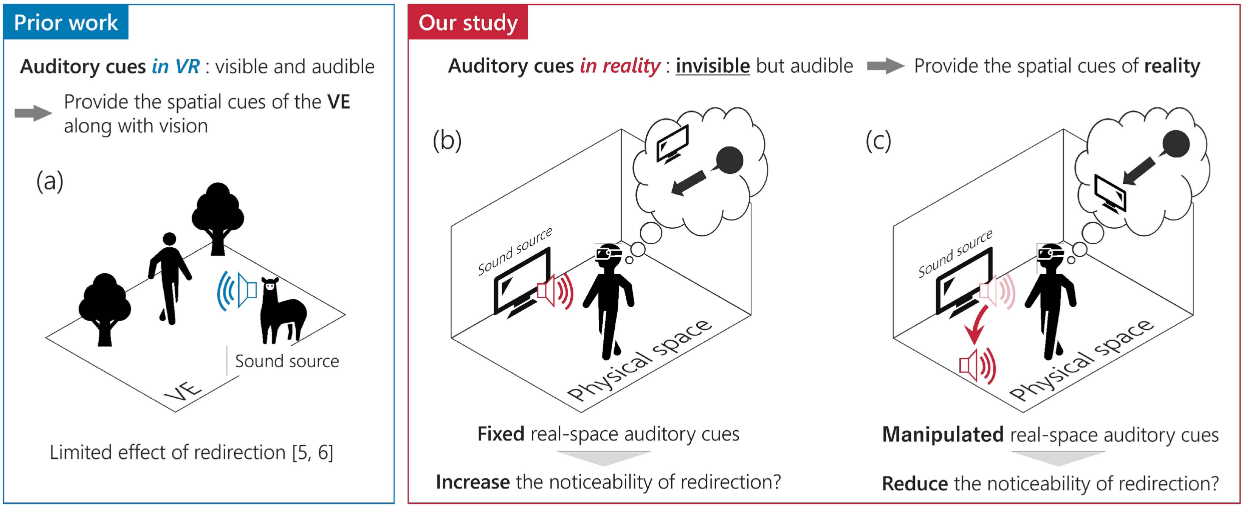

Fig. 1. Conceptual illustrations showing differences between auditory cues *in VR* (a) and *reality* (b)–(c) during the VR experience. (a) Most prior work focuses on auditory cues *in VR*, whose source objects are also visible; the auditory cues will provide spatial cues of the VE along with vision. (b)–(c) In contrast, this study focuses on auditory cues *in reality*, which are invisible during the VR experience and are supposed to be immobile. Therefore, (b) the auditory cues in reality will provide spatial cues of reality, potentially increasing the noticeability of redirection. Correspondingly, (c) manipulating the positions of the auditory cues may reduce the noticeability of redirection.

audition, and thus different results are expected from most of the existing visual-auditory RDW studies. Second, fixed real-space auditory cues can induce the user's preconception that the position of the sound source does not change, suggesting that manipulating its position (Fig. 1(c)) might lead to a less noticeable redirection (based on a study reporting that auditory cues from immobile objects could more strongly induce self-motion [10]). Despite these potential benefits, no previous study has investigated the effect of real-space auditory cues on RDW.

Therefore, in this study, we address the following two research questions as an initial attempt to investigate the effect of real-space auditory cues on redirection.

RQ1 Does presenting real-space auditory cues increase the noticeability of rotational RDW?

RQ2 Does manipulating the positions of real-space auditory cues reduce the noticeability of rotational RDW?

To address these questions, we implemented an experimental environment in which the perceived direction of real-space sound sources (i.e., ticking sounds of a metronome) could be manipulated by spatialized audio using binaural recording. With this environment, we conducted a user study ($N = 18$) to investigate the noticeability of rotational manipulation of RDW under three auditory conditions: No sound (NS), where no auditory cues were presented, Fixed sound (FS), where real-space auditory cues with fixed locations were presented, and Redirected sound (RDS), where real-space auditory cues were manipulated to make them consistent with visual rotational manipulation. We measured the participants' noticeability of redirection as well as the accuracy of sound localization by estimating the direction of the sound source.

The results showed, surprisingly, that the answer to both RQ1 and RQ2 was no. Regarding RQ1, the range of the DTs was significantly wider in FS than in NS, indicating that presenting the real-space auditory cues makes redirection less noticeable. Regarding RQ2, we did not find any difference in the ranges of the DTs between RDS and NS, indicating that manipulating the positions of the real-space auditory cues does not reduce the noticeability of redirection. In addition, after combining these results with those of sound localization accuracy, it was revealed that the user's perception of visual manipulation caused a shift in the sound localization of the auditory cues, which would lead to a less unnoticeable redirection in FS.

Furthermore, we conducted a follow-up experiment to measure the sound localization accuracy in FS where the sound sources were actually placed in a real setting, in order to compare it with the binaural-recorded FS used in the main experiment. The results show that the tendency of sound localization accuracy according to gain was generally aligned between the two conditions, suggesting that our binaural-recorded system had reasonably reproduced the real auditory space. Based on these findings, we discuss possible applications.

II. Related Work

A. Redirected Walking

RDW is a methodology of achieving locomotion with real walking in a large VE within a limited physical space [1]. It is often achieved by applying gain (i.e., upscale/downscale ratio of the amount of physical movement in VR) to the walking motion. The effectiveness of RDW is frequently discussed using DTs, which are metrics of noticeability estimated by psychophysical experiments using two-alternative forced choice (2AFC) tasks, where the participants are asked to answer whether the movement in the VE is larger or smaller than in reality. Steinicke et al. [3] first performed this type of experiment, and the estimated DTs showed that the users were less likely to notice the virtual rotation between the gains of 0.67 and 1.24. User perception to redirection is known to be affected by various factors other

TABLE I
SUMMARY OF PRIOR RDW STUDIES INVESTIGATING THE EFFECT OF AUDITORY CUES (THE STUDY OF STEINICKE ET AL. [3] IS FOR COMPARISON)

Gain Type	Source	VE	Audio Source	Audio Generation Method	Detection Threshold
Rotation	Steinicke et al. [3]	City	N/A	N/A	0.67 - 1.24
	Serafin et al. [11]	Blindfolded	Alarm clock	16 speakers (VBAP)	0.82 - 1.2
	Nogalski et al. [12]	Blindfolded	Barking dog	208 speakers (WFS)	N/A - 1.177
	Nilsson et al. [5]	Bridge, forest	Waterfall	16 speakers (VBAP)	0.77 - 1.10 (No audio) 0.80 - 1.11 (Static audio) 0.79 - 1.08 (Moving audio)
	Meyer et al. [4]	Desert w/oasis, etc.	Barking dog, etc.	208 speakers (WFS)	0.68 - 1.36
	Junker et al. [6]	Ruin, river, etc.	River, waterfall, etc.	Headphones	No audio: 0.77 - 1.13 (no fog) 0.78 - 1.09 (low fog) 0.81 - 1.11 (high fog) Spatialized audio: 0.78 - 1.12 (no fog) 0.75 - 1.09 (low fog) 0.80 - 1.09 (high fog)
Curvature	Steinicke et al. [3]	City	N/A	N/A	22 m
	Serafin et al. [11]	Blindfolded	Alarm clock	16 speakers (VBAP)	16 m
	Nogalski et al. [12]	Blindfolded	Ice cream truck	208 speakers (WFS)	3.64 m
	Meyer et al. [4]	Desert w/oasis, etc.	Fountain, etc.	208 speakers (WFS)	6.0 m
	Gao et al. [9]	Desert, train	Engine sound	Headphones	Congruent cues: 21.2 m (without fog) 29.2 m (with fog) Incongruent cues: 12.1 m (without fog) 9.8 m (with fog)

than gain (e.g., gender [13], [14], [15], HMD's field of view [14], and the complexity of the VE [16], [17]), and researchers have been actively exploring a wide range of approaches (obviously, not all of the vast amount of relevant research can be presented here, so please refer to the review articles, e.g., [18], [19] for details). However, current RDW techniques are still not able to compress the targeted experience within a practical size of physical space (e.g., 22-m [3] or 6.4-m [20] space is required to infinitely walk straight), thus further improvements in the efficacy of RDW or better combinations of several techniques would still be needed.

