2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 978-1-6654-0158-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISMAR52148.2021.00050

Understanding the Two-Step Nonvisual Omnidirectional Guidance for Target Acquisition in 3D Spaces

SeungA Chung, Kyungyeon Lee, Uran Oh*
Department of Computer Science and Engineering
Ewha Womans University, Seoul, South Korea
{ewhacsa, ruddus716}@ewhain.net, uran.oh@ewha.ac.kr

Abstract

Providing directional guidance is important especially for exploring unfamiliar environments. However, most studies are limited to two-dimensional guidance when many interactions happen in 3D spaces. Moreover, visual feedback that is often used to communicate the 3D position of a particular object may not be available in situations when the target is occluded by other objects or located outside of one's field of view, or due to visual overload or light conditions. Inspired by a prior finding that showed users' tendency of scanning a 3D space in one direction at a time, we propose two-step nonvisual omnidirectional guidance feedback designs varying the searching order where the guidance for the vertical location of the target (the altitude) is offered to the users first, followed by the horizontal direction of the target (the azimuth angle) and visa versa. To investigate its effect, we conducted the user study with 12 blind-folded sighted participants. Findings suggest that our proposed two-step guidance outperforms the default condition with no order in terms of task completion time and travel distance, particularly when the guidance in the horizontal direction is presented first. We plan to extend this work to assist with finding a target in 3D spaces in a real-world environment.

Index Terms: Human-centered computing—Empirical studies in interaction design—;——Human-centered computing—User interface design—;

1 Introduction

Understanding 3D positions of surrounding objects is important for users to decide whether to interact with or to avoid them (*e.g.*, grabbing a specific product displayed high or low on a shelf at a grocery store, finding a light switch or a door knob). However, visually locating particular objects can be inefficient and frustrating in a 3D space [29]. Moreover, this approach may not be feasible in some situations such as object being occluded by other objects or outside of one's field of view, or the environment is too dark, or for some users who have visual impairments.

For this reason, many researchers investigated how to design directional guidance with nonvisual feedback [14, 19, 20, 27, 32–34]. For example, audio feedback was explored to convey the touchscreen information such as coordinate information or direction on the screen to people with visual impairments [27, 28, 34]. Moreover, a portable navigation system called GPSTune proposed to help users to find their route through audio feedback. It conveys the remaining distance and walking direction towards the target destination with various levels of volume and panning sound [33]. On the other hand, vibration feedback has been studied as well [17, 17, 19, 20]. For instance, Hong *et al.* [19] designed a wrist-based wearable haptic device for people with visual impairments. It provides vibrations

*Corresponding author.

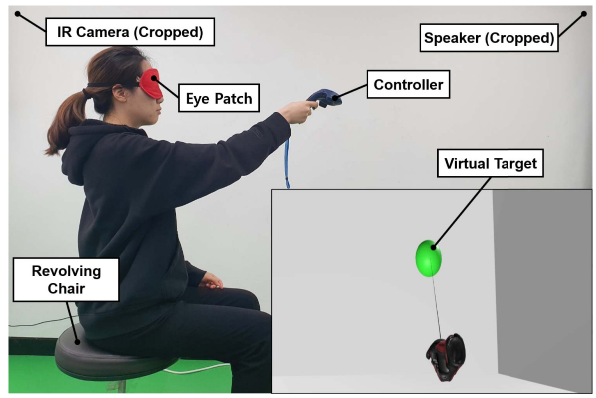


Figure 1: The experiment setting for the virtual target pointing task for the user study, and a screenshot of the virtual environment at the bottom right of the figure, which was not visible to the participants.

to users for guiding their hand to find the target position on a two-dimensional surface such as printed map. In addition, Katzschmann *et al.* [20] introduced a smart white cane system which can be worn as a belt. It conveys obstacles' direction and distance through vibrating motors. TactileGlove [17] also can convey up to eight different walking directions in addition to up and down directions.

Although many of the studies have contributed to spatial navigation such as way-finding or reaching for a specific location on a route or on a flat surface, the guidance is limited to the 2-dimensional space. Meanwhile, little has been studied on how to convey omnidirectional information in 3-dimensional space with nonvisual feedback except for Chung *et al.* [9] where they found users' tendency of scanning the 3D environment in a horizontal direction first and then in a vertical direction for finding a target regardless of nonvisual feedback conditions.

