IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 20, 2021, accepted December 6, 2021, date of publication December 20, 2021, date of current version December 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3136557

Virtual Walking With Omnidirectional Movies and Foot Vibrations: Scene-Congruent Vibrations Enhance Walking-Related Sensations and Ground Material Perceptions

JUNYA NAKAMURA[1], YUSUKE MATSUDA[1], TOMOHIRO AMEMIYA[2], (Member, IEEE), YASUSHI IKEI[2], AND MICHITERU KITAZAKI[1]

[1]Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi 4418580, Japan
[2]Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo 1138656, Japan

Corresponding authors: Junya Nakamura (nakamuraj@real.cs.tut.ac.jp) and Michiteru Kitazaki (mich@tut.jp)

This work was supported in part by the Japan Science and Technology Agency: Exploratory Research for Advanced Technology (JST ERATO) under Grant JPMJER1701, and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP18H04118 and Grant JP20H04489.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee for Human-Subject Research at Toyohashi University of Technology under Application No. H31-06.

ABSTRACT Walking is a human need. Even when physical disabilities or social situations, such as the COVID-19 pandemic, prevent people from walking, virtual reality can provide an opportunity for virtual walking or travel. In this study, a virtual walking system was developed in which users view omnidirectional movies while receiving scene-congruent vibrations to their left and right forefeet and heels. The timings of foot vibrations were generated by estimating the camera motion trajectory with visual simultaneous localization and mapping (visual SLAM) applied to four omnidirectional movies. Congruent vibration patterns were prepared for four ground scenes. Modulation of walking-related sensations and ground material perceptions by congruent and incongruent vibrations was verified using psychological measurements. The results showed that rhythmic foot vibration improved the sensations of self-motion, walking, leg action, and telepresence irrespective of scene–vibration congruency. Moreover, congruent vibrations were better than incongruent vibrations for walking-related sensations and telepresence in indoor corridors and snowy ground scenes. The perception of ground materials was enhanced by scene-congruent vibrations, whereas it was confused by scene-incongruent vibrations. These findings suggest that vibration patterns do not necessarily need to match the ground exactly to induce virtual walking sensations; however, scene-congruent or similar vibrations improve virtual walking sensations and ground material perception. By applying our methods, we can convert various public omnidirectional movies into realistic virtual walking experiences.

INDEX TERMS Psychology, tactile perception, virtual reality, virtual walking, visual perception.

I. INTRODUCTION

Humans move with their feet to explore environments, look for food, escape from threats, and communicate with others. Walking is a basic human activity that improves human physical and mental health [1]–[3]. Factors such as feasibility, accessibility, safety, comfort, and pleasure affect or can be limiting factors to walking [4]. In terms of feasibility, for example, persons with disabilities affecting their legs cannot walk. Others do not walk if the destination is too far, instead using other forms of mobility, such as driving. In terms of accessibility, for example, people walk if there are sidewalks or trails. In terms of safety, local crime rates affect decision-making about walking; moreover, the COVID-19 pandemic has discouraged people from walking outside [5]–[7]. Virtual reality can remove some of these limiting factors for walking, and hence, various virtual walking systems have been developed [8], [9]. In particular, we aim to remove physical limitation factors of uses so that we have developed a virtual walking system for seated observers.

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

Virtual walking systems can be categorized as active [10]–[28] and passive [29]–[39] systems. In active virtual walking systems, users move their body parts to produce locomotion in virtual environments. Gestures accompanying walking, such as virtual steps [10]–[13] or swinging arms [14], [15], are used in the walk-in-place (WIP) method. Locomotion by which the body leans in the appropriate direction is another WIP gesture [16], [17]. An omnidirectional treadmill makes natural walking motion possible [18]–[20]. Other devices to enable walking motions with actuators have been developed, such as foot-supporting arms [21], moving tiles [22], and rotating large spheres [23]. Mapping the cycling biomechanics of the legs of users to virtual walking has been proposed [24]. A walking system with low-friction shoes on a dish-shaped apparatus has been developed [25], [26] and made commercially available, such as Virtuix Omni One. The redirection technique manipulates the route of a user walking in a virtual environment to cause the virtual path to deviate from the real-world path, so that users can walk in infinite space within a limited physical space [27], [28]. These active virtual walking systems work by combining visual and other sensory displays, such as immersive or head-mounted displays. A rich sensation of walking can be induced by these systems because the motor commands and proprioception of users are included. However, people with walking disabilities cannot use these systems, and some of the systems are large and cannot be used at home. Passive walking systems provide a virtual walking opportunity for persons with walking disabilities and can be built inexpensively.

Passive virtual walking systems utilize sensory stimuli and/or passive movement of the limbs or body. Visually induced self-motion perception (vection) is one of the most popular sensory phenomena for locomotion in virtual environments [40], [41]. Vection has been investigated extensively in perceptual psychology [42], [43], and similar illusory self-motion perception can be induced by auditory stimuli [44]–[47] and by a combination of auditory and tactile stimulations [42]. Vection requires global visual motion in a large field [48]–[50]. It is dominated by background motion rather than foreground motion [51], [52] and unattended motion rather than attended motion [53]. Perspective jitter [54], [55] and naturalistic scenes [56] enhance vection strength. Most passive virtual walking systems for seated users combine vection with other sensory displays. Lécuyer *et al.* [29] simulated camera motion as perspective jitter to induce a walking sensation, and Terziman *et al.* [30] utilized artificial visual and tactile vibrations corresponding to footsteps during virtual walking (the ''King-Kong effect''). The FiveStar or Five Senses Theatre system [32] presents proprioceptive and tactile sensations of the body passively evoked by actuators, in addition to multisensory stimuli, such as vision, audition, air flow, and smells. Electrical stimulation of the muscles of the legs [33] and small motions of the lower limbs [34] or the entire body [35] by mechanical actuators can elicit a virtual walking sensation. Kruijff *et al.* [36] used visual, auditory, and tactile stimulation of the feet to simulate footsteps, and showed that vection can be enhanced in a leaning-based locomotion interface with these stimulations. Kitazaki *et al.* [37] captured visual motion images of an actual walker with footstep timings and presented the captured optic flow with foot vibrations that were synchronous to the oscillating optic flow and footstep to seated participants. In addition to the sensation of self-motion, the sensations of walking, leg action, and telepresence were enhanced by the synchronous foot vibrations. Walking avatars induced an illusory sense of agency over walking for seated participants [57] and enhanced virtual walking sensations [38], [39].

Omnidirectional movies or 360° movies [58]–[60] are becoming popular and are in the public domain on Internet websites and YouTube. These movies can be made using consumer 360° cameras, such as RICHO THETA and Insta360. If these omnidirectional movies are used in a virtual walking system, virtual travels to various places in the world can be offered to anyone. To realize this, it is necessary to estimate the motion trajectories of the camera/viewpoint in the movies. Visual simultaneous localization and mapping (visual SLAM) can be applied for this estimation [61]–[63].

Regarding vibration patterns during walking, Terziman *et al.* [30] compared two vibration models (one contact of heel strike and two contacts of both heels strike and toe strike) and found that the sensation of walking was better with only one contact of the foot rather than two. Regarding the ground materials during walking, Turchet *et al.* [31] used different vibration patterns depending on the ground material (snow, forest floor, and sand). Humans' haptic perception of ground compliance is affected by vibration amplitude, waveform, amplitude envelope, and frequency distribution of the feet during walking [64]. However, Kitazaki *et al.* [37] applied identical foot vibrations to three different ground scenes. Because humans optimally integrate visual and haptic information [65]–[68], the walking sensation can be enhanced by presenting foot vibrations congruent with the ground scene. However, whether vibration patterns congruent with the ground materials improve virtual walking sensations or scene-incongruent vibration patterns cause virtual walking sensations to deteriorate has not been investigated.

The aim of this study was to develop a virtual walking system for seated users, enabling them to experience virtual walking in various scenes, even if they have physical disabilities or are prevented from walking outside. For this purpose, a virtual walking system based on omnidirectional movies was developed by estimating the motion trajectory of a virtual viewpoint/camera with visual SLAM and adding rhythmic foot vibrations. Then, whether the walking-related sensations were modulated by the congruency between the visual ground scene and the vibration patterns was tested, as well as whether the congruent and incongruent vibration patterns affected the perception of the ground materials. We hypothesized that both the walking-related sensations and the perception of the ground materials would be improved when the vibration patterns were congruent with the visual ground scene.

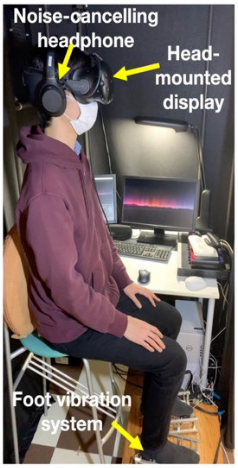


FIGURE 1. Overview of the virtual walking system. Visual stimuli were presented on an HMD controlled by a computer, which also controlled and presented tactile stimuli thought the vibration system on the feet. A noise cancelling headphone was used to eliminate sound from the vibration system.

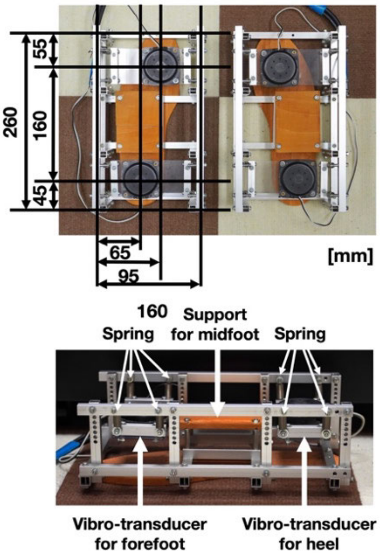


FIGURE 2. The foot vibration systems. Four vibro-transducers presented vibrations on the left and right forefeet and heels. The vibro-transducers were suspended from aluminum frames by springs to prevent vibrations from transmitting to other parts.

II. METHODS

A. PARTICIPANTS

Fourteen volunteers participated in the experiment — all male, mean 21.64 years old ±0.81 standard deviation (SD). Their anthropometric data were obtained from thirteen participants, whereas one participant declined to provide the data — mean height 1.71 m ± 0.04 SD, mean weight 61.69 kg ± 10.22 SD, mean Body Mass Index (BMI) 21.19 kg/m^2 ± 3.51 SD. The sample size was determined by a power analysis: a medium effect size f = 0.25, alpha = 0.05, power = 0.8, and repeated measures of analysis of variance (ANOVA) — four movie conditions × three vibration conditions — using G * Power 3.1 [69], [70]. All participants had normal binocular vision and physical abilities. They gave written informed consent before the experiment. All experiments were approved by the Ethical Committee for Human-Subject Research at Toyohashi University of Technology and were performed in accordance with the committee's guidelines and regulations.