B. Auditory Space Perception

Auditory information, as well as visual information, includes rich spatial information. Humans can extract the spatial information using their two ears. The difference between the binaural signals is used as a strong cue to localize a sound on the horizontal plane [21]. Since humans use both the left and right ears for hearing, the arrival time and the sound pressure level of a heard sound differ between the two ears, where these discrepancies are called interaural time difference (ITD) and interaural level difference (ILD), respectively. Furthermore, various factors are known to contribute to the perception of the distance from sound sources, such as the sound pressure level [22], the sound reflection [23], and the timbre [24].

As clear evidence that auditory information includes spatial information, auditory vection has been reported by various researchers. Vection originally refers to the self-motion sensation induced by moving visual stimuli [2], but auditory stimuli was also observed to induce self-motion [25]. The intensity of auditory vection is known to be affected by various factors, such as the number of sound sources [10] and the presentation of vibrations [26]. It is also known that sounds emanating from objects that are generally considered immobile (e.g., fountains) have a higher effect of inducing auditory vection [10].

The effect of audition on spatial perception has also been studied in the context of RDW. Table I shows relevant prior studies that examined the effect of auditory cues on RDW, and the following works in this subsection are those that focus on auditory redirection without vision. Serafin et al. [11] and Nogalski et al. [12] investigated how blindfolded users could be redirected using auditory cues. Serafin et al. [11] showed that the users' perceived direction and walking path curvature could be manipulated by controlling the position or direction of the sound source, although the manipulation was easily noticeable compared to visual-only RDW. Nogalski et al. [12] conducted a similar experiment using a different method of generating spatial audio and found that the user's walking path (curvature) and rotation angle could be manipulated, while the DTs were highly affected by the prior knowledge of the experiment. These studies are essentially different from ours in that they eliminate the influence of vision, but they still show the potential of auditory cues to enhance the effect of RDW.

C. Visual-Auditory Redirected Walking

Researchers have also examined the multimodal effect of combining visual and auditory cues for redirection to establish visual-auditory RDW. A frequent approach is to use audition as a complement to vision, that is, to increase the reliability of visual cues by providing auditory cues of virtual objects that are consistent with the visual manipulation. Several studies have attempted to demonstrate the effectiveness of this approach with rotational and curvature manipulation ([4], [5], [6], as shown in Table I), but their results are inconsistent. While Meyer et al. [4] reported that using visual-auditory cues gives a wider range of DTs than visual-only RDW [3], Nilsson et al. [5] and Junker et al. [6] reported no difference in detection threshold between usage with and without auditory cues. These results indicate that auditory cues have a limited effect on RDW when combined with visual cues. As a possible reason for this, Nilsson et al. [5] mentioned a phenomenon where vision is generally superior to audition when estimating the spatial location of objects [8].

More recent studies have explored reducing the potency of the high reliability of vision while seeking the optimal conditions of visual-auditory RDW. Junker et al. [6] attempted to lower the reliability of vision by adding fog to limit the visibility of the user's view. However, their results suggest that visibility does not affect the DTs for RDW. Gao et al. explored visual-auditory redirection based on maximum likelihood estimation [27], a multimodal integration model. Specifically, they attempted to manipulate the user's sensory reliability of vision and audition by presenting a virtual object with incongruent visual-auditory cues and by limiting the view's visibility using fog. Their results show that such incongruent visual-auditory cues can make curvature manipulation less noticeable, especially when using fog, and suggest that using incongruent multimodal cues is a promising application.

Aside from the approaches using gains, some studies have explored unique visual-auditory RDW techniques used for specific scenarios. Feigl et al. [28] used construction noise sound, to which people are likely to avoid listening, leading users to bend the walking path. Weller et al. [29] proposed a technique that manipulates the user's walking direction by changing the type of footstep sound in VEs with varied types of ground (e.g., grass and wet surfaces). Another approach is to add auditory cues to distractor-based RDW, which provides visual stimuli to take the user's attention away from redirection. Rekowski et al. [30] attempted this approach and showed that the distractors work effectively when combined with auditory feedback. Lee et al. [31] attempted to use olfactory stimuli in addition to auditory stimuli to make the distractors more effective. However, these approaches rely on specific scenarios or VEs, and thus they have limited applicability.

As described above, many approaches to combining visual-auditory cues for redirection have been studied, with the results revealing their effectiveness and limitations. However, the prior studies have only focused on auditory cues in VR, and there has been no study investigating their effect in reality. Since real-space auditory cues are invisible to the VR user, visual dominance over audition may not apply, leading us to expect different results from most of the conventional visual-auditory RDW studies. Moreover, since real-space sound sources are ubiquitous in our common VR play areas (e.g., a living room), they might be widely applicable regardless of VR content or narrative.

III. Experiment

We conducted an experiment to investigate how the user noticeability of rotational RDW is affected by presenting real-space auditory cues (RQ1) and by manipulating their emanated position (RQ2). The reason for choosing rotational manipulation is that it would have a greater influence on auditory cues for spatial perception, such as ITD and ILD, than translation and curvature manipulation. Another reason is that achieving translation/curvature manipulation in our approach requires a huge number of recording samples (i.e., hundreds of thousands) to reproduce the auditory space in arbitrary user positions and orientations, since we confirmed that software-based audio spatializers are insufficient (described in detail in 3.2). In the experiment, we asked the participants to rotate their body at a certain angle under a certain rotation gain manipulation. We then asked them to express their perception of the amount of rotation in VR compared with that in reality and their perceived direction of the auditory cue. Next, we measured the detection thresholds and the accuracy of sound localization under each auditory condition. We describe the design, procedure, and results of the experiment below.

TABLE II
Sensory Feedback Presented in Each Condition

	Vision	Audition	Somatic sensation
NS	Manipulated	(Not presented)	Unmanipulated
FS	Manipulated	Unmanipulated	Unmanipulated
RDS	Manipulated	Manipulated	Unmanipulated

Note that the auditory and somatic sensation are congruent in FS, while the vision and audition are congruent in RDS (underlined).

A. Experimental Design

The experiment was a two-factor within-participant design with the independent variables of auditory condition and rotation gain. Regarding the auditory condition, we prepared the following three conditions in order to address RQ1 and RQ2.

- *No sound (NS):* The baseline condition where no auditory cues were presented.
- *Fixed sound (FS):* The condition where real-space auditory cues with fixed locations were presented. This condition assumes that the sound source objects are simply placed somewhere in reality, implying that the amount of rotation inferred from the auditory cues is consistent with somatosensory perception. Comparing the results of this condition with NS would address RQ1.
- *Redirected sound (RDS):* The condition where real-space auditory cues were presented and manipulated to be consistent with the visual manipulation. In other words, under RDS, the amount of rotation inferred from the auditory cues is consistent with vision in VR. Such a condition does not occur in reality, and it was achieved with our custom-made system that virtually reproduces a given position of an auditory cue using binaural recording (see Section III-B2 for details). Comparing the results of this condition with NS would address RQ2.

Table II summarizes which sensory feedback was manipulated in each auditory condition, and Fig. 2 gives an overview of the visual and auditory movement relative to participant rotation for each auditory condition. For more details of each condition, please refer to the supplemental video where we visually described the movement of the auditory cues accordingly with the user rotation.

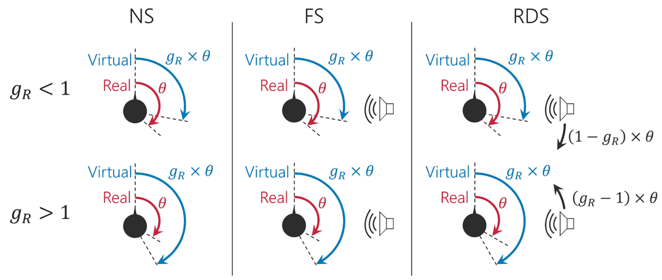

Fig. 2. Example of visual and auditory movement relative to participant rotation (represented as θ) for each auditory condition. g_R denotes the rotation gain (1).