Inspired by this finding, we proposed a two-step omnidirectional guidance where each step provides users with feedback for one particular direction at a time either the horizontal or the vertical direction. Then we conducted a user study with 12 blind-folded sighted participants to identify the effectiveness of the proposed design. As shown in Fig. 1, participants were asked to perform a target pointing task in a virtual environment with the following feedback designs: (1) *VF* condition where the vertical direction is provided first before the horizontal direction, (2) *HF* condition where the horizontal direction is conveyed first before the vertical direction, and (3) *NO* condition where both directions are provided at once. As a result, giving directional guidance in two steps showed positive effects on task performance in terms of task completion time and travel distance, especially when the horizontal direction is provided first to participants. In addition, the two-step guidance was preferred in terms of understanding, fatigue, and satisfaction. However, there was no significant difference on preference between

978-1-6654-0158-6/21/$31.00 ©2021 IEEE
DOI 10.1109/ISMAR52148.2021.00050

the two two-step guidance conditions (*i.e.*, *VF* and *HF*). Based on the findings, we discuss how future nonvisual omnidirectional guidance can be designed for locating specific objects in both virtual and physical environments.

The contributions of this research are as follows: (1) the proposal of the two-step nonvisual omnidirectional guidance for a 3D environment, (2) the empirical evaluation of the effectiveness of our proposed feedback type, (3) the implications for designing a better nonvisual directional feedback in 3D spaces.

2 Related Work

Our work is inspired by prior work on providing nonvisual directional guidance with audio and haptic feedback for people with visual impairments.

2.1 Directional Guidance with Auditory Feedback

As for auditory feedback, some researches focused on supporting navigation aid on two-dimensional (2D) surface, especially for touchscreen-based interactions [24,27,28,34]. Oh *et al.* [27,28], for instance, explored various types of sound parameters such as pitch, volume, and timbre for mapping different attributes of touchscreen gestures so that people with visual impairments can learn touchscreen gesture with auditory feedback. They recommended to use pitch and stereo panning for representing the movements on the x coordinates and y coordinates respectively based on their findings. Likewise, Su *et al.* [34] presented Timbremap which is a sonification interface that is designed to assist people with visual impairments for exploring the indoor layout such as floor plan on a touchscreen device. It also used pitch to indicate the upward or downward directions (*e.g.*, high pitch for upward) when conveying geometry information. In addition, Leplâtre and Brewster [24] investigated the effectiveness of providing non-speech audio feedback to the navigating mobile user interface with a touchscreen which had sophisticated menu structure with 150 different sounds.

Others used audio feedback for wayfinding [33,39,41,43]. For example, Strachan *et al.* [33] introduced a portable navigation system called GPSTune, which assists users with wayfinding with audio feedback which conveys information related to the target destination such as distance and direction. The remaining distance to the destination and walking direction are mapped to different levels of volume and panning sound, respectively. Furthermore, Zhao *et al.* [43] explored the design of wayfinding guidance for people with low vision using smartglasses. They conveyed turn-by-turn verbal instructions with spatialized audio cues for informing the direction of next turning point or the destination. While these work demonstrated how their audio feedback design can help people with low vision to find a way to their target destination, their directional guidance is limited to 2D surfaces such as a touchscreen or a floor.

Meanwhile, 3D directional guidance with nonvisual feedback is studied to assist people with visual impairments to aim the camera [1,38] or to convey spatial information [2]. For instance, Ahmetovic *et al.* [1] presented a mobile app called ReCog that supports people with visual impairments to recognize the personal object by training an object recognizer with photos they have taken. The authors found that providing camera-aiming guidance using audio feedback improves the recognition accuracy. In our study, we investigate how to better design the nonvisual feedback for providing 3D directional guidance for people with visual impairments.

2.2 Directional Guidance with Haptic Feedback

Haptic feedback is also studied to provide 2D guidance for people with visual impairments [17, 19, 21, 32]. Stearns *et al.*, for instance, [32] implemented an optical character recognition system which assists people with visual impairments to read printed text line by line using a finger-mounted camera through haptic feedback in addition to auditory feedback. Similarly, Hong *et al.* [19] proposed a wearable haptic device on the wrist which supports path tracing on a 2D surface such as touchscreen or paper by guiding the hands. Meanwhile, haptic feedback was investigated for supporting navigation tasks as well [13–15,20,25,30,35]. For example, Ertan *et al.* [14] proposed a wearable navigation system which is composed of 4-by-4 array of micromotors to convey four cardinal directions and a stop signal with haptic feedback. Similarly, Van Erp *et al.* [13] designed a wearable vibrotactile belt, which informs the direction and the distance to the next waypoint varying vibration location and rhythm, to assist people with wayfinding through nonvisual haptic feedback. In addition, Sun *et al.* [35] presented a handheld navigation system using vibration feedback and explored 11 different vibration feedback modes. The authors demonstrated that the system can reduce the workload of visual and auditory channels, and confirmed that the recognition accuracy of vibration feedback can be affected by the navigation speed. Moreover, Rümelin *et al.* [30] introduced NaviRadar, a mobile application designed using a radar metaphor, which provides the route guidance to users through various vibration modes varying the intensity, duration, rhythm, and roughness. Furthermore, Katzschmann *et al.* [20] proposed a smart white cane which delivers where obstacles are located in terms of two directions (*i.e.*, horizontal and vertical) as well as the distance of the obstacle and a user. While 3D directional guidance was studied as well with vibrotactile feedback [11,17,21], these require a custom hardware with vibrations motors. For example, de Jesus Oliveira *et al*. [11] proposed vibrotactile HMD with seven tactors which informs the azimuth direction of target and also conveys the target height by varying the vibration frequency.