B. APPARATUS

A computer (Intel Core i7 10700 CPU, NVIDIA GeForce RTX 2070 Super Graphics, DDR4 32GB memory) controlled the visual and tactile stimuli using Unity (2018.4.28f1). Visual stimuli were presented on a head-mounted display (HMD) (Figure 1): HTC VIVE, resolution 1080 [width] × 1200 [height] pixels for each eye, refresh rate of 90 Hz. Tactile stimuli were presented to the left and right forefeet and heels of the participants through four vibro-transducers (Acouve Lab Vp408). These vibro-transducers were mounted on acrylic plates separately and suspended from aluminum frames by springs to prevent vibrations from being transmitted to other parts, while the midfeet of participants were supported by wood plates that were connected to the aluminum frames (Figure 2). The vibro-transducers were driven by a multichannel power amplifier (Behringer EPQ304, 40 W (8 Ω × 4 ch) and a multichannel preamplifier (Behringer FCA1616, input 16 ch, output 16 ch) connected to the computer. The amplitude of vibration was fixed during the experiments and sufficiently strong for all participants to feel vibrations while wearing socks. White noise (70 dBA) was presented to the participants through a noise-canceling headphone (SONY WH-1000XM4) to prevent participants from hearing sounds from the vibro-transducers. The system latency was as high as 11.11 ms.

C. STIMULI AND CONDITIONS

1) VISUAL STIMULI

Four omnidirectional motion scenes with different grounds (indoor corridor, grassy ground, concrete paved road, and snowy ground) were used (Figure 3; see Supplementary video). We chose these different grounds based on previous studies on the perception of ground materials during walking [64], [71], [72]. Giordano *et al.* [71], [72] showed that humans can discriminate solid materials from aggregate materials using only haptic sensations on feet during walking, and the perception of ground compliance depends on the vibration patterns on the feet during walking [64]. Thus, we prepared a variety of grounds with different solidness and compliance; the indoor corridor was the most solid, followed by the concrete paved road. The grassy ground was less solid, and the snowy ground was the softest. We predicted that participants could discriminate between these ground materials using only haptic sensations.

Three scenes (indoor corridor, grassy ground, and concrete paved road) were captured on the campus of Toyohashi

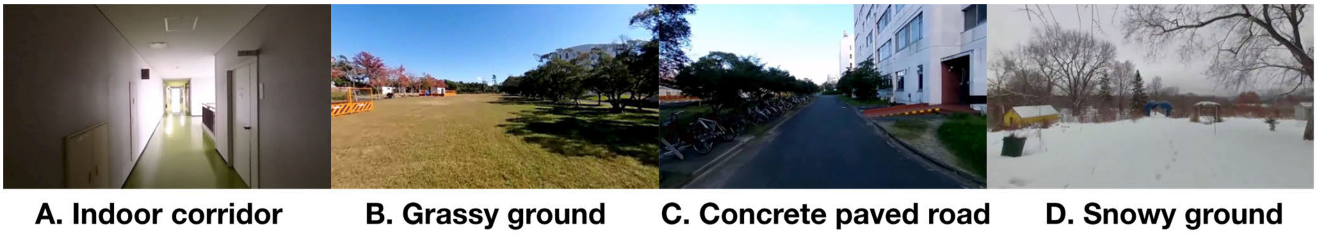


FIGURE 3. Example images of four visual stimuli: (A) indoor corridor scene, (B) grassy ground scene, (C) concrete paved road scene, and (D) snowy ground scene.

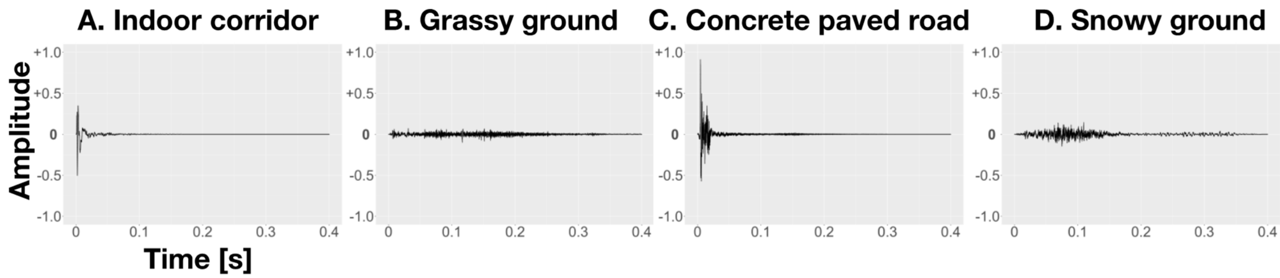


FIGURE 4. Amplitude profile of vibrations: (A) indoor corridor scene, (B) grassy ground scene, (C) concrete paved road scene, and (D) snowy ground scene.

University of Technology, Aichi, Japan, using a commercial 360° camera (RICOH THETA Z1, 3840 [width] × 1920 [height], pixel resolution, 29.97 fps) with a camera gimbal (DJI RONIN SC). The snow ground scene was taken from YouTube [73] because it was not possible to capture a snow scene at the test site. All movies were captured by a walking person holding the 360° camera in their hand. All scenes were forward-moving images at an approximately constant speed (1.26–1.41 m/s, depending on the scene) for 30 s. Omnidirectional motion images were presented on the inside surface of a sphere, Skybox (Unity 2018.4), and the viewpoint of the participant was located at the center of the sphere. There were no binocular disparities.

2) TACTILE STIMULI

The vibro-transducer presented vibrations by receiving sound inputs. Four footstep sounds were prepared that matched the ground materials of the scenes (indoor corridor, grassy ground, concrete paved road, and snowy ground, Figure 4). The sounds were chosen from the source asset of Unity — Floor of Classic Footstep SFX, Footstep (Grass), Ground of Classic Footstep SFX, and Footstep (Snow) respectively for the indoor corridor, grassy ground, concrete paved road, and snowy ground scenes. The duration of the vibrations was 400 ms. The transition of the camera position and rotation information in the movies was estimated using OpenVSLAM [63]. From the estimated camera motions, left and right (two-channel) footstep timings were generated for each movie with a constant step distance (81 cm/step). Thus, the walking speed depended on the transition speed of the scene (1.56–1.74 steps/s). This timing generation was not in real time but was automatically calculated before the experiment. These two-channel vibration timings were converted to four channels (the left and right forefeet and heels) by adding a 105-ms difference between respective contacts with the heel and forefoot.

3) CONDITIONS

The experimental conditions were combinations of four different ground scenes (indoor corridor, grassy ground, concrete paved road, and snowy ground) and three scene–vibration congruency conditions with within-subject design. The scene–vibration congruency conditions consisted of congruent, incongruent, and no-vibration conditions. In the congruent condition, each vibration was matched to the ground type in the scene. In the incongruent condition, one from the three non-matching vibrations was randomly chosen and presented to the scene. In the no-vibration condition, no vibration was presented, and only the visual scene was presented.

4) QUESTIONNAIRE

Participants were asked to complete two sets of questionnaires (Figure 5). The first consisted of descriptions regarding walking-related and telepresence sensations (Figure 5A).

W1. I felt as if my whole body was moving forward (self-motion).

W2. I felt as if I was walking forward (walking sensation).

W3. I felt as if my feet were striking the ground (leg action).

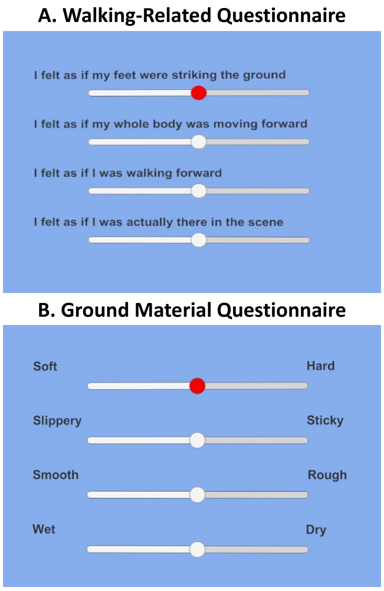


FIGURE 5. **Two questionnaires. Participants were asked to respond to each item with the visual analog scale (VAS). The order of items in each questionnaire was randomized for each trial. (A) An example of response screen of questionnaire items on walking-related and telepresence sensations. (B) An example of response screen of questionnaire items on ground material perception.**

W4. I felt as if I were actually there in the scene (telepresence).

These characterizations were based on previous studies [37], [38]. The order of items was randomized for each trial. The responses of participants were obtained using a visual analog scale (VAS). A line was presented on a screen, and the leftmost side of the line implied no sensation, whereas the right side of the line implied the same sensation as in the actual walking experience. The data were digitized from 0 to 100 for the analysis.

The second questionnaire consisted of choices between four pairs of contrasting adjective characterizations regarding ground materials (Figure 5B). These were selected based on the results of a preliminary experiment in which eight pairs of adjectives were used; Cronbach's coefficient α was used to eliminate similar items amongst the set "smooth–rough, soft–hard, wet–dry, slippery–sticky, fragile–tough, sinking–bouncing, shallow–deep, and sparse–dense," in order to improve the measure of scale reliability.

The four pairs that were used were as follows:

G1. Smooth–Rough

G2. Soft–Hard

G3. Wet–Dry

G4. Slippery–Sticky

The order of the items was randomized for each trial. The participants were asked to identify, via a VAS, the feeling of the ground materials that they experienced in each trial. A line was presented on a screen. The leftmost side of the line indicated one adjective, and the right side of the line indicated the other adjective. The data were digitized from −50 to +50 for the analysis.

D. PROCEDURE

The participants sat on a chair while wearing HMDs and placed their feet on a foot vibration system (see Figure 1). The position and height of the foot vibration system were adjusted for each participant to receive sufficiently strong tactile stimuli before the experiments. They wore noise-canceling headphones to prevent hearing the sounds of the vibro-transducers. Each trial consisted of a blank screen (5 s), a fixation cross (5 s), and a stimulus presentation (30 s). Participants performed 48 trials (four ground types × three vibration congruency conditions × four repetitions) in random order.

E. STATISTICAL ANALYSIS

The VAS data of walking-related sensations (self-motion, walking sensation, leg action sensation, and telepresence) were digitized between 0–100 for statistical analysis. The VAS data for ground perception (pairs of adjectives: G1. smooth–rough, G2. soft–hard, G3. wet–dry, G4. slippery–sticky) were digitized between −50 to +50 for statistical analysis. Statistical tests were performed with the R 4.1 software. First, normality of data was tested using the Shapiro–Wilk test $\alpha = 0.05$). If the data did not violate the normality ($p > 0.05$), a two-way repeated measures ANOVA was performed (four ground types × three vibration congruency conditions). If the data violated the normality test ($p < .05$), a two-way repeated measures ANOVA was performed with an aligned rank transformation (ANOVA with ART) procedure [74] as a non-parametric test. Then, we performed an analysis of simple main effects and a post-hoc multiple-comparison analysis as necessary. For parametric data, if there was a lack of sphericity with Mendoza's multisample sphericity test, the reported values were adjusted using the Greenhouse–Geisser correction [75]. Shaffer's modified sequentially rejective Bonferroni procedure was applied for post-hoc multiple comparisons. For non-parametric data, Tukey's method with Kenward–Roger degrees of freedom approximation [76] was applied for post-hoc multiple comparison analysis.