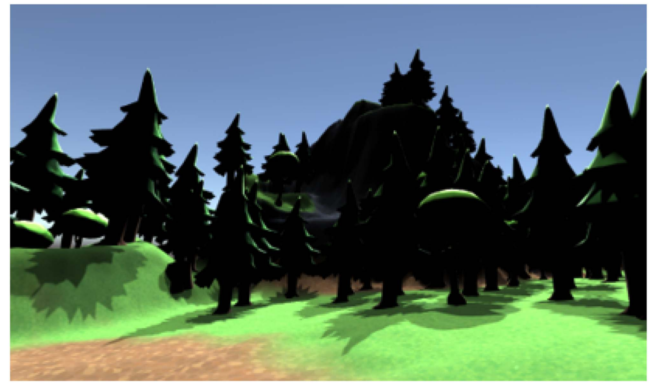

Fig. 3. VE used in the experiment.

We used the ticking sounds of a metronome as the real-space auditory cue that was assumed to be irrelevant to the VR experience. The reason for choosing the metronome was that it involves physical movement (i.e., oscillation of the pendulum bar) in the generation of ticking sound, which would make the participants more aware that the sound is coming from the physical object. At the beginning of the trial, the sound was presented from 90 degrees, either left or right, with a 1-m distance, and we instructed participants to rotate toward the sound of the auditory cue. This instruction was given to make the participants reasonably aware of the position of the metronome. These directions of the sound were used because it would be easier for the participants to perceive the direction of the sound source than when it comes from the front or back direction. The metronome sound was played at a tempo of 60 times per minute.

We applied rotation gains ranging from 0.6 to 1.4 at an interval of 0.1 (nine gains in total), referring to several prior studies (e.g., [14], [32]). The experiment was divided into three blocks for each auditory condition. In each block, a total of 36 trials (9 gains $\times$ 4 repetitions) were performed. The four repetitions included two trials each in the clockwise and counterclockwise rotational directions. The number of repetitions was determined based on our preliminary experiment with three participants, considering the duration of the experiment and participants' fatigue. Furthermore, the order of auditory conditions was counterbalanced among the participants.

Regarding the participants' rotation angle in each trial, we set this to a random value between 100 and 130 degrees, as used by Congdon et al. [33]. This was because we observed in the preliminary experiment that using a fixed rotation angle (set to 90 degrees) led to the participants inferring the amount of rotation from cues irrelevant to somatic sensations, such as memorizing the number of steps.

In each trial, we asked participants to give their perceived rotational amount using the 2AFC method and perceived direction of the auditory cue (i.e., metronome sound), which led to measuring DTs and the accuracy of sound localization, respectively. The reason for measuring the accuracy of sound localization was to investigate how visual manipulation and the auditory conditions affect the participant's auditory space perception. Another reason was to make the user reasonably aware of the auditory cues. At the end of each block, we employed the Slater-Usoh-Steed Presence Questionnaire (SUS PQ [34]) to measure the subjective sense of presence, as well as an open-ended questionnaire about the general impression of the experience in each block (i.e., auditory condition). In the questionnaire for the last block, we additionally asked about the impression throughout the three blocks, including the degree of perceived simulator sickness, how the auditory cue affected the VR experience, and whether the participants noticed that the auditory cues were virtual (we did not employ SSQ [35] in order to avoid lengthening the experimental time). The experiment was officially approved by our university's ethics committee.

B. Implementation

1) Visual Stimuli: We presented a forest VE as visual stimuli to cue the participants' rotation, as shown in Fig. 3. The algorithm for the visual manipulation we used has been commonly used in many RDW studies (e.g., [3]), in which the rotation gain (g_R) is represented by the division of rotation angle in VR ($R_{virtual}$) by that in reality (R_{real}), as shown in the equation below.

$$g_R = \frac{R_{virtual}}{R_{real}} \tag{1}$$

A gain value larger than 1 ($g_R > 1$) means that the rotation in VR is larger than in reality, which would lead to the user's perceiving the visual rotation as faster, and vice versa.

2) Auditory Stimuli: The key to this experiment was how the auditory cue was generated due to the introduction of RDS; in that condition we needed to manipulate the perceived positions of auditory cues while making the participants believe the position was fixed in reality. Since it was not realistic to physically move the source of the auditory cue (i.e., metronome), we adopted a method to control it virtually using spatialized audio. We first implemented several software prototypes that generate the spatialized audio using Unity and its plugins (e.g., Oculus Spatializer, Steam Audio), which are based on a generalized head-related transfer function (HRTF). However, our informal tests with three testers suggest that the implemented auditory cues in these prototypes could not offer a sufficient quality of spatialization. More specifically, the testers stated that the generated auditory cues were not judged as very spatialized and were clearly distinguishable from the original metronome sound. This was probably due to the individual differences in HRTF and/or the limitations of the engines of spatialized audio generation.

Therefore, to reproduce the sound of the auditory cue with more reality, we implemented our custom-made system of generating virtual auditory cues using binaural recording. We adopted binaural recording because it could inherently reproduce the original sound field recorded. To provide appropriate auditory cues according to the participant's head rotation angles, our system successively selected one of the binaural sound sources pre-recorded at the angle closest to the participant's current head orientation, and this was done without performing audio interpolations. Specifically, we recorded 36 binaural sound sources (every 10 degrees) using a pair of binaural microphones (Roland CS-10EM) worn in each ear of the experimenter in the same environment used in the experiment. The recording interval (i.e., 10 degrees) might be a bit too sparse given the human sound localization ability [36], but it would be adequate considering that the sound localization accuracy decreases during self-motion [37] and that many of the studies in acoustic engineering investigating sound localization accuracy have employed an interval greater than 10 degrees (e.g., [38]). The volume of the replayed sound was set to be consistent among participants, and its level was determined so that the metronome sounds were clearly heard and considerably louder than the ambient noise of the experimental room (47.5 dB, obtained by a sound level meter, ONO SOKKI LA-5110). From the three testers in another informal test, we obtained the response that the generated sound was felt sufficiently spatialized and was almost indistinguishable from the real one.

Although we originally built the sound generation system to achieve RDS, we also used the same system in FS to make the hearing of the sound as consistent as possible. FS was implemented by switching the binaural sound sources at the corresponding angle based on the actual orientation the user's head (i.e., HMD), while RDS was also implemented based on the orientation of the user's view in VR.

C. Apparatus

Fig. 4 shows the apparatus used in the experiment. The participants wore an HMD (Meta Quest 2) and held a controller in their right hand. The HMD had a resolution of $1,832 \times 1,920$ per eye and a refresh rate of 72 Hz. We used Unity (Ver. 2019.3.6) for our software implementation. The experimental program was run on the experimenter's PC (Windows 10 Education), and the PC and HMD were wirelessly connected using Air Link. In addition, the participants wore a pair of earphones (Apple AirPods Pro) to present the auditory cue. The noise-canceling functionality of the earphone was turned off because it would be unnatural for surrounding noises other than the auditory cues to disappear. The size of the tracking space was 2.25 m $\times$ 3.00 m, and the participants stood in the middle of the space. A 1-meter-high platform was placed on either side of the participant, and a metronome was placed on top of each. The horizontal distance between the participant and each platform was set to 1 m. The experimenter and an assistant were seated near the metronomes to convince the participants that the metronomes were operated by them during trials (in fact, the start/stop of the metronome's sound was controlled by the experimenter's keyboard, since the sound was virtual). A video camera was placed in the corner of the environment, and we recorded the entire experiment with the participants' consent.

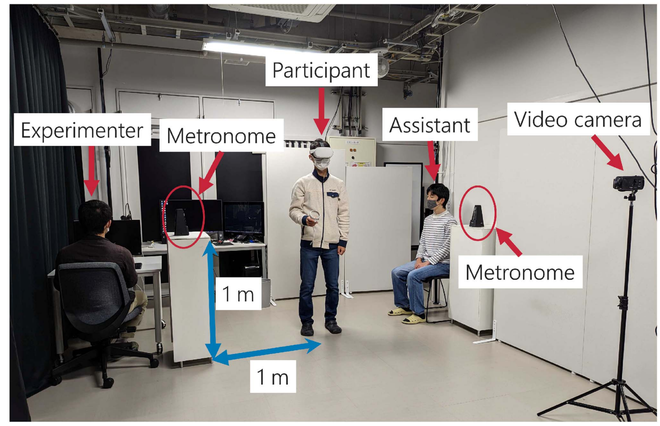

Fig. 4. Experimental setup.

To prevent COVID-19 infection, the experimenter, the assistant, and the participant kept a sufficient distance from each other and wore masks during the experiment. In addition, we attached a replaceable silicon cover to the HMD, and the participants wore a disposable VR mask for use with the HMD. The devices used in the experiment (e.g., silicon cover, controllers, PC) were disinfected before and after the experiment.