In this study, we aim to investigate various feedback designs to support two-step nonvisual omnidirectional guidance in 3D spaces using both audio and haptic feedback than can be conveyed without specialized hardware devices.

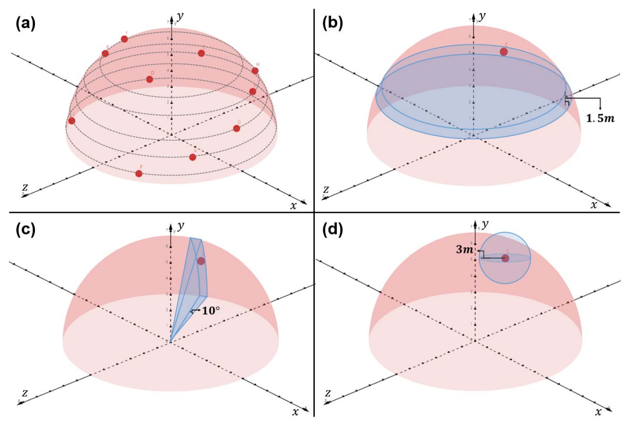


Figure 2: (a) The red dots on the surface of the hemisphere represent the position of 10 targets which have the same distance from participants' seated locations (the origin). The blue regions indicate (b) the target height for *VF* condition, (c) the target direction for *HF* condition and (d) the near-target region for *NO* condition where the handheld controller vibrates.

3 Experiments

To explore the effects of different nonvisual omnidirectional guidance on the performance of the target pointing task, we conducted a single-session user study with 12 blind-folded sighted participants. They were asked to point virtual targets using a handheld controller as shown in Fig. 1 where the targets appeared at one of the 10 positions in random order; see Fig. 2(a).

3.1 Conditions

In a previous study, it is found that there is a tendency of scanning the 3D environment in horizontal direction (the azimuth angle) first followed by the virtual direction (the altitude; height) when finding a target in a 3D space [9]. Inspired by this, we designed a two-step nonvisual omnidirectional guidance so that horizontal or vertical direction is provided one at a time based on the shortest Cartesian distance from the current position to the target location instead of guiding users based on the Euclidean distance, which is shorter. Then we investigated three different feedback conditions for conveying omnidirectional guidance to users: *Vertical First*, *Horizontal First*, and *No Order*:

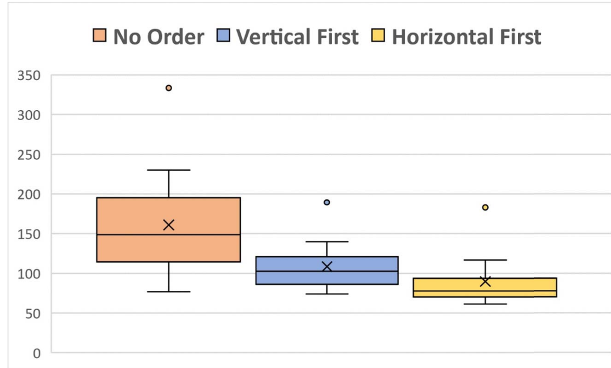
- **Vertical First (VF)**: It provides directional guidance to users in two steps: vertical direction first followed by horizontal direction. To be specific, it plays a beeping sound repeatedly to guide users towards the target height (*i.e.*, altitude), the vertical position of the target (y-coordinate) as shown in the disk-shaped blue region in Fig. 2(b); the height of the disk is set to the diameter of the target (*i.e.* 1.5 m). Then when the region is reached, the same beeping sound is played again to guide a user to the exact 3D position of the target. In both steps, the interval of the beeping sound decreases as the laser raycast from the controller gets closer to the target height in the first step or the exact target position in the second step. Meanwhile, the handheld controller vibrates as the user stays within the region.
- **Horizontal First (HF)**: Similar to *VF* condition, the guidance is provided to users in two steps but in the reversed order; horizontal direction first followed by vertical direction. A beeping sound is played to guide a user to the target direction (*i.e.*, azimuth angle), the horizontal position of the target, which is the blue pie-like region shown in Fig. 2(c). The angle of the region is set to 10° in xz-plane which is close to the angle of two tangent lines of the target sphere from the position of a user. When a user reaches the region, the same beeping sound is then played to guide a user to the exact 3D position of the target. The beeping interval and the vibration feedback from the controller are the same except that the feedback is based on the target's direction instead of its height.
- **No Order (NO)**: Unlike the first two conditions, there is no certain order when finding the target objects. The beeping interval decreases as the pointing direction gets closer to the exact 3D position of the target object (x-, y-, z-coordinates). Similarly, the handheld controller vibrates while the laser raycast from the controller is located within the 3 m from the center of the target as shown in Fig. 2(d). Note that the region was set to indicate the target position in various directions, not only vertical or horizontal directions.