III. RESULTS

A. WALKING-RELATED AND TELEPRESENCE SENSATIONS

1) SELF-MOTION SENSATION

For the self-motion sensation (Figure 6A), ANOVA with ART revealed a significant main effect of scene–vibration congruency ($F(2, 26) = 19.00$, $p < .0001$, $\eta_p^2 = 0.60$) and the interaction between the ground scene and scene–vibration congruency ($F(6, 78) = 3.23$, $p = .007$, $\eta_p^2 = 0.20$).

The analysis of the simple main effects showed that the effects of scene–vibration congruency were significant in all ground scenes (indoor: $F(2, 26) = 18.28$, $p < .0001$, $\eta_p^2 = 0.58$; grass: $F(2, 26) = 9.69$, $p = .0007$, $\eta_p^2 = 0.43$; concrete: $F(2, 26) = 8.92$, $p = .001$, $\eta_p^2 = 0.41$; snow: $F(2, 26) = 23.75$, $p < .0001$, $\eta_p^2 = 0.65$), while the effect of the different ground scenes was significant in the congruent condition

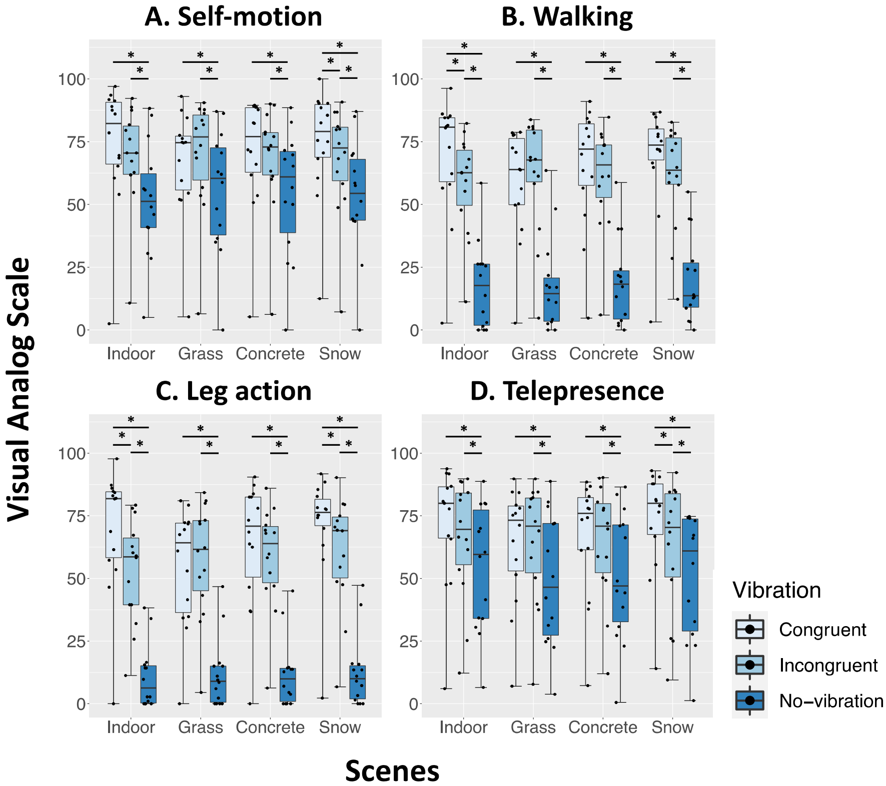


FIGURE 6. Results of walking-related and telepresence sensations: (A) self-motion sensation, (B) walking sensation, (C) leg action sensation, and (D) telepresence. The horizontal axis indicates four different ground scenes. The vertical axis indicates the value of digitized VAS (0–100). A box indicates the range between Q1 (25%) and Q3 (75%), the thick line in the box indicates the median, and a set of whiskers indicates the maximum and minimum values. * indicates p < .05.

(congruent: F (3, 39) = 5.60, p = .0027, $\eta_p^2 = 0.30$), but not significant in the incongruent and the no-vibration conditions (incongruent: F (3, 39) = 0.47, p = .71, $\eta_p^2 = 0.035$; no-vibration: F (3, 39) = 1.31, p = .29, $\eta_p^2 = 0.091$).

For all ground scenes, the scores for the congruent and incongruent conditions were significantly higher than those for the no-vibration condition. The score for the congruent condition was significantly higher than for the incongruent condition only for the snowy ground scene (Table 1A). The scores for the indoor-scene condition and snowy-scene condition were significantly higher than for the grassy-scene condition only in the scene–vibration-congruent condition (Table 2A).

2) WALKING SENSATION

For the walking sensation (Figure 6B), the ANOVA with ART revealed a significant main effect of scene–vibration congruency (F (2, 26) = 49.45, p < .0001, $\eta_p^2 = 0.79$) and the interaction (F (6, 78) = 2.98, p = .011, $\eta_p^2 = 0.19$).

The analysis of the simple main effects showed that the effects of scene–vibration congruency were significant in all the ground scenes (indoor: F (2, 26) = 48.65, p < .0001, $\eta_p^2 = 0.79$; grass: F (2, 26) = 32.07, p < .0001, $\eta_p^2 = 0.71$; concrete: F (2, 26) = 28.50, p < .0001, $\eta_p^2 = 0.69$; snow: F (2, 26) = 46.97, p < .0001, $\eta_p^2 = 0.78$), whereas the effects of the different ground scenes were significant in the congruent vibration conditions (congruent: F (3, 39) = 5.79, p = .002, $\eta_p^2 = 0.31$) but not significant in the incongruent and no-vibration conditions (incongruent: F (3, 39) = 0.83, p = .48, $\eta_p^2 = 0.060$; no-vibration: F (3, 39) = 0.85, p = .47, $\eta_p^2 = 0.062$).

For all ground scenes, the scores for the congruent and incongruent conditions were significantly higher than those for the no-vibration condition. For the indoor-scene, the score for the congruent condition was significantly higher than that for the incongruent condition (Table 1B). The scores for the indoor-scene and snowy-scene conditions were significantly higher than for the grassy-scene condition only in the scene–vibration-congruent condition (Table 2B).

3) LEG ACTION SENSATION

For leg action sensation (Figure 6C), ANOVA with ART revealed significant main effects of the ground scene (F (3, 39) = 5.51, p = .0030, $\eta_p^2 = 0.30$) and scene–vibration congruency (F (2, 26) = 78.47, p < .0001, $\eta_p^2 = 0.86$), and the interaction (F (6, 78) = 4.31, p = .0008, $\eta_p^2 = 0.25$).

The analysis of the simple main effects showed that the effects of scene–vibration congruency were significant in all

TABLE 1. The post hoc multiple-comparison analysis of the simple effects of the scene-vibration congruency condition for each ground scene.

Question	Scene	Pair	t	adj.p	Cohen's d
A. Self-motion	Indoor	Cong - Incong	1.579	0.2726	0.597
		Cong - None	5.844	< 0.0001 ***	2.209
		Incong - None	4.265	0.0007 ***	1.612
	Grass	Cong - Incong	-1.183	0.4739	-0.447
		Cong - None	3.081	0.0130 *	1.165
		Incong - None	4.264	0.0007 ***	1.612
	Concrete	Cong - Incong	0.496	0.8739	0.187
		Cong - None	3.880	0.0018 **	1.467
		Incong - None	3.384	0.0062 **	1.279
	Snow	Cong - Incong	2.517	0.0466 *	0.951
		Cong - None	6.815	< 0.0001 ***	2.576
		Incong - None	4.298	0.0006 ***	1.624
B. Walking	Indoor	Cong - Incong	2.616	0.0376 *	0.989
		Cong - None	9.544	< 0.0001 ***	3.607
		Incong - None	6.928	< 0.0001 ***	2.619
	Grass	Cong - Incong	-1.025	0.5678	-0.388
		Cong - None	6.366	< 0.0001 ***	2.406
		Incong - None	7.392	< 0.0001 ***	2.794
	Concrete	Cong - Incong	0.983	0.5936	0.372
		Cong - None	6.975	< 0.0001 ***	2.636
		Cong - None	5.991	< 0.0001 ***	2.265
	Snow	Cong - Incong	2.165	0.0963	0.818
		Cong - None	9.264	< 0.0001 ***	3.502
		Incong - None	7.099	< 0.0001 ***	2.683
C. Leg-Action	Indoor	Cong - Incong	3.363	0.0066 **	1.271
		Cong - None	10.458	< 0.0001 ***	3.953
		Incong - None	7.095	< 0.0001 ***	2.682
	Grass	Cong - Incong	-1.054	0.5507	-0.398
		Cong - None	7.834	< 0.0001 ***	2.961
		Incong - None	8.887	< 0.0001 ***	3.359
	Concrete	Cong - Incong	1.128	0.5061	0.426
		Cong - None	8.472	< 0.0001 ***	3.202
		Incong - None	7.345	< 0.0001 ***	2.776
	Snow	Cong - Incong	2.624	0.0370 *	0.992
		Cong - None	10.030	< 0.0001 ***	3.791
		Incong - None	7.406	< 0.0001 ***	2.799
D.Telepresence	Indoor	Cong - Incong	1.484	0.3149	0.561
		Cong - None	4.279	0.0006 ***	1.617
		Incong - None	2.796	0.0252 *	1.057
	Grass	Cong - Incong	-0.270	0.9605	-0.102
		Cong - None	3.364	0.0065 **	1.272
		Incong - None	3.635	0.0033 **	1.374
	Concrete	Cong - Incong	0.754	0.7336	0.285
		Cong - None	4.184	0.0008 ***	1.581
		Incong - None	3.430	0.0056 **	1.296
	Snow	Cong - Incong	2.513	0.0470 *	0.950
		Cong - None	5.510	< 0.0001 ***	2.083
		Incong - None	2.997	0.0158 *	1.133

* indicates p < 0.05, ** indicates p < 0.01, ***indicates p < 0.001

TABLE 2. The post hoc multiple-comparison analysis of the simple effects of the scene difference for the scene at the congruent condition.