D. Participants

A total of 18 university students (12 males and 6 females) participated in the experiment. The mean age of the participants was 21.5 ($SD = 1.04$). All participants had normal or corrected vision and hearing (self-reported). As for the participants' prior VR experience, eight had no experience, six had fewer than five times, and four had more than five times. We instructed the participants on the basic idea of RDW before the experiment, but they had no prior knowledge of this study. Each participant was paid for their participation (approx. 20 USD) in accordance with university regulations.

E. Procedure

We first guided the participants to the experimental space and then explained the basic idea of redirection techniques, followed by an overview of the experiment. We next asked them to sign an informed consent form. After that, we instructed them in the detailed procedure of the experiment, explaining that the metronome sound emanated from the physical space while activating it. We then asked the participants to wear an HMD and a pair of earphones. We explained that the reason for wearing earphones was to hear the sound feedback when pressing the controller's button during trials; accordingly, the true reason for listening to the virtual auditory cues was not conveyed to them.

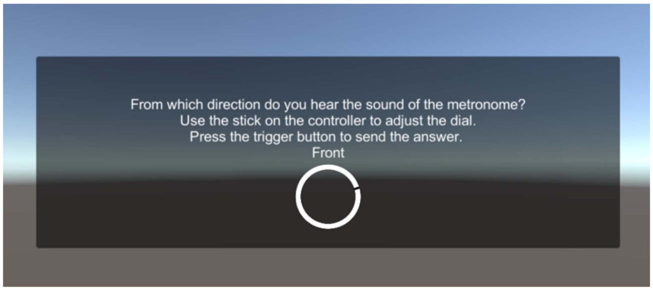

Fig. 5. Message window showing the question and dial-type UI for the sound localization task.

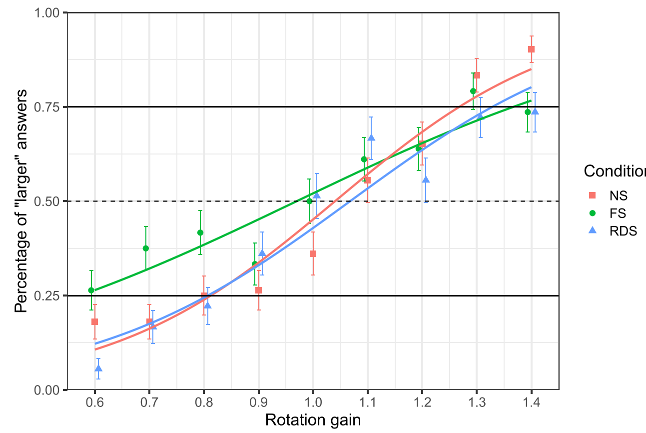

Fig. 6. Fitted psychometric curves in each auditory condition.

After giving the above explanation of the experiment, we started the practice phase. The purpose of this phase was to help the participants get accustomed to the trials and understand the meaning of each question in the trials. In this phase, the participants experienced the trial under NS with a clearly noticeable gain (i.e., 0.5 and 1.5), and the trial could be repeated for as many times as the participant wanted. For the rotation, we instructed them to rotate the entire body as slowly as possible to prevent rotating more than the specified angle (i.e., a random value between 100 and 130 degrees in VR). We did not record the answers to the questions in the practice phase. After the phase was finished, we moved on to the experiment phase.

Each trial included the following steps. First, after confirming the participant's safety, the experimenter started the trial by pressing a key on the keyboard. Immediately after that, the experimenter started playing the metronome sound by pressing another key. When the trial began, the forest VE and an instruction message showing the rotational direction was displayed. In NS, the direction of the participants' rotation was indicated by showing an arrow with the instruction message. In FS and RDS, the participants were instructed to rotate in the direction of the metronome sound heard, which was either the left or right direction at 90 degrees to the participants' initial orientation. The participants could start the trial at their own timing by pressing a button on the controller. After the participants pressed the button, the instruction message disappeared, and they started rotating. Their whole body slowly rotated until the view changed from the forest VE into an empty space without any directional cues (i.e., only horizon and sky) with a message window showing a question from the questionnaire. The participants then answered the question using the controller. The questions consisted of one about the sound localization (except for NS), followed by one about the redirection noticeability. The exact question about the sound localization was "From which direction do you hear the sound of the metronome?" They answered with the direction of the auditory cue relative to their facing direction using a stepless dial-type user interface (Fig. 5). The question about redirection noticeability was "Was the rotation in VR larger or smaller than in reality?," and the participants had to answer either "larger" or "smaller." After answering these questions, they turned back to the initial orientation, while the VE remained empty. The trial was then finished and the metronome sound was stopped.

The experiment phase was divided into three blocks for each auditory condition. After finishing each block, the participants took off the HMD and answered the questionnaire. All of the questionnaires were administrated through Google Forms. After completing the questionnaire, they could take a five-minute break before the next block started. The entire experiment took about 90 minutes per participant.

F. Hypotheses

We set the following three hypotheses for the experiment.

- H1 The range of DTs under FS is narrower than that under NS;
- H2 The range of DTs under RDS is wider than that under NS;
- H3 The sense of presence under NS is higher than that under the other two conditions.

We set H1 (corresponding to RQ1) because we thought that FS's unmanipulated auditory cue would help participants infer their facing direction in reality, which would lead to a higher noticeability of redirection. We set H2 (corresponding to RQ2) because we thought that RDS's auditory cue, which was manipulated consistently with the visual manipulation, would increase the sense of feeling the rotation in the VE as real, which in turn would lead to a lower noticeability of redirection. We set H3 because we thought the auditory cue, which was irrelevant to the VR contents, could negatively affect the quality of the experience.

G. Results

1) Detection Thresholds: All participants successfully completed all trials, and a total of 1,944 responses for the redirection noticeability were obtained without any loss. The data were calculated as the percentage of responses saying that the rotation in VR was "larger" than that in reality, and they were fitted into a logistic function $y = 1/(1 + \exp(-ax + b))$ as a psychometric function. Fig. 6 shows the psychometric curves of each auditory condition (error bars indicate standard errors). In addition, we estimated the lower and higher limits of DTs (25% threshold DT_L and 75% threshold DT_H) and the point of subjective

TABLE III
ESTIMATED DTs AND PSE IN EACH AUDITORY CONDITION

	DT_L	PSE	DT_H	ΔDT
NS	0.81 [0.75, 0.87]	1.04 [0.99, 1.10]	1.27 [1.20, 1.31]	0.46
FS	0.57 [0.47, 0.72]	0.97 [0.90, 1.02]	1.37 [1.22, 1.46]	0.79
RDS	0.81 [0.76, 0.86]	1.07 [1.02, 1.11]	1.33 [1.26, 1.39]	0.52

[ ] indicates 95% CI.

equality (PSE) as well as their 95% confidence intervals (CIs) in each auditory condition using bootstrapping (100 repetitions for resampling). We also defined and calculated ΔDT as a metric to indicate the range of DTs in each auditory condition.

$$\Delta DT = DT_H - DT_L \tag{2}$$

Table III summarizes the estimated values related to the auditory conditions. We found the range of DTs (ΔDT) was wider in the order of FS, RDS, and NS. Regarding the lower limit of DT (DT_L), FS gave a significantly lower value than the other two conditions, and RDS and NS gave almost equivalent values. As for the higher limit of DT (DT_H), the mean value was larger in the order of FS, RDS, and NS, although no significant difference was observed between any of the three conditions considering the CIs. These results suggest that FS decreased the noticeability of redirection, which was contrary to H1. Moreover, RDS was shown to give equivalent noticeability to NS, which did not support H2.

2) Sound Localization Accuracy: A total of 1,294 responses for the questions on sound localization were obtained (two responses in RDS were excluded from the analysis because of data acquisition failure due to the participants' operation error). We first analyzed the *absolute* angular error of sound localization, defined as the difference in the angles between what the system actually presented and what the participants answered. The mean value of the absolute angular error was 21.3 degrees ($SD = 6.7$) in FS and 17.1 degrees ($SD = 3.5$) in RDS. A Wilcoxon signed-rank test revealed that the absolute angular error in FS was significantly greater than that in RDS ($r = .57,\ p < .05$).