3.2 Participants

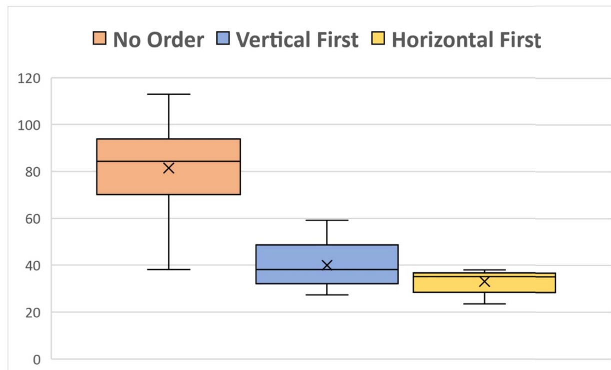
We recruited 12 participants (all female) for the study, and their average age was 23.5 (SD = 1.78; range 21-27). Based on their self-reported information, all of them were right-handed and had a virtual reality experience. Also, none of them had visual or auditory difficulties.

3.3 Apparatus

For the experiment, we built software with Unity (version: 2019.2.17f1). It was run on a desktop computer with an AMD Ryzen 7 1700 CPU, 16GB of RAM, and an RTX2080 graphics card. As for the hardware for simulating different target locations in a virtual environment, we used a handheld controller for HTC VIVE Pro Eye to track participants' pointing direction while they are holding it and to convey vibration feedback to the participants during the experiment. As for audio feedback, we had an external speaker. As for the target finding task, all 10 targets have the same shape and size (*i.e.*, a sphere with a radius of 0.75 m), and the distance



(a) Task Completion Time (s)



(a) Travel Distance (m)

Figure 3: Task performance results for pointing 10 targets in a row per condition: (a) the box plot of task completion time in seconds, and (b) the box plot of travel distance in meters. Each X represents the average value.

between the position of the participant and each target was set to 6 m. The position of the targets were set varying the degree in horizontal direction (*i.e.*, every 30° from 30° to 330° in xz-plane except for 180°) and height (*i.e.*, five different heights from 0.5 m to 4.5 m in y-coordinate with the interval of 1.0) as shown in Fig. 2(a); note that none of the successive targets appear at the same height while the order of the target position was randomly set.

We played beeping sound for audio feedback where its interval decreases as the pointer gets closer to the target region or the target itself. Also, we turned on the vibration motor for haptic feedback when the pointer is in the blue region to inform users that they found the target's vertical or horizontal position or the target is near. In addition, the pitch for beeping sound was set to reciprocal of the half of distance so that higher pitch sound is played. Note that we updated the interval and pitch of the beeping sound using Unity AudioMixer. Moreover, when the laser pointer enters the exact target position, participants receive short vibration feedback along with a chime sound for confirmation where the intensity is doubled compared to the vibration feedback that is provided to participants when they stay within the blue regions in Fig. 2(b-d). We saved event logs with timestamps for the analysis.

3.4 Procedure

The user study began by signing the consent form for the experiment, followed by collecting the personal information of participants such as age, prior experience with VR, and dominant hand. Then we briefly introduced the devices and tasks. For each condition, we demonstrated feedback such as beeping and chime sounds for participants to get familiar with it. Next, we asked them to sit on a revolving chair that can be rotated 360° at a fixed location. They were also asked to hold the controller with their dominant hand with their eyes closed during the experiment. The task was to point each target as quickly as possible with directional guidance feedback given for each condition. They were instructed to follow the directional guidance feedback given for each condition and point the target which appears in a random position in the virtual 3D space. The conditions were presented in a counterbalanced order using a balanced Latin square. Before the actual task with 10 targets, we offered a practice session to participants to get familiar with the feedback for each condition with five targets varied in terms of height and direction, where none of them shared the same 3D coordinates as any of the targets used in the actual task. Participants were allowed to take a five-minute break at the end of each condition. Subjective feedback was collected after the completion of all conditions by asking them to rate the level of understanding, fatigue, and satisfaction as well as their preference.

3.5 Data and Analysis

We collected the task completion time and trajectory of 360 trials (12 participants x 10 targets x 3 conditions). One-way ANOVA was used to assess the main effects of three guidance conditions in terms of task completion time and travel distance. Then pairwise comparisons were conducted for the post hoc analyses. Also, subjective feedback including fatigue and satisfaction for each condition were examined.

3.6 Findings

Here, we present the results of the experiment. Note that there was no ordering effect.