Question	Pair	t	adj.p	Cohen's d
A. Self-motion	Concrete - Grass	1.725	0.3249	0.652
	Concrete - Indoor	-1.870	0.2575	-0.707
	Concrete - Snow	-1.652	0.3622	-0.625
	Grass - Indoor	-3.595	0.0048 **	-1.359
	Grass - Snow	-3.377	0.0087 **	-1.276
	Indoor - Snow	0.217	0.9963	0.082
B. Walking	Concrete - Grass	2.326	0.1095	0.879
	Concrete - Indoor	-1.585	0.3990	-0.599
	Concrete - Snow	-0.877	0.8169	-0.331
	Grass - Indoor	-3.911	0.0019 **	-1.478
	Grass - Snow	-3.203	0.0138 *	-1.211
	Indoor - Snow	0.708	0.8933	0.268
C. Leg-Action	Concrete - Grass	3.013	0.0225 *	1.139
	Concrete - Indoor	-1.577	0.4030	-0.596
	Concrete - Snow	-1.719	0.3280	-0.650
	Grass - Indoor	-4.591	0.0003 ***	-1.735
	Grass - Snow	-4.732	0.0002 ***	-1.789
	Indoor - Snow	-0.141	0.9990	-0.053
D. Telepresence	Concrete - Grass	1.168	0.6499	0.442
	Concrete - Indoor	-2.193	0.1432	-0.829
	Concrete - Snow	-2.553	0.0672	-0.965
	Grass - Indoor	-3.361	0.0091 **	-1.270
	Grass - Snow	-3.722	0.0034 **	-1.407
	Indoor - Snow	-0.361	0.9837	-0.136

* indicates p < 0.05, ** indicates p < 0.01, ***indicates p < 0.001

the ground scenes (indoor: F (2, 26) = 57.01, $p < .0001$, $\eta_p^2 = 0.81$; grass: F (2, 26) = 47.15, $p < .0001$, $\eta_p^2 = 0.78$; concrete: F (2, 26) = 42.33, $p < .0001$, $\eta_p^2 = 0.77$; snow: F (2, 26) = 54.11, $p < .0001$, $\eta_p^2 = 0.81$), whereas the effects of the different ground scenes were significant in the congruent vibration conditions (congruent: F (3, 39) = 9.67, $p < .0001$, $\eta_p^2 = 0.43$) but not significant in the incongruent and no-vibration conditions (incongruent: F (3, 39) = 1.82, p = .16, $\eta_p^2 = 0.12$; no-vibration: F (3, 39) = 0.20, p = .90, $\eta_p^2 = 0.015$).

For all ground scenes, the scores for the congruent and incongruent conditions were significantly higher than those for the no-vibration condition. For the indoor and snowy ground scenes, the scores for the congruent condition were significantly higher than those for the incongruent condition (Table 1C). The scores for the indoor-scene, concrete-scene, and snowy-scene conditions were significantly higher than for the grassy-scene condition only in the scene–vibration-congruent condition (Table 2C).

4) TELEPRESENCE SENSATION

For telepresence (Figure 6D), ANOVA with ART revealed significant main effects for the ground scene (F (3, 39) = 4.70, p = .0068, $\eta_p^2 = 0.27$), scene–vibration congruency (F (2, 26) = 13.07, p = .0001, $\eta_p^2 = 0.50$), and interaction (F (6, 78) = 2.77, p = .017, $\eta_p^2 = 0.18$).

The analysis of the simple main effects showed that the effects of the scene–vibration congruency were significant in all the ground scenes (indoor: F (2, 26) = 9.44, p = .0008, $\eta_p^2 = 0.42$; grass: F (2, 26) = 8.20, p = .0017, $\eta_p^2 = 0.39$; concrete: F (2, 26) = 9.94, p = .0006, $\eta_p^2 = 0.43$; snow: F (2, 26) = 15.22, $p < .0001$, $\eta_p^2 = 0.54$), whereas the effects of the different ground scenes were significant in the congruent vibration conditions (congruent: F (3, 39) = 6.33, p = .001, $\eta_p^2 = 0.33$) but not significant in the incongruent and no-vibration conditions (incongruent: F (3, 39) = 0.39, p = 0.76, $\eta_p^2 = 0.029$; no-vibration: F (3, 39) = 1.61, p = .20, $\eta_p^2 = 0.11$).

For all ground scenes, the scores for the congruent and incongruent conditions were significantly higher than those for the no-vibration condition. The score for the congruent condition was significantly higher than for the incongruent condition only for the snowy ground scenes (Table 1D). The scores for the indoor-scene and snowy-scene conditions were significantly higher than for the grassy-scene condition only in the scene–vibration-congruent condition (Table 2D).

5) SUMMARY OF WALKING-RELATED SENSATIONS AND TELEPRESENCE

In summary, the rhythmic foot vibration improved sensations of self-motion, walking, leg action, and telepresence regardless of its congruency with the ground scene; the congruent vibrations were better than the incongruent (scene unmatched) vibrations for self-motion sensation in the snowy scene, for walking sensation in the indoor scene, for leg-action sensation in both the indoor and snowy scenes, and for telepresence in the snowy scene. In the scene-congruent vibration condition, the indoor and snowy scenes were better than the grassy scene for self-motion, walking, and telepresence sensations, and the indoor, concrete, and snowy scenes were better than the grassy scene for the leg-action sensations.

B. GROUND PERCEPTION

1) SMOOTH–ROUGH PERCEPTION

For the ground perception of smoothness and roughness (Figure 7A), ANOVA revealed significant main effects of the ground scene ($F(3, 39) = 26.77$, $p < .0001$, $\eta_p^2 = 0.67$), scene–vibration congruency ($F(1.22, 15.86) = 7.09$, $p = .013$, $\eta_p^2 = 0.35$), and interaction ($F(6, 78) = 10.63$, $p < .0001$, $\eta_p^2 = 0.45$).

An analysis of the simple main effects showed that the effects of the scene–vibration congruency were significant in all the ground scenes (indoor: $F(2, 26) = 12.20$, $p = .0002$, $\eta_p^2 = 0.48$; grass: $F(1.37, 17.82) = 6.58$, $p = .013$, $\eta_p^2 = 0.34$; concrete: $F(2, 26) = 5.04$, $p = .014$, $\eta_p^2 = 0.28$; snow: $F(1.38, 17.99) = 11.20$, $p = .0018$, $\eta_p^2 = 0.46$), whereas the effects of the different ground scenes were significant in the congruent and the no-vibration conditions (congruent: $F(1.86, 24.21) = 35.80$, $p < .0001$, $\eta_p^2 = 0.73$; no-vibration: $F(1.22, 15.91) = 9.58$, $p = .0049$, $\eta_p^2 = 0.42$) but not significant in the incongruent vibration condition (incongruent: $F(3, 39) = 1.99$, $p = .13$, $\eta_p^2 = 0.13$).

The indoor scene was perceived as generally smooth, rather than rough and smoother in the scene–vibration-congruent and no-vibration conditions than in the incongruent condition. The grassy ground scene was perceived to be rough in the congruent condition, neutral in the incongruent condition, and smooth in the no-vibration condition. The concrete road scene was perceived as generally smooth rather than rough, and there was no significant difference between the vibration conditions ($p > .05$). The snowy scene was perceived to be rough in the congruent condition, neutral in the incongruent condition, and smooth in the no-vibration condition. The details are listed in Table 3A.

The grassy ground and the snowy ground scenes were perceived as significantly rougher than the indoor and concrete road scenes in the scene–vibration-congruent condition (Table 4A).

2) SOFT–HARD PERCEPTION

For the ground perception regarding softness and hardness (Figure 7B), ANOVA with ART revealed a significant main effect of the ground scene ($F(3, 39) = 47.07$, $p < .0001$, $\eta_p^2 = 0.78$) and the interaction ($F(6, 78) = 25.14$, $p < .0001$, $\eta_p^2 = 0.66$).

The analysis of the simple main effects showed that the effects of the scene–vibration congruency were significant in the all ground scenes (indoor: $F(2, 26) = 18.53$, $p < .0001$, $\eta_p^2 = 0.59$; grass: $F(2, 26) = 17.11$, $p < .0001$, $\eta_p^2 = 0.57$; concrete: $F(2, 26) = 12.37$, $p = .0002$, $\eta_p^2 = 0.49$; snow: $F(2, 26) = 33.72$, $p < .0001$, $\eta_p^2 = 0.72$), whereas the effect of the different ground scenes was significant both in the congruent and the no-vibration conditions (congruent: $F(3, 39) = 58.10$, $p < .0001$, $\eta_p^2 = 0.82$; no-vibration: $F(3, 39) = 22.70$, $p < .0001$, $\eta_p^2 = 0.64$) but not significant in the incongruent vibration condition (incongruent: $F(3, 39) = 1.58$, $p = .21$, $\eta_p^2 = 0.11$).

TABLE 3. The post hoc multiple-comparison analysis of the simple effects of the scene-vibration congruency condition for each ground scene.

Question	Scene	Pair	t	adj.p	Cohen's d
A. Smothness and Roughness	Indoor	Cong - Incong	-4.686	0.0013 **	-1.515
		Cong - None	-1.087	0.2968	-0.284
		Incong - None	3.224	0.0067 **	1.241
	Grass	Cong - Incong	3.088	0.0260 *	0.823
		Cong - None	2.916	0.0260 *	1.556
		Incong - None	1.663	0.1202	0.739
	Concrete	Cong - Incong	-2.052	0.0609	-0.650
		Cong - None	1.305	0.2144	0.298
		Incong - None	2.660	0.0589	1.094
	Snow	Cong - Incong	4.763	0.0011 **	1.377
		Cong - None	4.347	0.0011 **	2.250
		Incong - None	1.523	0.1516	0.780
B. Softness and Hardness	Indoor	Cong - Incong	6.013	< 0.0001 ***	2.273
		Cong - None	3.826	0.0021 **	1.446
		Incong - None	-2.187	0.0924	-0.826
	Grass	Cong - Incong	-5.835	< 0.0001 ***	-2.205
		Cong - None	-3.282	0.0080 **	-1.240
		Incong - None	2.553	0.0431 *	0.965
	Concrete	Cong - Incong	4.183	0.0008 ***	1.581
		Cong - None	4.421	0.0004 ***	1.671
		Incong - None	0.238	0.9693	0.090
	Snow	Cong - Incong	-8.204	< 0.0001 ***	-3.101
		Cong - None	-4.418	0.0004 ***	-1.670
		Incong - None	3.785	0.0023 **	1.431
C. Wetness and Dryness	Indoor	Cong - Incong	3.906	0.0017 **	1.476
		Cong - None	2.509	0.0475 *	0.948
		Incong - None	-1.396	0.3574	-0.528
	Concrete	Cong - Incong	3.279	0.0080 **	1.239
		Cong - None	4.250	0.0007 ***	1.606
		Incong - None	0.971	0.6012	0.367
	Snow	Cong - Incong	-4.212	0.0008 ***	-1.592
		Cong - None	-3.162	0.0107 *	-1.195
		Incong - None	1.050	0.5526	0.397
D. Slipperiness and Stickiness	Indoor	Cong - Incong	-1.738	0.2105	-0.657
		Cong - None	2.910	0.0194 *	1.100
		Incong - None	4.648	0.0002 ***	1.757
	Grass	Cong - Incong	1.277	0.4207	0.482
		Cong - None	2.901	0.0198 *	1.097
		Incong - None	1.625	0.2535	0.614
	Concrete	Cong - Incong	-1.044	0.5564	-0.395
		Cong - None	2.390	0.0611	0.903
		Incong - None	3.434	0.0055 **	1.298
	Snow	Cong - Incong	3.126	0.0116 *	1.182
		Cong - None	4.767	0.0002 ***	1.802
		Incong - None	1.641	0.2470	0.620