To gain a deeper insight into the results of sound localization, we then examined the *relative* angular error, defined as a signed value of sound localization error where the participant's rotational direction in each trial was set to be positive. Fig. 7 shows the mean relative angular error related to each gain for each auditory condition. From this graph, the relative angular error in FS appeared to be considerably dependent on gain, while that in RDS was generally within the range of 0 to +5 degrees regardless of gain. We conducted correlation analyses and found that there was a significant negative correlation between gain and relative angular error in FS ($r = -0.90,\ p < .01$), whereas there was no correlation in RDS ($p > .05$).

3) SUS PQ: The mean score of SUS PQ was 3.47 ($SD = 1.16$) in NS, 3.33 ($SD = 1.23$) in FS, and 3.44 ($SD = 1.39$) in RDS. A Wilcoxon signed-rank test revealed no significant

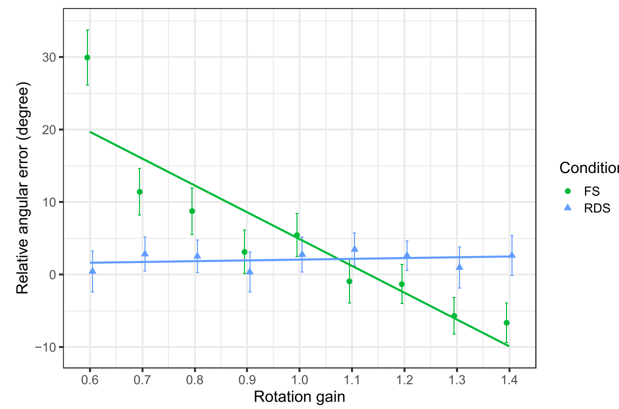

Fig. 7. Results of relative angular error of sound localization [degree] related to gain in FS and RDS.

difference between any conditions ($p > .05$). Therefore, H3 was not supported.

4) Subjective Comments: Regarding the participants' main criteria for judging the amount of rotation, ten participants mentioned the rotational speed and/or amount of the VE. Four participants stated they focused on their sensation of turning their feet. Three participants stated they focused on the metronome sound. One participant answered that his judgement was based on a comparison with the trial before.

Regarding the question of whether they noticed that the metronome sound was virtual, five participants answered they had noticed it, and the remaining thirteen answered they had not. As for the reasons why they noticed it, three participants mentioned discomfort with the quality of the sound (e.g., inclusion of noises, absence of reverberation), one mentioned an inconsistency in the perceived metronome sound as physically moving, despite the fact that the experimenters' footsteps were not heard, and one mentioned an unnaturalness in that the start timing of the metronome sound at the beginning of the trials was felt to be constant (note that this was not actually the case because the start/stop of the metronome sound was operated manually by the experimenter). We compared the results of DTs and sound localization accuracy between the participants who noticed and who did not notice the sound was virtual, but the details are not reported here because no significant differences were found or no meaningful comparisons could be made due to the imbalance of the sample numbers.

Regarding the question of how the auditory cues affected the VR experience, four participants commented that they used the sound as a cue to infer their amount of rotation. On the contrary, three participants commented that answering the question of sound localization distracted them and made answering the question on noticeability difficult or forgettable. In addition, five participants commented that they felt a higher sense of motion sickness when the auditory cues were presented (such comments were obtained for both FS and RDS).

H. Discussion

1) Validity of DTs: Comparing our noticeability results with those of prior studies, the range of DTs in NS (0.81–1.27) was slightly narrower (i.e., more noticeable) than that of Steinicke et al. (0.67–1.24) [3] with a similar experimental design. The reason for this is unclear, but it might be due to the differences in the specifications of the HMD used (e.g., resolution and field of view), the complexity of the VE, or the participants' individual differences (e.g., age, gender, VR experience).

As for the conditions with auditory cues, notably, FS gave a clearly wide range of DTs (0.57-1.37) compared to the prior visual-auditory RDW studies [4], [5], [6] as shown in Table I. Although a simple comparison should not be attempted due to the many differences in the experimental designs, this result still suggests that FS provided the participants with a perception that was distinct from the previous studies. We believe this is mainly due to the presence of the sound localization tasks, as discussed in more detail in the next section.

Nevertheless, the obtained DTs should still be further validated. One reason for this is the small number of trials for each participant (our study had 648 trials per auditory condition, while that of Steinicke et al. [3] had 1,540 trials). Another reason is that the range of DTs obtained in FS was beyond the gain range examined, indicating that noticeability should be further clarified over a wider range of gains.

2) Noticeability and Sound Localization Accuracy: The obtained noticeability results did not support either H1 or H2, which could not be reasonably explained solely from subjective comments. Since most of the participants answered they did not notice that the auditory cue was virtual, we believe that our experimental setup generally succeeded in convincing them that the auditory cue was emanating from the position of the real sound source. In addition, some participants stated that they used the auditory cues as a criterion to infer their facing direction, and others even commented regarding NS that *"it was difficult to infer the rotational amount without the sound."* These comments indicate that the participants estimated the rotational amount while giving much consideration to the auditory cues. More interestingly, contrary to the results of DTs, some comments partially supporting H1 and H2 were given, such as a comment regarding FS that *"I felt a stronger difference between the sense of my own rotation and the visual rotation in VR than [NS]"* and a comment regarding RDS that *"I became unsure of the facing direction as I repeated trials."*

Meanwhile, we found a reasonable interpretation of the noticeability results by combining them with the results of the sound localization accuracy. The results of the relative angular error (Fig. 7) showed unexpected tendencies in which the auditory cue was perceived as shifted in FS where its position was fixed, while it was quite accurately perceived in RDS where its position was manipulated. We interpret these findings as disproving our initial assumption that real-space auditory cues could help the user's spatial perception of the real space; rather, the localization of the real-space auditory cues was affected by the user's vision of the VE according to the visual manipulation. The following sections provide a more detailed interpretation in each condition.

Fixed sound: The reason for the lower noticeability in FS was probably because the inaccurately perceived auditory cue interfered with the user's spatial perception of the real space, rather than helped it. To explain this interpretation, let us exemplify the case where the gain is less than 1, as shown in Fig. 8. In this case, the rotational amount inferred from vision is smaller than the actual amount in reality, whereas the rotational amount inferred from audition corresponds to the actual amount in reality (Fig. 8(a)). This would have made a visual-auditory inconsistency in the participant-estimated position of the sound source (Fig. 8(b)), which would lead to a shift in the sound localization task toward the position inferred from vision (Fig. 8(c)). Similar phenomena to this have been frequently reported in earlier studies in acoustical engineering [39], [40], [41] (though not in VR contexts), and these are called "visual capture." Those studies also reported that the sound localization shifts toward the direction of the motion of the visual cue [39], [40], [41], and the amount of such a shift depends on the velocity of the motion [40]; both of these findings are consistent with our results. Moreover, our interpretation is also similar to that of Gao et al. [9], who took the approach of applying the maximum-likelihood estimation for localizing a target object perceived from vision and audition, but we believe our approach is slightly different because the target object (i.e., a sound source) was invisible in our study.

Moreover, since many participants would have used the auditory cue as a criterion to infer their facing direction, as suggested by the subjective comments, the shifted sound localization of the auditory cues would have led to inaccuracy in their answer to the noticeability question. In the example case under a gain of less than 1, based on the perceived auditory cues shifting toward the participant's rotational direction (Fig. 8(c)), they would have underestimated their rotational amount, thereby reducing the difference from the visually perceived rotational amount in VR, making it difficult to choose the right answer (Fig. 8(d)).

One reason for the wide range of DTs would be the presence of the sound localization task and its order in each trial. More specifically, the sound localization task was conducted right before asking the noticeability question, which might have made the participants more aware of the positional relationship between the auditory cue and themselves, or have just distracted them from inferring their rotational amount. If we had not conducted the sound localization task or swapped the order of the questions, the auditory cues would not have been used to infer the facing direction as much as they were in this experiment, and thus we would expect a narrower range of DTs. Follow-up studies are required to clarify this issue.