3.6.1 Task Completion Time

An one-way ANOVA revealed that the difference in task completion time between three conditions were statistically significant ($F_{(2)}$ = 7.18, p = .003). As shown in Fig. 3(a), the average time was 108.44 s for *VF* (*SD* = 31.5), 89.64 s for *HF* (*SD* = 32.9), and 160.96 s for *NO* (*SD* = 69.1). Post hoc pairwise comparisons showed that participants were significantly faster in *VF* and *HF* conditions compared to *NO* condition (p = .014 for *HF* vs. *NO*, and p = .026 for *VF* vs. *NO*), confirming that two-step guidance conditions (*i.e., VF* and *HF*) were more efficient than one-step guidance condition (*i.e., NO*). However, the time difference between *VF* and *HF* is not found to be statistically significant.

3.6.2 Travel Distance

We also analyzed the travel distance. As shown in Fig. 3(b), the differences between three conditions were found to be significant ($F_{(2)}$ = 39.33, $p < .001$). As shown in Fig. 3(b), the average distance was 39.93 m for *VF* (*SD* = 9.7), 33.05 m for *HF* (*SD* = 4.8), and 81.41 m for *NO* (*SD* = 22.6). Pairwise post hoc analyses showed that participants moved significantly shorter in *VF* and *HF* conditions compared to *NO* condition ($p < .001$ for both) although the ideal distance was guided with *NO* condition. Moreover, we found that the travel distance is shorter with *HF* than *VF* condition (p = .038).

3.6.3 The Effect of Searching Direction on Performance

We conducted a follow-up analysis examining the duration of each step of two-step guidance for each condition. For example, the duration of the *horizontal search* (the first step) in *HF* condition is from the start of the task to the first entrance of the blue region around the correct target direction, and that of the *vertical* search (the second step) is from the first entrance of blue region to the first entrance of the target. We conducted a two-way ANOVA with factors of *feedback condition* (*VF* vs. *HF*) and *search direction*

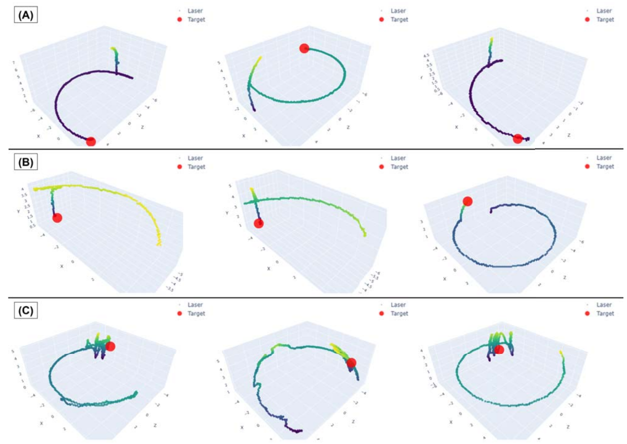


Figure 4: Trace examples during target pointing task in (a) *VF* condition for P1, P6, and P9 (from left to right), (b) *HF* condition for P1, P5, and P9 (from left to right), and (c) *NO* condition for P1, P3, P7 (from left to right). Red circles indicate the target positions.

(*horizontal search* vs. *vertical search*). The test revealed that there is an interaction effect between two factors ($F_{(1,1)}$ = 85.2, $p < .001$).

As for *horizontal search*, participants were significantly faster under *HF* condition (*M* = 50.69 s, *SD* = 20.91) compared to *VF* condition (*M* = 85.33 s, *SD* = 24.66); t = 13.77, p = .001. However, as for *vertical search*, the average duration of *VF* (*M* = 23.38 s, *SD* = 11.31) was shorter than that of *HF* (*M* = 38.66 s, *SD* = 13.05); t = 9.40, p = .006.

Likewise, the interaction effect also exists for travel distance ($F_{(1,1)}$ = 26.0, $p < .001$). As for horizontal search, participants found the target with shorter distance under *HF* condition (*M* = 22.31 m, *SD* = 11.96) compared to *VF* condition (*M* = 31.30 m, *SD* = 17.81); t = 21.0, $p < .001$. However, as for *vertical search*, the average travel distance of *VF* (*M* = 8.63 m, *SD* = 6.42) was shorter than that of *HF* (*M* = 10.74 m, *SD* = 8.07); t = 5.0, p = .026.

3.6.4 Trace Analysis

In addition to the main analysis, we analyzed participants' traces. Reflecting the results of the task performance particularly in terms of the travel distance, participants tend to make less number of sudden changes in directions in the two-step directional guidance conditions (*i.e.*, *VF* and *HF* as shown in Fig. 4(a)-(b)) than one-step condition, *NO*, as shown in Fig. 4(c). To be specific, participants were confused about which direction to move once they reach the target height in *VF* condition or at the beginning in *HF* condition where they received guidance for the horizontal direction. On the other hand, the number of sudden directional changes was more prevalent at the end when finding the exact target location in *NO* condition. In addition, the tendency of finding the target direction first before finding the exact position in 3D was confirmed in *NO* condition as found in [9].