* indicates $p < 0.05$, ** indicates $p < 0.01$, ***indicates $p < 0.001$

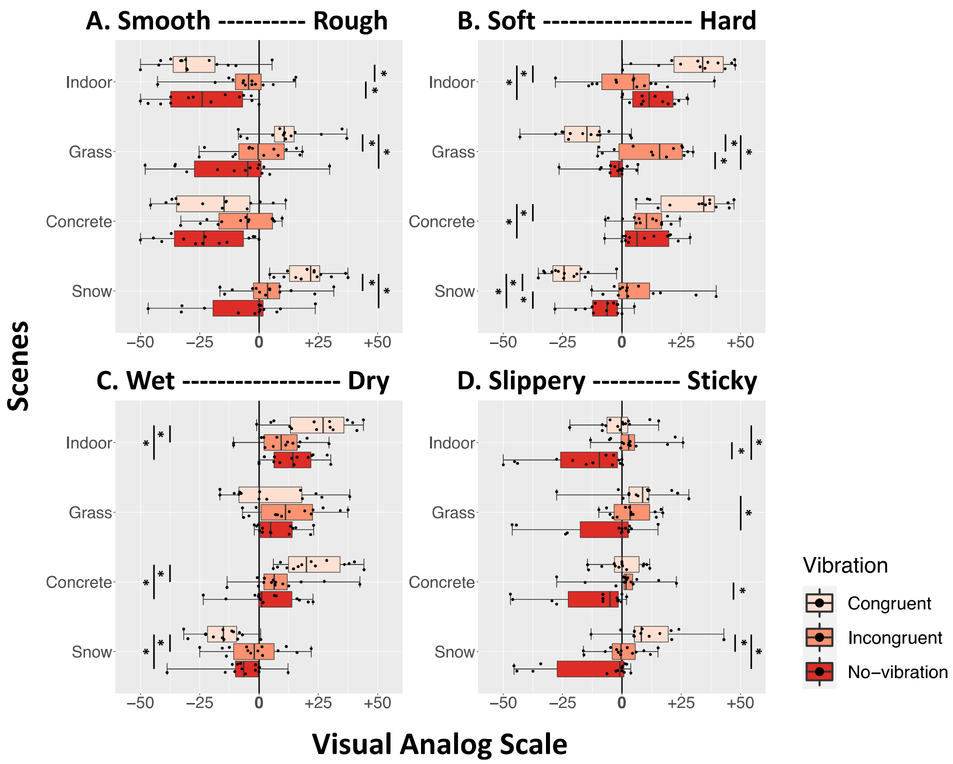


FIGURE 7. Results of ground material perception for each pair of adjectives: (A) smooth–rough, (B) soft–hard, (C) wet–dry, and (D) slippery–sticky. The horizontal axis indicates the value of digitized VAS (−50 to +50) for each pair of adjectives. The vertical axis indicates four ground scenes. A box indicates the range between Q1 (25%) and Q3 (75%), the thick line in the box indicates the median, and a set of whiskers indicates the maximum and minimum values. * indicates p < .05.

TABLE 4. The post hoc multiple-comparison analysis of the simple effects of the scene difference for the scene at the congruent condition.

Question	Pair	t	adj.p		Cohen's d
A. Smothness and Roughness	Concrete - Grass	-4.139	0.0035	**	-1.812
	Concrete - Indoor	2.469	0.0563		0.598
	Concrete - Snow	-6.336	0.0001	***	-2.373
	Grass - Indoor	-7.206	< 0.0001	***	-2.743
	Grass - Snow	-2.347	0.0563		-0.633
	Indoor - Snow	-8.663	< 0.0001	***	-3.296
B. Softness and Hardness	Concrete - Grass	8.205	< 0.0001	***	3.101
	Concrete - Indoor	-0.166	0.9983		-0.063
	Concrete - Snow	10.104	< 0.0001	***	3.819
	Grass - Indoor	-8.371	< 0.0001	***	-3.164
	Grass - Snow	1.900	0.2448		0.718
	Indoor - Snow	10.271	< 0.0001	***	3.882
C. Wetness and Dryness	Concrete - Grass	3.463	0.0069	**	1.309
	Concrete - Indoor	-0.337	0.9867		-0.127
	Concrete - Snow	7.218	< 0.0001	***	2.728
	Grass - Indoor	-3.799	0.0027	**	-1.436
	Grass - Snow	3.755	0.0030	**	1.419
	Indoor - Snow	7.555	< 0.0001	***	2.855
D. Slipperiness and Stickiness	Concrete - Grass	-3.012	0.0226	*	-1.138
	Concrete - Indoor	0.240	0.9951		0.091
	Concrete - Snow	-3.546	0.0055	**	-1.340
	Grass - Indoor	3.252	0.0122	*	1.229
	Grass - Snow	-0.534	0.9502		-0.202
	Indoor - Snow	-3.786	0.0028	**	-1.431

* indicates p < 0.05, ** indicates p < 0.01, ***indicates p < 0.001

The indoor scene was perceived as generally hard, rather than soft and harder in the scene–vibration congruent condition than in the incongruent and no-vibration conditions. The grassy ground scene was perceived to be soft in the congruent condition, neutral in the no-vibration condition, and hard in the incongruent condition. The concrete road scene was perceived as generally hard, rather than soft and harder in the congruent condition than in the incongruent and no-vibration conditions. The snowy scene was perceived to be softer in the congruent condition than in the no-vibration condition and slightly harder in the incongruent condition. The details are listed in Table 3B.

The indoor and concrete road scenes were perceived as significantly harder than the grassy ground and snowy ground scenes both in the scene–vibration-congruent condition and in the no-vibration condition (Table 4B).

3) WET–DRY PERCEPTION

For the ground perception regarding wetness and dryness (Figure 7C), ANOVA with ART revealed a significant main effect of the ground scene ($F(3, 39) = 19.64$, $p < .0001$, $\eta_p^2 = 0.60$) and the interaction ($F(6, 78) = 9.31$, $p < .0001$, $\eta_p^2 = 0.42$).

Analysis of the simple main effects showed that the effects of the scene–vibration congruency were significant in all the ground scenes except for the grassy ground scene (indoor: $F(2, 26) = 7.83$, $p = .0022$, $\eta_p^2 = 0.38$; grass: $F(2, 26) = 2.42$, $p = .11$, $\eta_p^2 = 0.16$; concrete: $F(2, 26) = 9.92$, $p = .0006$, $\eta_p^2 = 0.43$; snow: $F(2, 26) = 9.62$, $p = .0007$, $\eta_p^2 = 0.43$), whereas the effects of the different ground scenes were significant in all vibration conditions (congruent: F (3, 39) = 24.97, p < .0001, η_p^2 = 0.66; incongruent: F (3, 39) = 2.98, p = .043, η_p^2 = 0.19; no-vibration: F (3, 39) = 11.90, p < .0001, η_p^2 = 0.48).

The indoor scene was perceived as generally dry, rather than wet, and dryer in the scene–vibration-congruent condition than in the incongruent and the no-vibration conditions. The grassy ground scene was generally perceived as dry, rather than wet, and there were no differences between the vibration conditions. The concrete road scene was perceived to be dryer in the congruent condition than in the incongruent and no-vibration conditions. The snowy scene was perceived to be wetter in the congruent condition than in the no-vibration and incongruent conditions. The details are listed in Table 3C.

The indoor and concrete road scenes were perceived as significantly dryer than the grassy and snowy ground scenes, and the grassy ground scene was perceived to be dryer than the snowy ground scene in the scene–vibration-congruent condition. In the incongruent vibration condition, the grassy ground scene was perceived to be dryer than the snowy ground scene. In the no-vibration condition, the indoor and grassy ground scenes were perceived to be dryer than the snowy ground scene, and the concrete road scene was perceived to be dryer than the snowy ground scene. The details are presented in Table 4C.

4) SLIPPERY–STICKY PERCEPTION

For the ground perception of slipperiness and stickiness (Figure 7D), ANOVA with ART revealed significant main effects of the ground scene (F (3, 39) = 4.83, p = .006, η_p^2 = 0.27), scene–vibration congruency (F (2, 26) = 13.75, p < .0001, η_p^2 = 0.51), and interaction (F (6, 78) = 4.19, p = .001, η_p^2 = 0.24).

The analysis of the simple main effects showed that the effects of the scene–vibration congruency were significant in all the ground scenes (indoor: F (2, 26) = 11.03, p = .0003, η_p^2 = 0.46; grass: F (2, 26) = 4.23, p = .026, η_p^2 = 0.25; concrete: F (2, 26) = 6.20, p = .006, η_p^2 = 0.32; snow: F (2, 26) = 11.73, p = .0002, η_p^2 = 0.47), whereas the effect of the different ground scenes was significant in the congruent and the no-vibration conditions (congruent: F (3, 39) = 7.82, p = .0003, η_p^2 = 0.38; no-vibration: F (3, 39) = 7.79, p = .0003, η_p^2 = 0.37) but not significant in the incongruent vibration condition (incongruent: F (3, 39) = 0.47, p = .71, η_p^2 = 0.035).

The indoor scene was perceived as more slippery in the no-vibration condition than in the scene–vibration congruent and incongruent conditions. The grassy ground scene was perceived to be stickier in the congruent condition than in the no-vibration condition. The concrete road scene was perceived to be more slippery in the no-vibration condition than in the incongruent condition. The snowy scene was perceived as stickier in the congruent condition, neutral in the incongruent condition, and slippery in the no-vibration condition. The details are presented in Table 3D.

The grassy and snowy ground scenes were perceived as significantly stickier than the indoor and concrete road scenes in the scene–vibration-congruent condition. In the no-vibration condition, the concrete road scene was perceived to be more slippery than the grassy ground scene, and the indoor scene was perceived to be more slippery than the grassy and snowy ground scenes. The details are presented in Table 4D.