Redirected sound: The reason for RDS's equivalent noticeability to that of NS was probably because the participants' perceptions were substantially similar to those of prior visual-auditory RDW techniques [5], [6]. Unlike FS, RDS showed quite accurate sound localization of less than 5 degrees, given the recording interval of 10 degrees. This could be due to the consistency between the visual and auditory cues; more specifically, since there was a consistency between the amount of rotation inferred from vision in VR and audition in reality, the participants would more easily estimate the sound location, even though they faced confusion over how the sound source in reality was moving. One participant's comment regarding RDS would also support this interpretation, commenting that *"I felt the sensation in reality and VR was more consistent*

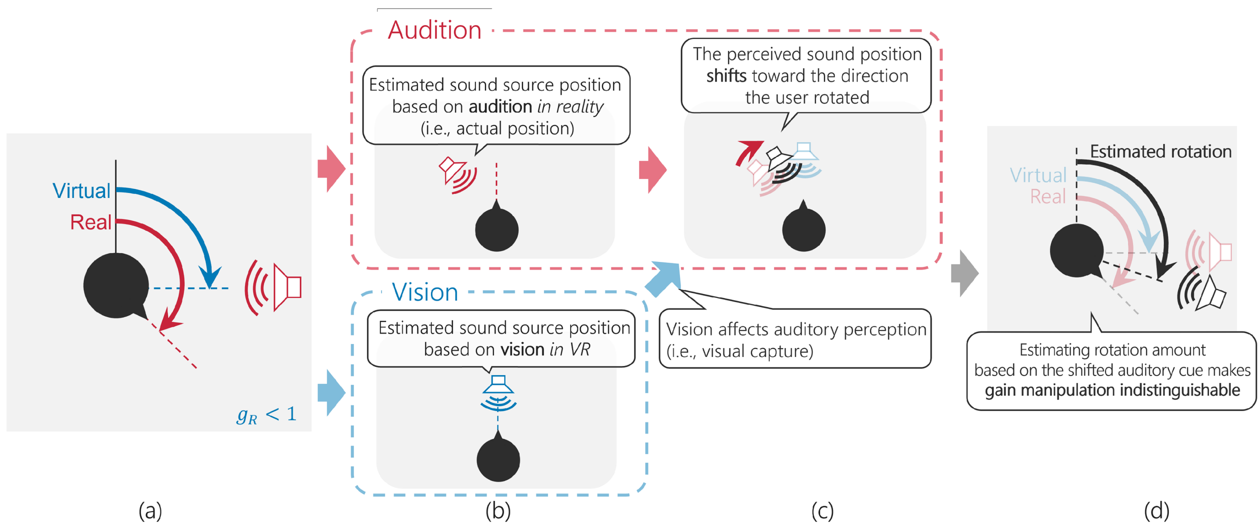

Fig. 8. Our assumed process of the participant's perception in a trial (FS, $g_R < 1$, clockwise): (a) The rotational amount in VR is smaller than that in reality due to the gain; (b) the sound position is estimated by the participant independently from audition and vision, which shows incongruence; (c) the visual-auditory incongruence makes the sound localization shift toward the position inferred from vision; (d) The participant estimates his/her actual rotational amount based on the shifted auditory cue, which reduces the difference from the visually perceived rotational amount in VR and makes it difficult to distinguish the gain manipulation.

than in [FS]." Therefore, the multimodal interaction between vision and audition in RDS would have been similar to the prior visual-auditory RDW studies [5], [6] in which vision and audition are aligned, which might have led to similar results showing the limited effect of the auditory cues on noticeability.

Another interpretation of RDS's noticeability results is that the participants did not rely on the auditory cue. Some participants commented that *"I felt like the metronome was moving together with my rotation"* and *"It was hard to localize the metronome sound compared to [FS],"* suggesting that they felt discomfort because of the auditory and somatosensory inconsistency. Due to this, the participants might not have used such possibly unreliable auditory cues for their spatial perception of the real space, resulting in similar noticeability results to those of NS. The validity of the above two interpretations should be further studied in future work.

3) Sense of Presence: The results showed no significant differences in SUS PQ scores among the three auditory conditions. This suggests that the presentation of the auditory cues did not negatively affect the sense of presence in our experiment, even though the auditory cues were irrelevant to the context in VR or their positions were manipulated. Interestingly, some participants even commented that they felt a higher sense of being in the VE when the sound was presented. One possible reason for such results is that the auditory cues were felt to be a part of the main content rather than external noises because the sound localization task was performed, and thus the participants might not have felt the sound interfered with the VR experience. Another possible reason is that the VR contents used in the experiment did not include any auditory feedback (e.g., conversation, gunshot sound, ambient noise) or narratives. The results of the sense of presence might be different in more realistic applications involving these elements. In addition, some participants commented that they felt higher motion sickness with the auditory cues in both FS and RDS. This might be due to the perceived incongruence among vision, audition, and somatic sensation, which should be elaborated in future work.

IV. Follow-Up Experiment

A. Overview

We conducted a follow-up experiment to clarify the obtained results of the sound localization accuracy. In the main experiment, described in the previous section, the auditory cues using binaural recording succeeded in convincing most participants that the auditory cues emanated from reality, but various factors other than realistic auditory cues (e.g., discrete recording interval of 10 degrees and using non-individualized HRTF) could have affected the results. Consequently, we prepared another auditory condition named *FS-Real*, which was assumed to be substantially the same as FS (called *FS-Binaural* in this section for clarity); however, it differed in that the sound sources (i.e., two metronomes) were actually placed in the real experimental environment. The procedure of the experiment was common to the sound localization task in the main experiment: We asked participants to rotate their entire bodies and then answer the direction from which the sound source was heard. Other conditions, such as the experimental setup, the VE used, the gain presented, and the rotation angle in each trial, were also made consistent with the main experiment. We then compared the obtained results of sound localization accuracy with those in the main experiment. Furthermore, we attempted to recruit all of the participants from the main experiment to also participate in this experiment, and 12 of them (9 males and 3 females) responded.

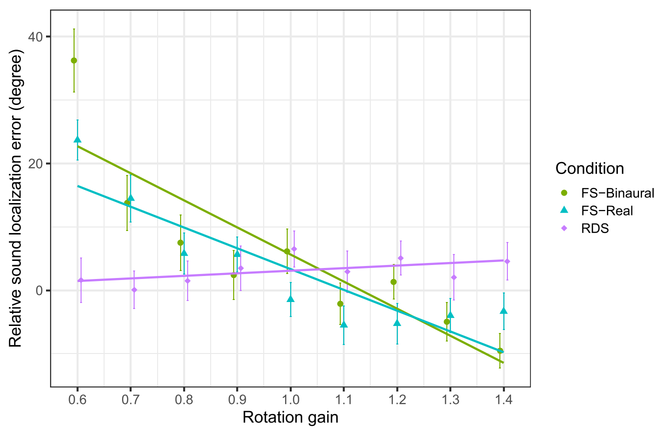

Fig. 9. Results of relative angular error of sound localization [degree] related to gain in FS-Binaural, FS-Real and RDS.

B. Results

The data for a total of 432 trials were obtained. We excluded one participant's data from the analysis because the noise-canceling functionality of the earphone was unintentionally activated during his experiment. Fig. 9 shows the results of relative angular error of the sound localization task in FS-Real related to the gain for the remaining 11 participants' data (396 trials). The graph also shows FS-Binaural and RDS conditions of the corresponding 11 participants for comparison. From the results, we found significant correlations in both FS-Binaural ($r = -0.87,\ p < .01$) and in FS-Real ($r = -0.89,\ p < .01$). However, the correlation in RDS was not significant ($p > .05$).

C. Discussion

The results show that the tendency of relative sound localization accuracy according to gain in FS-Real was generally similar to that of FS-Binaural, suggesting the validity of the auditory cues used in the main experiment. Accordingly, the results also appear to reinforce our interpretation that the shift in sound localization according to gain in both FS conditions was due to visual capture induced by visual manipulation, not to the inherent nature of binaural recording. However, the effect should be further studied in detail in future work, because this experiment was preliminary with some uncontrolled factors such as the order effect (i.e., all participants tried FS-Real after the other conditions).

V. General Discussion

A. Summary of Findings and Possible Applications

Here we summarize our findings from the experiment and discuss possible applications that could leverage these findings.