3.6.5 Subjective Evaluation and Preference

Finally, we analyzed the participants' feedback including subjective evaluation and preference collected after the task completion. Participants were asked to evaluate each condition in a 5-point Likert scale in terms of *understanding*, *fatigue*, and *satisfaction*. The results are shown in Fig. 5.

- *Understanding: It was easy to understand the guidance.*
- *Fatigue: The task was physically tiring with the guidance.*
- *Satisfaction: I am satisfied with the guidance.*

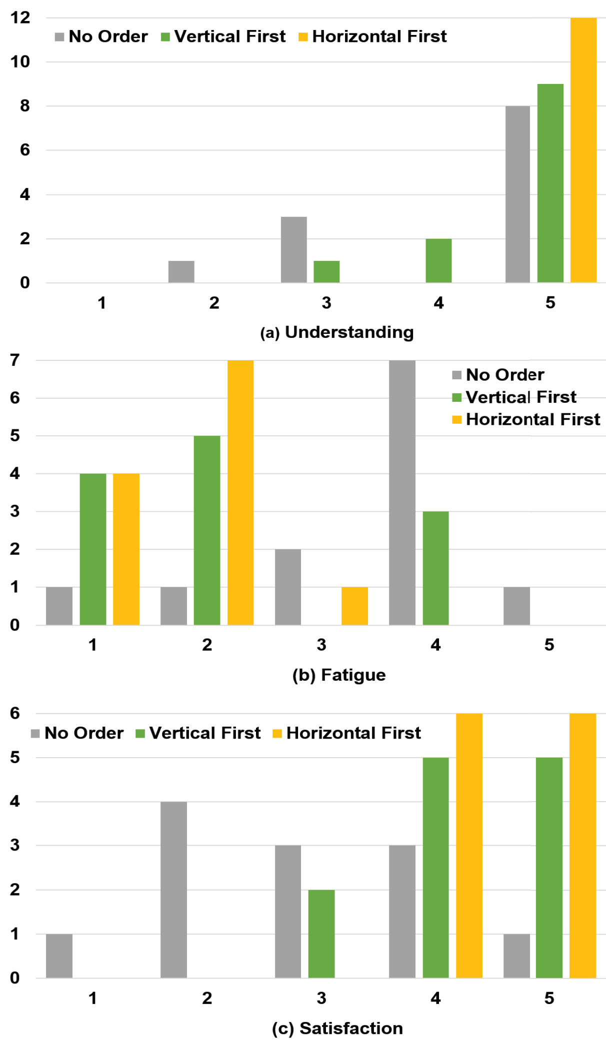


Figure 5: The histograms for the subjective evaluation of each condition in terms of three metrics: (a) understanding, (b) fatigue, and (c) satisfaction. X-axis represents the rating of 5-point Likert scale (*1-strongly disagree, 5-strongly agree*).

Most participants perceived that all three feedback conditions are easy to understand, particularly for *HF* condition, which received the ratings of five from all participants (M = 5.0; SD = 0.0). The average ratings were 4.67 for *VF* (SD = 0.65) and 4.25 for *NO* (SD = 1.14). Similarly, when asked about fatigues, again, *HF* condition received the best ratings with the average of (M = 1.75; SD = 0.62), followed by *VF* (M = 2.17; SD = 1.19), and *NO* (M = 3.50; SD = 1.09). The results are similar in terms of satisfaction. The average ratings were the highest for *HF* (M=4.50; SD = 0.52), followed by *VF* (M=4.25; SD = 0.75) and *NO* (M=2.92; SD = 1.16) condition.

Reflecting the results of subjective evaluation, *HF* conditions received the highest number of votes when asked for preference (N = 7), while *VF* received the remaining five votes. The most dominant reason for choosing *HF* as their favorite was that following the horizontal directional guidance first is less exhausting since they do not have to stay the arm in specific heights, unlike *VF* condition. Meanwhile, participants who chose *VF* condition commented that it was comfortable and efficient since they felt they moved their arms less to find the vertical position of the target.

As for *NO* condition, which did not receive any vote, one participant commented that it was confusing to interpret both the vertical and horizontal directions at the same time as it only informs whether it is relatively closer or farther to the target.

4 Discussions

Here we discuss study findings to address suggestions for designing nonvisual omnidimensional directional feedback in 3D spaces.

4.1 Two-Step over One-Step Guidance

The findings showed that our two-step guidance is significantly faster with a shorter travel distance than one-step guidance (*NO*). This is interesting since the guided path of *NO* condition is based on Euclidean distance, and thus shorter than the other two two-step guidance based on the shortest Cartesian distance. Moreover, even when one-step guidance is provided to participants, we observed the tendency of dividing the tasks into two-steps as in prior study [9]. This could be due to the low cognitive load since the participants could focus on one direction at a time, either horizontal or vertical, with two-step guidance. Although further investigation is needed in terms of the optimal feedback design for directional guidance in 3D spaces, we recommend giving two-step directional guidance in cardinal directions rather than diagonal direction.