5) SUMMARY OF GROUND PERCEPTION

Scene-congruent vibrations enhanced the perception of ground materials. The indoor scene was perceived to be harder and dryer with congruent vibration than in the no-vibration condition (Figure 8A). The grassy ground scene was perceived as softer and stickier with congruent vibration than in the no-vibration condition (Figure 8B). The concrete road scene was perceived to be harder and dryer with congruent vibration than in the no-vibration condition (Figure 8C). The snowy ground scene was perceived as rougher, softer, wetter, and stickier with congruent vibration than in the no-vibration condition (Figure 8D). However, the scene-incongruent vibration caused the perception of the ground materials to deteriorate. For example, grassy and snowy ground scenes were perceived to be harder with incongruent vibration than with congruent vibration or without vibration.

IV. DISCUSSION

A. SUMMARY OF RESULTS

The timings of footsteps were generated using motion trajectory estimation with OpenVSLAM applied to four omnidirectional movies. Scene-congruent/matched vibrations were prepared for the four ground scenes. Then, the effects of congruent and incongruent vibrations on walking-related sensations and ground material perceptions were investigated using psychological measurements. The results showed that rhythmic foot vibration improved sensations of self-motion, walking, leg action, and telepresence with congruent and incongruent vibrations. Moreover, congruent vibrations were better than incongruent vibrations for self-motion sensation in snowy scenes, walking sensation in indoor scenes, leg action sensation in indoor and snowy scenes, and telepresence in snowy scenes. The perception of ground materials was enhanced by scene-congruent vibrations. The indoor scene was perceived to be harder and dryer, the grassy ground scene was perceived as softer and stickier, the concrete road scene was perceived as harder and dryer, and the snowy ground scene was perceived to be rougher, softer, wetter, and stickier with the congruent vibration than the no-vibration condition. However, ground perception was confused by scene-incongruent vibrations, so that the grassy and snowy ground scenes were perceived to be harder with

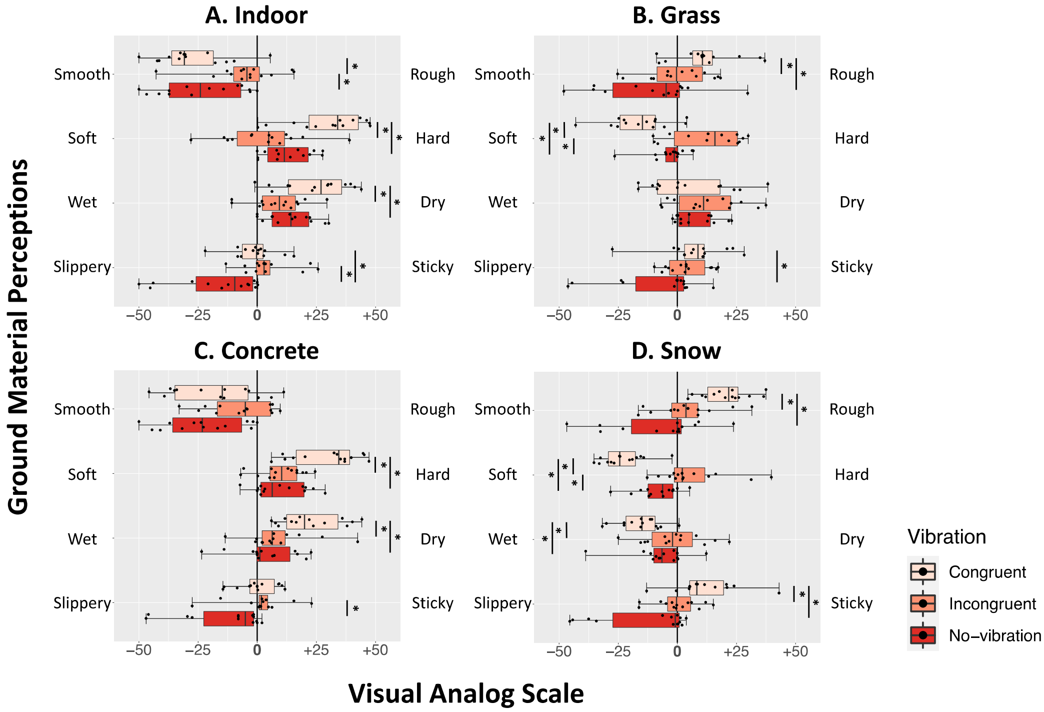


FIGURE 8. Results of ground material perception for each ground scene (replot of Figure 7): (A) indoor corridor scene, (B) grassy ground scene, (C) concrete paved road scene, and (D) snowy ground scene. The horizontal axis indicates the value of digitized VAS (−50 to +50) for each pair of adjectives. The vertical axis indicates different pairs of adjectives. A box indicates the range between Q1 (25%) and Q3 (75%), the thick line in the box indicates the median, and a set of whiskers indicates the maximum and minimal values. * indicates $p < .05$.

incongruent vibration than congruent vibration or without vibration.

B. SCENE-CONGRUENT VIBRATIONS IMPROVE VIRTUAL WALKING

The sensations of self-motion, walking, leg action, and telepresence were improved by scene-congruent vibrations over incongruent vibrations depending on the scene. This effect was strongest in the snowy ground scene. The vibrations of the snowy scene contained a prolonged, moderate component (0.02–0.15 s) followed by a second, weaker component (0.23–0.36 s) (see Figure 4D). It is speculated that the vibration pattern of the snowy-scene vibrations was characteristic in comparison with the other vibrations, so that its effect was stronger than in the other scenes. Generally, rhythmic foot vibrations enhance walking-related sensations and telepresence regardless of scene–vibration congruency. This result suggests that it is unnecessary for virtual walking experience to present exact vibration patterns matched to the ground materials; rather, rhythmic timings simulating the foot striking the ground are important, as shown in previous studies [30], [37], [38]. Kitazaki *et al.* [37] reported that sensations of self-motion, walking, leg action, and telepresence were perceived better from an oscillating visual flow with synchronized foot vibrations than with randomized-timing vibrations, and that a 250-ms delay of foot vibrations with respect to the scenes caused the walking-related sensations to deteriorate. Thus, the present results are consistent with those of previous studies. In addition, it is suggested that scene-congruent vibration patterns improve virtual walking experiences.

C. SCENE-CONGRUENT VIBRATIONS ENHANCE PERCEPTION OF GROUND MATERIALS

The results of this study indicated that scene-congruent vibrations modulated the perception of ground materials appropriately, as expected, in different scenes. This could be caused by the ground compliance perception based on the foot vibration patterns [64] and a human multimodal integration process [65], [66]. Tactile sensations on the feet during walking depend on the materials of the shoes and socks, conditions of the skin, and types of leg action. Thus, it is difficult to present precise or exact foot vibration patterns that match the ground. Nevertheless, in this study, scene-congruent vibrations enhanced ground perceptions. The effect of congruent vibrations was most explicit in snowy ground scenes. This result suggests that the characteristic vibration pattern of the snowy scene is reliable or informative and has an advantage over the other scenes. If visual scenes deteriorate, such as become blurred, the effect of vibration congruency may be more explicit because the reliability of the vibration increases relative to the vision, according to the statistically optimal integration model [65], [67]. This should be investigated in future studies.

D. LIMITATIONS

The scene-congruent vibrations were chosen arbitrarily from libraries of footstep sounds. Thus, scenes and sounds were not obtained in identical situations. Foot vibrations were presented by driving vibro-transducers with sound sources. This is a simple and easy method to present different vibrations depending on the ground scene; however, the tactile sensations on the foot and footstep sounds are not exactly the same. Thus, the validity of congruency between scenes and vibrations should be discounted. However, the results showed the enhancement of walking-related sensations and ground perception by congruent vibrations, so that the selection of vibrations would be appropriate to compare the congruent, incongruent, and no-vibration conditions. If the vibration patterns are automatically generated from visual scenes by estimating ground materials and/or compliance, any omnidirectional movies that are already available on the Internet can be converted to virtual walking. This requires further investigation.

It is another limitation that only healthy people were involved in the study. We should investigate the virtual walking system with persons who have walking disabilities in future study. Since only four common scenarios were employed in the experiment, it is necessary to verify whether the system can be applicable to a variety of further scenes and scenarios. The system is not only limited to walking experiences, but also applicable to jogging or running. Thus, we would like to further consider investigating other activities such as jogging.

Only vision and foot vibrations were presented without sound to investigate the effect of scene–vibration congruency. However, footstep sounds also contribute to virtual walking experiences [30], [31]. Other modalities, such as audition, should be added to improve virtual walking in future research.

V. CONCLUSION

A virtual walking system for seated users using omnidirectional movies with foot vibrations was developed. The virtual viewpoint/camera transition was estimated using visual SLAM. The timings of foot strikes were generated, and vibrations were presented to the left and right forefeet and heels rhythmically. Scene-congruent vibrations were found to improve walking-related sensations and telepresence and enhance the perception of ground materials. These results suggest that we can provide effective virtual walking experiences, including ground-material perception, by presenting foot vibrations that are well matched to the ground scene. Based on our virtual walking system, we can convert various public omnidirectional movies into realistic virtual walking with visual–tactile integration and provide virtual travel and walking experiences at various locations around the world. It could also contribute to improving the mental health and well-being of people even if they have walking disabilities or are unable to walk due to physical limitations or social limitations such as the COVID-19 pandemic.

ACKNOWLEDGMENT

The authors are would like to thank Editage (www.editage.com) for English language editing. An earlier version of this paper was presented at the IEEE VR Conference [77] [DOI: 10.1109/VRW52623.2021.00186].

REFERENCES

[1] P. Kelly, C. Williamson, A. G. Niven, R. Hunter, N. Mutrie, and J. Richards, "Walking on sunshine: Scoping review of the evidence for walking and mental health," *Brit. J. Sports Med.*, vol. 52, no. 12, pp. 800–806, Jun. 2018, doi: 10.1136/bjsports-2017-098827.

[2] J. A. Giandonato, V. M. Tringali, and R. C. Thoms, "Improving mental health through physical activity: A narrative literature review," *Phys. Activity Health*, vol. 5, no. 1, pp. 146–153, Aug. 2021, doi: 10.5334/paah.108.

[3] M. Mau, A. Aaby, S. H. Klausen, and K. K. Roessler, "Are long-distance walks therapeutic? A systematic scoping review of the conceptualization of long-distance walking and its relation to mental health," *Int. J. Environ. Res. Public Health*, vol. 18, no. 15, p. 7741, Jul. 2021, doi: 10.3390/ijerph18157741.

[4] M. A. Alfonzo, "To walk or not to walk? The hierarchy of walking needs," *Environ. Behav.*, vol. 37, no. 6, pp. 808–836, Nov. 2005, doi: 10.1177/0013916504274016.