First, *our initial assumption that VR users could infer their actual direction from external sounds in reality appeared to be incorrect*. The noticeability results of the experiment showed no support for either H1 or H2, suggesting that the real-space auditory cues do not help users perceive the amount of rotation. The reason would be that the perceived direction of the auditory cues in reality was shifted by the visual manipulation in VR, resulting in a less accurate spatial perceptual cue. Therefore, it would not be a promising approach to enhance the redirection effect by manipulating the real-space sound like RDS.

Second, *making users aware of and localizing real-space auditory cues might reduce the noticeability of visual rotational manipulation*. The results of the large sound localization error in FS and the small sound localization error in RDS suggest that the user perception of the real-space auditory cues was shifted by the visual manipulation. Particularly in FS, affected by the shifted auditory perception in the sound localization task, the user would have become less noticeable to the visual manipulation. To exploit this phenomenon in applications, inserting sound localization tasks of real-space sound sources into the VR experience may enhance the effects of redirection.

Third, *real-space auditory cues do not lower the sense of presence of the VR experience, within the scope of our experimental conditions.* However, further studies are required on the effects of the real-space auditory cues on the quality of the VR experience (e.g., testing other types of auditory cues, the existence of in-VR sounds, and the effects on motion sickness).

Given these findings, the applicability of the real-space auditory cues for redirection is rather limited. We initially assumed external noises irrelevant to the VR experience as the real-space auditory cues, but such external noises may not be suitable for applications. The reason is that making the user aware of (and localizing) the auditory cues will be required for redirection, but being aware of such external noises probably interferes with the VR experience. Therefore, we alternatively suggest the use of the real-space sounds *relevant* to the VR scenario as a possible application. Examples include an application in which a speaker placed in reality generates sounds related to a VR scenario (in particular, sounds whose sound source location is not directly seen, such as a conversational voice heard behind a wall in VR). In this application, the user knows the location of the sound source (i.e., speaker) in advance and the sound source is not visible during the VR experience, enabling us to replicate auditory conditions similar to FS in our experiment. In addition, the use of sounds relevant to the VR scenario would naturally encourage the user to localize the sound.

B. Limitations and Future Work

Although we obtained some useful findings for using auditory cues in reality for redirection, our study still has several limitations. One major limitation is the validity of the obtained results. As discussed in Section III-H1, the obtained noticeability results might be framed within a range of initial investigations, and thus they should be further determined with a greater number of trials for each participant and a wider range of gains. Moreover, the generalizability of this study should be further examined: There remain several factors that possibly affect the sound localization accuracy, such as hearing devices, type of auditory cues, and participants' backgrounds (e.g., gender, culture, and experience). Similarly, this study only used metronome sounds as auditory cues in the experiment, but to demonstrate the results in applications (as described in Section V-A), we need to consider a more practical design of auditory cues according to VR narratives.

Another limitation is that the influence of introducing the sound localization task on the obtained results is unclear. As discussed in Section III-H2, we hope to investigate the noticeability results without the sound localization task or swapping the order of questions. Similarly, it is also unclear how the results would be influenced by convincing the participants that the auditory cues were real. Therefore, we intend to examine the results of noticeability and sound localization accuracy in this study in relation to those without giving instructions on it to users.

Finally, there are still challenges in generating spatial auditory cues in response to the user's movements. The follow-up experiment suggests the validity of the binaural recorded auditory cues, but further investigation under more controlled conditions is needed. In addition, to gain deeper insight into RDW utilizing real-space auditory cues, other types of RDW algorithms than rotational manipulation (e.g., translation and curvature manipulation) should be studied in future work.

VI. Conclusion

In this study, we explored an approach to visual-auditory RDW using auditory cues in reality, not in VR. The results of the user experiment ($N = 18$) showed that presenting the auditory cues fixed in reality makes redirection less noticeable, and that manipulating the position of the auditory cues does not reduce the noticeability of redirection. The reason for the reduced noticeability with the auditory cues fixed in reality was interpreted as the visual manipulation in VR causing the participants' shifted localization of the sound source in reality, and then the shifted auditory localization was used to infer their facing direction. Based on these findings, we suggested an application in which the redirection noticeability could be decreased through using a fixed loudspeaker in reality playing sounds relevant to a VR scenario.

References

[1] S. Razzaque, Z. Kohn, and M. C. Whitton, "Redirected walking," in *Proc. 22nd Annu. Conf. Eur. Assoc. Comput. Graph.*, 2001, pp. 289–294.

[2] J. Dichgans and T. Brandt, "Visual-vestibular interaction: Effects on self-motion perception and postural control," in *Perception*. Berlin, Germany: Springer, 1978, pp. 755–804. [Online]. Available: https://doi.org/10.1007/978-3-642-46354-9_25

[3] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe, "Estimation of detection thresholds for redirected walking techniques," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 1, pp. 17–27, Jan./Feb. 2010.

[4] F. Meyer, M. Nogalski, and W. Fohl, "Detection thresholds in audio-visual redirected walking," in *Proc. Sound Music Comput. Conf.*, 2016, pp. 293–299.

[5] N. C. Nilsson, E. Suma, R. Nordahl, M. Bolas, and S. Serafin, "Estimation of detection thresholds for audiovisual rotation gains," in *Proc. IEEE VR*, IEEE, 2016, pp. 241–242.

[6] A. Junker et al., "Revisiting Audiovisual Rotation Gains for Redirected Walking," in *Proc. IEEE Virtual Reality*, 2021, pp. 358–359.

[7] J. Lackner, "Induction of illusory self-rotation and nystagmus by a rotating sound-field," *Aviation, Space, Environ. Med.*, vol. 48, no. 2, pp. 129–131, Feb. 1977.

[8] E. Goldstein, *Sensation and Perception*. Belmont, CA, USA: Wadsworth, 2010.

[9] P. Gao, K. Matsumoto, T. Narumi, and M. Hirose, "Visual-auditory redirection: Multimodal integration of incongruent visual and auditory cues for redirected walking," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2020, pp. 639–648.

[10] P. Larsson, D. Västfjäll, and M. Kleiner, "Perception of self-motion and presence in auditory virtual environments," in *Proc. 7th Annu. Int. Workshop Presence*, 2004, pp. 252–258.

[11] S. Serafin, N. C. Nilsson, E. Sikstrom, A. D. Goetzen, and R. Nordahl, "Estimation of detection thresholds for acoustic based redirected walking techniques," in *Proc. IEEE Virtual Reality*, Mar. 2013, pp. 161–162. [Online]. Available: https://doi.org/10.1109/vr.2013.6549412

[12] M. Nogalski and W. Fohl, "Acoustic redirected walking with auditory cues by means of wave field synthesis," in *Proc. IEEE Virtual Reality*, Mar. 2016, pp. 245–246. [Online]. Available: https://doi.org/10.1109/vr.2016.7504745

[13] A. Nguyen, Y. Rothacher, B. Lenggenhager, P. Brugger, and A. Kunz, "Individual differences and impact of gender on curvature redirection thresholds," in *Proc. 15th ACM Symp. Appl. Percep.*, New York, NY, USA, 2018, pp. 5:1–5:4. [Online]. Available: https://doi.org/10.1145/3225153.3225155

[14] N. L. Williams and T. C. Peck, "Estimation of rotation gain thresholds considering FOV, gender, and distractors," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 11, pp. 3158–3168, Nov. 2019.