4.2 Horizontal Guidance before Vertical Guidance

We assumed that the task completion time of HF condition would be shorter than that of VF condition, reflecting the tendency of searching for a target in the horizontal direction first [9]. However, we found no significant difference in terms of task completion time. Yet, the travel distance was shorter under HF condition than VF condition. In addition, HF received more positive feedback from the participants compared to other feedback conditions particularly in terms of fatigue and satisfaction. This could be due to the gorilla arm problem [18], as finding the horizontal direction of the target while lifting one's arm at the target height in *VF* is physically more difficult compared to finding the target's horizontal direction while holding the controller at one's comfortable height in *HF* condition. Thus, considering the experiment result of relatively shorter travel distance and positive subjective ratings of HF condition, guiding users towards a target in the horizontal direction first followed by vertical direction is recommended for informing the target's 3D position.

4.3 A Ballistic Step Followed by A Corrective Step

Our findings show that the participants were faster in the first step than the second step when receiving two-step guidance feedback regardless of the feedback condition. One possible explanation is the the extra cognitive load during the second step as participants had to focus on performing two subtasks with multiple output modalities (*i.e.*, pointing the controller towards the closer direction of the target given the proximity-based beeping sound, while staying in the blue region given vibration feedback). This can be also explained by ballistic and correction phases of goal-directed 3D movements observed in prior studies [3, 4, 26] where rough movements were made in the first step and fine adjustments were made in the last step. Although there was no intention when designing two-step guidance and a further investigation is needed, this implies that our two-step approach is suitable for supporting users' searching behavior.

4.4 The Limitation of Proximity-based Feedback

We have successfully refined the feedback design for nonvisual omnidirectional guidance for targets in 3D spaces by splitting the task into two steps. However, we noticed sudden changes in direction from the trace analysis, and this confusion was observed when participants are receiving horizontal directions in both HF and VF

conditions. While we did not use stereo sound reflecting the prior finding that conveying different types of information through the same channel can be distracting [31] and used different modalities for conveying proximity-based feedback instead, our findings suggest the importance of spatialized sound. The absence of spatialized sound requires users to move their hands to different directions in order to identify correction direction by listening to the changes in the beeping interval which is inefficient. The effect of spatialized sound on the performance of two-step omnidirectional guidance would be interesting to investigate.

4.5 Setting the Optimal Size for Vibrating Regions

When deciding the vibrating regions (blue regions in Fig. 2) for providing an additional cue, we tried to map the region based on the size (*i.e.*, the diameter) and the distance (*i.e.*, the angle between two tangent lines) of the target. However, we believe that the region can be optimized. To be specific, if the region is set too big, it region itself can be found at ease but the search space within the region is still too wide to find the accurate target position. As shown in Fig. 4(c), for instance, the traces for *NO* condition have zigzag motion near the target, which suggests that the size of the blue region for *NO* condition may have been set as too wide. On the other hand, if the range of the region is too small, users would not benefit from the additional cue since pointing the region can be difficult as much as pointing the target.

4.6 VR Applications for People With Visual Impairments

While a number of VR applications had been designed for sighted people, recent studies began to investigate VR for training people with visual impairments for daily activities (*e.g.*, walking) [22, 23, 36, 37, 41]. In addition, researchers have conducted several studies that track and observe the behavior of how people with visual impairments search for specific objects in virtual or augmented reality space as a simulated real-world [7, 10, 42, 44]. Our nonvisual feedback design can also be used in various 3D environments in virtual reality (*e.g.*, underwater swimming, piloting an aircraft, spacewalking) for people with visual impairments for training and entertainment in addition to orientation and mobility training. The system can be used to help people with visual impairments to find the direction of a specific target in any 3D space before initiating interactions.

4.7 Implementation for Real World Tasks

The main feedback modality we used to inform the remaining distance towards the target was audio. For this reason, if a user is at a place in a real world environment with loud noises, relying only on sound feedback alone could be challenging. Moreover, listening to the sound of the surrounding environment is important for one's safety. While a bone-conductive headset is an option, haptic feedback can be reconsidered as the main feedback modality for real world settings as in [16]. In addition, while we assumed that the target position is known and that users' pointing direction is detected with high precision by utilizing the capabilities of a virtual environment focusing on the feedback design, the existing technology has limited accuracy to be supported for real word use. Thus, we plan to investigate and improve existing state-of-art technologies for recognizing a specific object in a 3D environment along with its position including the depth [12, 40, 45], and tracking one's pointing finger or head orientation in an actual environment [8, 41], then apply our feedback design for real word tasks as a future work. As for the implementation of such system, a wearable camera that can generate 360° equirectangular panoramic images with a wide vertical field of view should be used to detect surrounding objects from floor-, to head-height as in [16] that are predefined or set on the fly via voice command.