[5] A. Castañeda-Babarro, A. Arbillaga-Etxarri, B. Gutiérrez-Santamaría, and A. Coca, "Physical activity change during COVID-19 confinement," *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, p. 6878, Sep. 2020, doi: 10.3390/ijerph17186878.

[6] Y. Wang, Y. Zhang, K. Bennell, D. K. White, J. Wei, Z. Wu, H. He, S. Liu, X. Luo, S. Hu, C. Zeng, and G. Lei, "Physical distancing measures and walking activity in middle-aged and older residents in Changsha, China, during the COVID-19 epidemic period: Longitudinal observational study," *J. Med. Internet Res.*, vol. 22, no. 10, Oct. 2020, Art. no. e21632, doi: 10.2196/21632.

[7] A. Doubleday, Y. Choe, T. B. Isaksen, S. Miles, and N. A. Errett, "How did outdoor biking and walking change during COVID-19?: A case study of three U.S. cities," *PLoS ONE*, vol. 16, no. 1, Jan. 2021, Art. no. e0245514, doi: 10.1371/journal.pone.0245514.

[8] H. Iwata, "Locomotion interfaces," in *Human Walking Virtual Environments*. New York, NY, USA: Springer, 2013, pp. 199–219.

[9] N. C. Nilsson, S. Serafin, F. Steinicke, and R. Nordahl, "Natural walking in virtual reality: A review," *Comput. Entertainment*, vol. 16, no. 2, pp. 1–22, Apr. 2018, doi: 10.1145/3180658.

[10] M. Slater, M. Usoh, and A. Steed, "Taking steps: The influence of a walking technique on presence in virtual reality," *ACM Trans. Comput.-Hum. Interact.*, vol. 2, no. 3, pp. 201–219, Sep. 1995, doi: 10.1145/210079.210084.

[11] J. Templeman, P. Denbrook, and L. Sibert, "Virtual locomotion: Walking in place through virtual environments," *Presence*, vol. 8, no. 6, pp. 598–617, Dec. 1999, doi: 10.1162/105474699566512.

[12] J. Feasel, M. C. Whitton, and J. D. Wendt, "LLCM-WIP: Low-latency, continuous-motion walking-in-place," in *Proc. IEEE Symp. 3D User Interfaces*, Mar. 2008, pp. 97–104.

[13] N. C. Nilsson, S. Serafin, M. H. Laursen, K. S. Pedersen, E. Sikstrom, and R. Nordahl, "Tapping-in-place: Increasing the naturalness of immersive walking-in-place locomotion through novel gestural input," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2013, pp. 31–38.

[14] M. McCullough, H. Xu, J. Michelson, M. Jackoski, W. Pease, W. Cobb, W. Kalescky, J. Ladd, and B. Williams, "Myo arm: Swinging to explore a VE," in *Proc. ACM SIGGRAPH Symp. Appl. Perception*, Sep. 2015, pp. 107–113.

[15] P. T. Wilson, W. Kalescky, A. MacLaughlin, and B. Williams, "VR locomotion: Walking > walking in place > arm swinging," in *Proc. 15th ACM SIGGRAPH Conf. Virtual-Reality Continuum Its Appl. Ind.*, Dec. 2016, pp. 243–249.

[16] A. Harris, K. Nguyen, P. T. Wilson, M. Jackoski, and B. Williams, "Human joystick: Wii-leaning to translate in large virtual environments," in *Proc. 13th ACM SIGGRAPH Int. Conf. Virtual-Reality Continuum Appl. Ind. (VRCAI)*, 2014, pp. 231–234.

[17] T. Nguyen-Vo, B. E. Riecke, W. Stuerzlinger, D.-M. Pham, and E. Kruijff, "NaviBoard and NaviChair: Limited translation combined with full rotation for efficient virtual locomotion," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 1, pp. 165–177, Jan. 2021, doi: 10.1109/TVCG.2019.2935730.

[18] R. P. Darken, W. R. Cockayne, and D. Carmein, "The omni-directional treadmill: A locomotion device for virtual worlds," in *Proc. 10th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, New York, NY, USA, 1997, pp. 213–221, doi: 10.1145/263407.263550.

[19] H. Iwata, "Walking about virtual environments on an infinite floor," in *Proc. IEEE Virtual Reality*, Houston TX, USA, Mar. 1999, pp. 286–293, doi: 10.1109/VR.1999.756964.

[20] J. L. Souman, P. R. Giordano, M. Schwaiger, I. Frissen, T. Thümmel, H. Ulbrich, A. De Luca, H. H. Bülthoff, and M. O. Ernst, "CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments," *ACM Trans. Appl. Perception*, vol. 8, no. 4, pp. 1–22, Nov. 2011, doi: 10.1145/2043603.2043607.

[21] H. Iwata, H. Yano, and F. Nakaizumi, "Gait master: A versatile locomotion interface for uneven virtual terrain," in *Proc. IEEE Virtual Reality*, Yokohama, Japan, Mar. 2001, pp. 131–137, doi: 10.1109/VR.2001.913779.

[22] H. Iwata, H. Yano, H. Fukushima, and H. Noma, "CirculaFloor: A locomotion interface using circulation of movable tiles," in *Proc. IEEE Virtual Reality Conf. (VR)*, Bonn, Germany, Mar. 2005, pp. 223–230, doi: 10.1109/VR.2005.11.

[23] E. Medina, R. Fruland, and S. Weghorst, "Virtusphere: Walking in a human size VR 'hamster ball,'" in *Proc. Hum. Factors Ergon. Soc. Annu. Meeting*, New York, NY, USA, vol. 27, no. 27, Sep. 2008, pp. 2102–2106, doi: 10.1177/154193120805202704.

[24] J. P. Freiwald, O. Ariza, O. Janeh, and F. Steinicke, "Walking by cycling: A novel in-place locomotion user interface for seated virtual reality experiences," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–12, doi: 10.1145/3313831.3376574.

[25] T. Cakmak and H. Hager, "Cyberith virtualizer: A locomotion device for virtual reality," in *Proc. ACM SIGGRAPH Emerg. Technol. (SIGGRAPH)*, 2014, p. 6, doi: 10.1145/2614066.2614105.

[26] D. Swapp, J. Williams, and A. Steed, "The implementation of a novel walking interface within an immersive display," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2010, pp. 71–74.

[27] S. Razzaque, D. Swapp, M. Slater, M. C. Whitton, and A. Steed, "Redirected walking in place," in *Eight Eurographics Workshop on Virtual Environments*, S. Müller and W. Stüzlinger, Eds. May 2002, pp. 123–130.

[28] E. A. Suma, G. Bruder, F. Steinicke, D. M. Krum, and M. Bolas, "A taxonomy for deploying redirection techniques in immersive virtual environments," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2012, pp. 43–46.

[29] A. Lecuyer, J.-M. Burkhardt, J.-M. Henaff, and S. Donikian, "Camera motions improve the sensation of walking in virtual environments," in *Proc. IEEE Virtual Reality Conf. (VR)*, Alexandria VA, USA, Mar. 2006, pp. 11–18, doi: 10.1109/VR.2006.31.

[30] L. Terziman, M. Marchal, F. Multon, B. Arnaldi, and A. Lecuyer, "The King-Kong effects: Improving sensation of walking in VR with visual and tactile vibrations at each step," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Costa Mesa, CA, USA, Mar. 2012, pp. 19–26, doi: 10.1109/3DUI.2012.6184179.

[31] L. Turchet, P. Burelli, and S. Serafin, "Haptic feedback for enhancing realism of walking simulations," *IEEE Trans. Haptics*, vol. 6, no. 1, pp. 35–45, 1st Quart., 2013, doi: 10.1109/TOH.2012.51.

[32] Y. Ikei, S. Shimabukuro, S. Kato, K. Komase, Y. Okuya, K. Hirota, M. Kitazaki, and T. Amemiya, "Five senses theatre project: Sharing experiences through bodily ultra-reality," in *Proc. IEEE Virtual Reality (VR)*, Arles, France, Mar. 2015, pp. 195–196, doi: 10.1109/VR.2015.7223362.

[33] H. Kaneko, T. Amemiya, V. Yem, Y. Ikei, K. Hirota, and M. Kitazaki, "Leg-jack: Generation of the sensation of walking by electrical and kinesthetic stimuli to the lower limbs," in *Proc. SIGGRAPH Asia Emerg. Technol.*, Tokyo, Japan, Dec. 2018, pp. 1–2.

[34] K. Yamaoka, R. Koide, T. Amemiya, M. Kitazaki, V. Yem, and Y. Ikei, "Rendering of walking sensation for a sitting user by lower limb motion display," in *Proc. ICAT-EGVE*, Tokyo, Japan, Sep. 2019, p. 14.

[35] T. Amemiya, M. Kitazaki, and Y. Ikei, "Pseudo-sensation of walking generated by passive whole-body motions in heave and yaw directions," *IEEE Trans. Haptics*, vol. 13, no. 1, pp. 80–86, Jan. 2020, doi: 10.1109/TOH.2020.2965937.

[36] E. Kruijff, A. Marquardt, C. Trepkowski, R. W. Lindeman, A. Hinkenjann, J. Maiero, and B. E. Riecke, "On your feet!: Enhancing vection in leaning-based interfaces through multisensory stimuli," in *Proc. Symp. Spatial User Interact.*, Oct. 2016, pp. 149–158, doi: 10.1145/2983310.2985759.

[37] M. Kitazaki, T. Hamada, K. Yoshiho, R. Kondo, T. Amemiya, K. Hirota, and Y. Ikei, "Virtual walking sensation by prerecorded oscillating optic flow and synchronous foot vibration," *i-Perception*, vol. 10, no. 5, Sep. 2019, Art. no. 204166951988244, doi: 10.1177/2041669519882448.

[38] Y. Matsuda, J. Nakamura, T. Amemiya, Y. Ikei, and M. Kitazaki, "Enhancing virtual walking sensation using self-avatar in first-person perspective and foot vibrations," *Frontiers Virtual Reality*, vol. 2, Apr. 2021, Art. no. 654088, doi: 10.3389/frvir.2021.654088.

[39] D. R. Labbe, K. Kouakoua, R. Aissaoui, S. Nadeau, and C. Duclos, "Proprioceptive stimulation added to a walking self-avatar enhances the illusory perception of walking in static participants," *Frontiers Virtual Reality*, vol. 2, p. 73, Jun. 2021, doi: 10.3389/frvir.2021.557783.

[40] B. E. Riecke, "Compelling self-motion through virtual environments without actual self-motion: Using self-motion illusions ('vection') to improve user experience in VR," in *Virtual Reality*, J. Kim, Ed. London, U.K.: IntechOpen, 2011, pp. 149–176.

[41] B. E. Riecke and J. Schulte-Pelkum, "Perceptual and cognitive factors for self-motion simulation in virtual environments: How can self-motion illusions ('vection') be utilized?" in *Human Walking Virtual Environments*. New York, NY, USA: Springer, 2013, pp. 27–54.