[15] A. Nguyen et al., "Effect of cognitive load on curvature redirected walking thresholds," in *Proc. 26th ACM Symp. Virtual Reality Softw. Technol.*, New York, NY, USA, 2020, pp. 1–5. [Online]. Available: https://doi.org/10.1145/3385956.3418950

[16] L. Kruse, E. Langbehn, and F. Steinicke, "I can see on my feet while walking: Sensitivity to translation gains with visible feet," in *Proc. IEEE Virtual Reality*, 2018, pp. 305–312. [Online]. Available: https://doi.org/10.1109/vr.2018.8446216

[17] A. Paludan et al., "Disguising rotational gain for redirected walking in virtual reality: Effect of visual density," in *Proc. IEEE Virtual Reality*, 2016, pp. 259–260. [Online]. Available: https://doi.org/10.1109/vr.2016.7504752

[18] N. C. Nilsson et al., "15 years of research on redirected walking in immersive virtual environments," *IEEE Comput. Graph. Appl.*, vol. 38, no. 2, pp. 44–56, Mar./Apr. 2018. [Online]. Available: https://doi.org/10.1109/mcg.2018.111125628

[19] Y.-J. Li, F. Steinicke, and M. Wang, "A comprehensive review of redirected walking techniques: Taxonomy, methods, and future directions," *J. Comput. Sci. Technol.*, vol. 37, no. 3, pp. 561–583, May 2022. [Online]. Available: https://doi.org/10.1007/s11390--022-2266-7

[20] T. Grechkin, J. Thomas, M. Azmandian, M. Bolas, and E. Suma, "Revisiting detection thresholds for redirected walking," in *Proc. SAP. Assoc. Comput. Machinery*, Jul. 2016, pp. 113–120. [Online]. Available: https://doi.org/10.1145/2931002.2931018

[21] L. Rayleigh, "On our perception of sound direction," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 13, no. 74, pp. 214–232, Feb. 1907. [Online]. Available: https://doi.org/10.1080/14786440709463595

[22] M. B. Gardner, "Distance estimation of 0 or apparent 0-oriented speech signals in anechoic space," *J. Acoustical Soc. Amer.*, vol. 45, no. 1, pp. 47–53, Jan. 1969. [Online]. Available: https://doi.org/10.1121/1.1911372

[23] T. Gotoh, Y. Kimura, A. Kurahashi, and A. Yamada, "A consideration of distance perception in binaural hearing," *J. Acoustical Soc. Jpn.*, vol. 33, no. 12, pp. 667–671, 1977.

[24] S. P. Thompson, "On the function of the two ears in the perception of space," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 13, no. 83, pp. 406–416, Jun. 1882. [Online]. Available: https://doi.org/10.1080/14786448208627205

[25] R. Dodge, "Thresholds of rotation," *J. Exp. Psychol.*, vol. 6, no. 2, pp. 107–137, Apr. 1923. [Online]. Available: https://doi.org/10.1037/h0076105

[26] B. E. Riecke, D. Feuereissen, and J. J. Rieser, "Auditory self-motion illusions ("circular vection") can be facilitated by vibrations and the potential for actual motion," in *Proc. 5th Symp. Appl. Percep. Graph. Visual.*, 2008, pp. 147–154. [Online]. Available: https://doi.org/10.1145/1394281.1394309

[27] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, Jan. 2002. [Online]. Available: https://doi.org/10.1038/415429a

[28] T. Feigl, E. Kõre, C. Mutschler, and M. Philippsen, "Acoustical manipulation for redirected walking," in *Proc. Symp. Virtual Reality Softw. Technol.*, Nov. 2017, pp. 45:1–45:2. [Online]. Available: https://doi.org/10.1145/3139131.3141205

[29] R. Weller, B. Brennecke, and G. Zachmann, "Redirected walking in virtual reality with auditory step feedback," *Vis. Comput.*, vol. 38, no. 9–10, pp. 3475–3486, Jul. 2022. [Online]. Available: https://doi.org/10.1007/s00371-022-02565-4

[30] N. Rewkowski, A. Rungta, M. Whitton, and M. Lin, "Evaluating the effectiveness of redirected walking with auditory distractors for navigation in virtual environments," in *Proc. IEEE Virtual Reality*, Mar. 2019, pp. 395–404. [Online]. Available: https://doi.org/10.1109/vr.2019.8798286

[31] J. Lee, S. Hwang, K. Kim, and S. Kim, "Auditory and olfactory stimuli-based attractors to induce reorientation in virtual reality forward redirected walking," in *Proc. CHI Conf. Hum. Factors Comput. Syst. Extended Abstr.*, Apr. 2022, pp. 446:1–446:7. [Online]. Available: https://doi.org/10.1145/3491101.3519719

[32] J. Zhang, E. Langbehn, D. Krupke, N. Katzakis, and F. Steinicke, "Detection thresholds for rotation and translation gains in 360 video-based telepresence systems," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 4, pp. 1671–1680, Apr. 2018. [Online]. Available: https://doi.org/10.1109/tvcg.2018.2793679

[33] B. J. Congdon and A. Steed, "Sensitivity to rate of change in gains applied by redirected walking," in *Proc. Symp. Virtual Reality Softw. Technol.*, Nov. 2019, pp. 3:1–3:9. [Online]. Available: https://doi.org/10.1145/3359996.3364277

[34] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence: Teleoperators Virtual Environ.*, vol. 9, no. 5, pp. 497–503, Oct. 2000. [Online]. Available: https://doi.org/10.1162/105474600566989

[35] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *Int. J. Aviation Psychol.*, vol. 3, no. 3, pp. 203–220, 1993. [Online]. Available: https://doi.org/10.1207/s15327108ijap0303_3

[36] T. Z. Strybel and K. Fujimoto, "Minimum audible angles in the horizontal and vertical planes: Effects of stimulus onset asynchrony and burst duration," *J. Acoustical Soc. Amer.*, vol. 108, no. 6, pp. 3092–3095, Dec. 2000. [Online]. Available: https://doi.org/10.1121/1.1323720

[37] A. Honda, S. Tsunokake, Y. Suzuki, and S. Sakamoto, "Auditory subjective-straight-ahead blurs during significantly slow passive body rotation," *i-Perception*, vol. 13, no. 1, Jan. 2022, Art. no. 204166952110706. [Online]. Available: https://doi.org/10.1177/20416695211070616

[38] Y. Iwaya, Y. Suzuki, and D. Kimura, "Effects of head movement on front-back error in sound localization," *Acoustical Sci. Technol.*, vol. 24, no. 5, pp. 322–324, 2003. [Online]. Available: https://doi.org/10.1250/ast.24.322

[39] W. R. Thurlow and T. P. Kerr, "Effect of a moving visual environment on localization of sound," *Amer. J. Psychol.*, vol. 83, no. 1, Mar. 1970, Art. no. 112. [Online]. Available: https://doi.org/10.2307/1420861

[40] S. Mateeff, J. Hohnsbein, and T. Noack, "Dynamic visual capture: Apparent auditory motion induced by a moving visual target," *Perception*, vol. 14, no. 6, pp. 721–727, Dec. 1985. [Online]. Available: https://doi.org/10.1068/p140721

[41] N. Kitajima and Y. Yamashita, "Dynamic capture of sound motion by light stimuli moving in three-dimensional space," *Perceptual Motor Skills*, vol. 89, no. 3_suppl, pp. 1139–1158, Dec. 1999. [Online]. Available: https://doi.org/10.2466/pms.1999.89.3f.1139

Kumpei Ogawa received the BS and MS degrees from Tohoku University, Japan, in 2021 and 2023, respectively. He is currently working toward the PhD degree with the Graduate School of Information Science, Tohoku University. His research interests include human computer interaction and virtual reality.

Kazuyuki Fujita received the PhD in information science and technology from Osaka University, in 2013. He worked for ITOKI from 2013 to 2018, and he has been working as an Assistant Professor of Research Institute of Electrical Communication, Tohoku University from 2018. He has also been granted the title of Prominent Research Fellow at Tohoku University from 2023. His research interests include human-workspace interaction and virtual reality.

Shuichi Sakamoto received the BS, MSc, and PhD degrees from Tohoku University, Japan, in 1995, 1997, and 2004, respectively. He is currently a professor with the Research Institute of Electrical Communication, Tohoku University. His research interests include human multisensory information processing including hearing, speech perception, and the development of high-definition three-dimensional audio recording systems. He is a member of ASJ, ASA, IEICE, VRSJ, ASA, and several other professional societies.

Kazuki Takashima received the PhD degree from the Graduate School of Information Science and Technology, Osaka University, in 2008. He then worked as an Assistant Professor with Osaka University, and joined Tohoku University's Research Institute of Electrical Communication as an Assistant Professor, in 2011. He was promoted to the rank of Associate Professor with Tohoku University, in 2018.

Yoshifumi Kitamura is professor and deputy director of the Research Institute of Electrical Communication, Tohoku University, and director of the Interdisciplinary ICT Research Center for Cyber and Real Spaces. His research interests include interactive content design, human computer interactions, 3D user interfaces, virtual reality, telecommunication with nonverbal information, and related fields. He has served in positions such as the Japan Liaison of IFIP TC-13, Japan Liaison and chair of ACM SIGCHI Asian Development Committee, chair of Japan ACM SIGCHI Chapter, Steering Committee chair of ACM VRST, and Conference chair of ACM SIGGRAPH Asia 2015, general chair of CHI 2021.