[42] J. Dichgans and T. Brandt, "Visual-vestibular interactions: Effects on self-motion perception and postural control," in *Handbook of Sensory Physiology 8*, R. Held, H. W. Leibowitz H. L. Teuber, Eds. Berlin, Germany: Springer, 1978, pp. 755–804.

[43] S. Palmisano, R. S. Allison, M. M. Schira, and R. J. Barry, "Future challenges for vection research: Definitions, functional significance, measures, and neural bases," *Frontiers Psychol.*, vol. 6, p. 193, Feb. 2015, doi: 10.3389/fpsyg.2015.00193.

[44] S. Sakamoto, Y. Osada, Y. Suzuki, and J. Gyoba, "The effects of linearly moving sound images on self-motion perception," *Acoust. Sci. Technol.*, vol. 25, no. 1, pp. 100–102, 2004, doi: 10.1250/ast.25.100.

[45] A. Väljamäe, "Auditorily-induced illusory self-motion: A review," *Brain Res. Rev.*, vol. 61, no. 2, pp. 240–255, 2009, doi: 10.1016/j.brainresrev.2009.07.001.

[46] B. E. Riecke, A. Väljamäe, and J. Schulte-Pelkum, "Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality," *ACM Trans. Appl. Perception*, vol. 6, no. 2, pp. 1–27, Feb. 2009, doi: 10.1145/1498700.1498701.

[47] B. E. Riecke, D. Feuereissen, and J. J. Rieser, "Auditory self-motion simulation is facilitated by haptic and vibrational cues suggesting the possibility of actual motion," *ACM Trans. Appl. Perception*, vol. 6, no. 3, pp. 1–22, Aug. 2009, doi: 10.1145/1577755.1577763.

[48] T. Brandt, J. Dichgans, and E. Koenig, "Differential effects of central versus peripheral vision on egocentric and exocentric motion perception," *Exp. Brain Res.*, vol. 16, no. 5, pp. 476–491, Mar. 1973, doi: 10.1007/BF00234474.

[49] R. Held, J. Dichgans, and J. Bauer, "Characteristics of moving visual scenes influencing spatial orientation," *Vis. Res.*, vol. 15, no. 3, pp. 357–365, 1975, doi: 10.1016/0042-6989(75)90083-8.

[50] B. E. Riecke and J. D. Jordan, "Comparing the effectiveness of different displays in enhancing illusions of self-movement (vection)," *Frontiers Psychol.*, vol. 6, p. 713, Jun. 2015, doi: 10.3389/fpsyg.2015.00713.

[51] T. Brandt, E. R. Wist, and J. Dichgans, "Foreground and background in dynamic spatial orientation," *Perception Psychophys.*, vol. 17, no. 5, pp. 497–503, Sep. 1975, doi: 10.3758/BF03203301.

[52] M. Ohmi, I. P. Howard, and J. P. Landolt, "Circular vection as a function of foreground-background relationships," *Perception*, vol. 16, no. 1, pp. 17–22, Feb. 1987, doi: 10.1068/p160017.

[53] M. Kitazaki and T. Sato, "Attentional modulation of self-motion perception," *Perception*, vol. 32, no. 4, pp. 475–484, Apr. 2003, doi: 10.1068/p5037.

[54] S. Palmisano, B. J. Gillam, and S. G. Blackburn, "Global-perspective jitter improves vection in central vision," *Perception*, vol. 29, no. 1, pp. 57–67, Jan. 2000, doi: 10.1068/p2990.

[55] S. Palmisano, D. Burke, and R. S. Allison, "Coherent perspective jitter induces visual illusions of self-motion," *Perception*, vol. 32, no. 1, pp. 97–110, Jan. 2003, doi: 10.1068/p3468.

[56] B. E. Riecke, J. Schulte-Pelkum, F. Caniard, and H. H. Bulthoff, "Towards lean and elegant self-motion simulation in virtual reality," in *Proc. IEEE VR*, Mar. 2005, pp. 131–138, doi: 10.1109/VR.2005.1492765.

[57] E. Kokkinara, K. Kilteni, K. J. Blom, and M. Slater, "First person perspective of seated participants over a walking virtual body leads to illusory agency over the walking," *Sci. Rep.*, vol. 6, no. 1, p. 28879, Jul. 2016, doi: 10.1038/srep28879.

[58] S. K. Nayar, "Catadioptric omnidirectional camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 482–488, doi: 10.1109/CVPR.1997.609369.
[59] S. K. Nayar, "Omnidirectional vision," in *Robotics Research*, Y. Shirai and S. Hirose, Eds. London, U.K.: Springer, 1998.
[60] M. B. Campos, A. M. G. Tommaselli, J. M. Junior, and E. Honkavaara, "Geometric model and assessment of a dual-fisheye imaging system," *Photogramm. Rec.*, vol. 33, no. 162, pp. 243–263, Jun. 2018, doi: 10.1111/phor.12240.
[61] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015, doi: 10.1109/TRO.2015.2463671.
[62] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Jun. 2017, doi: 10.1109/TRO.2017.270510.
[63] S. Sumikura, M. Shibuya, and K. Sakurada, "OpenVSLAM: A versatile visual SLAM framework," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2292–2295.
[64] Y. Visell, B. L. Giordano, G. Millet, and J. R. Cooperstock, "Vibration influences haptic perception of surface compliance during walking," *PLoS ONE*, vol. 6, no. 3, Mar. 2011, Art. no. e17697, doi: 10.1371/journal.pone.0017697.
[65] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, Jan. 2002, doi: 10.1038/415429a.
[66] B. De Gelder and P. Bertelson, "Multisensory integration, perception and ecological validity," *Trends Cognit. Sci.*, vol. 7, no. 10, pp. 460–467, Oct. 2003, doi: 10.1016/j.tics.2003.08.014.
[67] H. B. Helbig and M. O. Ernst, "Optimal integration of shape information from vision and touch," *Exp. Brain Res.*, vol. 179, no. 4, pp. 595–606, Jun. 2007, doi: 10.1007/s00221-006-0814-y.
[68] H. B. Helbig, M. O. Ernst, E. Ricciardi, P. Pietrini, A. Thielscher, K. M. Mayer, J. Schultz, and U. Noppeney, "The neural mechanisms of reliability weighted integration of shape information from vision and touch," *NeuroImage*, vol. 60, no. 2, pp. 1063–1072, Apr. 2012, doi: 10.1016/j.neuroimage.2011.09.072.
[69] F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner, "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, 2007, doi: 10.3758/bf03193146.
[70] F. Faul, E. Erdfelder, A. Buchner, and A. G. Lang, "Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses," *Behav. Res. Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.
[71] B. L. Giordano, S. McAdams, Y. Visell, J. Cooperstock, H.-Y. Yao, and Vincent Hayward, "Non-visual identification of walking grounds," *J. Acoust. Soc. Amer.*, vol. 123, no. 5, p. 3412, 2008, doi: 10.1121/1.2934136.
[72] B. L. Giordano, Y. Visell, H.-Y. Yao, V. Hayward, J. R. Cooperstock, and S. McAdams, "Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions," *J. Acoust. Soc. Amer.*, vol. 131, no. 5, pp. 4002–4012, May 2012, doi: 10.1121/1.3699205.
[73] (Aug. 30, 2021). *Preston Wright. 360 Snow Walk*. [Online]. Available: https://youtu.be/7od8MpjOmP0
[74] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2011, pp. 143–146.
[75] S. Geisser and S. W. Greenhouse, "An extension of box's results on the use of the F distribution in multivariate analysis," *Ann. Math. Statist.*, vol. 29, no. 3, pp. 885–891, 1958, doi: 10.1214/aoms/1177706545.
[76] M. G. Kenward and J. H. Roger, "Small sample inference for fixed effects from restricted maximum likelihood," *Biometrics*, vol. 53, no. 3, pp. 983–997, 1997, doi: 10.2307/2533558.
[77] J. Nakamura, Y. Matsuda, T. Amemiya, Y. Ikei, and M. Kitazaki, "Virtual walking generator from omnidirectional video with ground-dependent foot vibrations," in *Proc. IEEE Virtual Reality*, Mar. 2021, pp. 607–608, doi: 10.1109/VRW52623.2021.00186.

JUNYA NAKAMURA received the B.S. degree in computer science and engineering from the Toyohashi University of Technology, Aichi, Japan, in 2020, where he is currently pursuing the M.S. degree in computer science and engineering.

YUSUKE MATSUDA received the M.S. and Ph.D. degrees in interdisciplinary science and engineering from the Tokyo Institute of Technology, in 2011 and 2014, respectively.

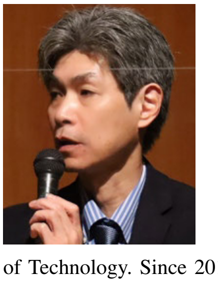
From January 2015 to October 2015, he was a Research Assistant at the Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology. From 2015 to 2019, he was a Research Assistant at the Tokyo University of Marine Science and Technology. From 2019 to 2020, he was a Research Assistant at the Department of Computer Science and Engineering, Toyohashi University of Technology. Since 2020, he has been an Assistant Professor at the Department of Computer Science and Engineering, Toyohashi University of Technology.

TOMOHIRO AMEMIYA (Member, IEEE) received the B.S. and M.S. degrees in mechano-informatics from The University of Tokyo, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in bioinformatics from Osaka University, Osaka, Japan, in 2008.

From 2004 to 2019, he was a Researcher at the NTT Corporation Communication Science Laboratories, Atsugi, Japan. Since 2020, he has been an Associate Professor at the Graduate School of Information Science and Technology, The University of Tokyo.

YASUSHI IKEI received the Ph.D. degree in mechanical engineering from The University of Tokyo, Tokyo, Japan, in 1988.

From 1993 to 1994, he was a Lecturer at Tokyo Metropolitan University, Tokyo. From 1997 to 2009, he was an Associate Professor at Tokyo Metropolitan University. From 2010 to 2019, he was a Professor at Tokyo Metropolitan University. Since 2020, he has been a Project Professor with the Graduate School of Information Science and Technology, The University of Tokyo.



MICHITERU KITAZAKI received B.A., M.A., and Ph.D. degrees in experimental psychology from The University of Tokyo, Tokyo, Japan, in 1992, 1994, and 1997, respectively.

From 1997 to 2000, he was a Research Associate at the Department of Psychology, The University of Tokyo. From 2000 to 2003, he was a Lecturer at the Toyohashi University of Technology, Aichi, Japan. From 2003 to 2016, he was an Associate Professor at the Toyohashi University of Technology. Since 2017, he has been a Professor at the Department of Computer Science and Engineering, Toyohashi University of Technology.