5 Limitations

As a controlled lab study, our study has several limitations. First, the effectiveness of other feedback designs for providing nonvisual guidance to users including spatialize sound such as echolocation [2] was not investigated. Also, while we had the simplest proximity-based design for both audio and haptic feedback focusing on two-step guidance approaches, it can be optimized to maximize the effect. In addition, user-specific features such as arm length or were not considered, which might affect the results. Also, we used ray casting method and we did not examine the movements in depth; the positions of the participants and the target were fixed in our study. A future study should examine how relative depth information can be conveyed in addition to relative direction. Also, while the pitch of the audio feedback [5, 6] is known to affect the reaction time for target finding tasks in 3D, we did not consider pitch in our study. Moreover, although we were able to find statistical significance, the sample size is small (N=12). Lastly, although people with visual impairments would benefit the most from our nonvisual feedback than sighted people in special situations such as objects being occluded or under poor light condition, none of our participants in our study had visual impairments. Thus, a future study should be conducted with a larger number of participants including people with visual impairments to generalize the findings with stronger statistical results.

6 Conclusion and Future Work

We proposed two-step omnidirectional nonvisual guidance for finding a target in 3D spaces, and conducted a user study with 12 blindfolded sighted participants to assess its effectiveness compared to a default one-step guidance. The findings show that our two-step guidance outperforms the default condition, especially when the horizontal guidance is provided to participants first before the vertical guidance. Likewise, the subjective evaluation shows that our feedback design is easy to understand with less fatigue and that it is satisfying. Based on the findings, we plan to extend our work by applying and evaluating the design focusing on conveying the frequent changes in objects' direction with respect to users' location in both virtual and real world environments.

7 Acknowledgments

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2020-0-01460) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation).

References

[1] D. Ahmetovic, D. Sato, U. Oh, T. Ishihara, K. Kitani, and C. Asakawa. Recog: Supporting blind people in recognizing personal objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

[2] R. Andrade, J. Waycott, S. Baker, and F. Vetere. Echolocation as a means for people with visual impairment (pvi) to acquire spatial knowledge of virtual space. *ACM Trans. Access. Comput.*, 14(1), Mar. 2021. doi: 10.1145/3448273

[3] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 327–334, 2018. doi: 10.1109/VR.2018.8446317

[4] O. J. Ariza N., M. Lange, F. Steinicke, and G. Bruder. Vibrotactile assistance for user guidance towards selection targets in vr and the cognitive resources involved. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 95–98, 2017. doi: 10.1109/3DUI.2017.7893323

[5] A. U. Batmaz and W. Stuerzlinger. The effect of pitch in auditory error feedback for fitts' tasks in virtual reality training systems. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 85–94. IEEE, 2021.

[6] A. U. Batmaz and W. Stuerzlinger. Effects of different auditory feedback frequencies in virtual reality 3d pointing tasks. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 189–194. IEEE, 2021.

[7] C. R. Bennett, E. S. Bailin, T. K. Gottlieb, C. M. Bauer, P. J. Bex, and L. B. Merabet. Virtual reality based assessment of static object visual search in ocular compared to cerebral visual impairment. In *International Conference on Universal Access in Human-Computer Interaction*, pp. 28–38. Springer, 2018.

[8] R. Boldu, A. Dancu, D. J. Matthies, T. Buddhika, S. Siriwardhana, and S. Nanayakkara. Fingerreader2. 0: designing and evaluating a wearable finger-worn camera to assist people with visual impairments while shopping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–19, 2018.

[9] S. Chung, K. Lee, and U. Oh. Investigating three-dimensional directional guidance with nonvisual feedback for target pointing task. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 206–210. IEEE, 2020.

[10] S. Chung, S. Park, S. Park, K. Lee, and U. Oh. Improving mealtime experiences of people with visual impairments. In *Proceedings of the 18th International Web for All Conference*, W4A '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3430263.3452421

[11] V. A. de Jesus Oliveira, L. Brayda, L. Nedel, and A. Maciel. Designing a vibrotactile head-mounted display for spatial awareness in 3d spaces. *IEEE transactions on visualization and computer graphics*, 23(4):1409–1417, 2017.

[12] G. P. de La Garanderie, A. A. Abarghouei, and T. P. Breckon. Eliminating the blind spot: Adapting 3d object detection and monocular depth estimation to 360 panoramic imagery. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 789–807, 2018.

[13] J. B. V. Erp, H. A. V. Veen, C. Jansen, and T. Dobbins. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP)*, 2(2):106–117, 2005.

[14] S. Ertan, C. Lee, A. Willets, H. Tan, and A. Pentland. A wearable haptic navigation guidance system. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)*, pp. 164–165. IEEE, 1998.

[15] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 68–82, 2019.

[16] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In *The 21st International ACM SIGACCESS conference on computers and accessibility*, pp. 68–82, 2019.

[17] S. Günther, F. Müller, M. Funk, J. Kirchner, N. Dezfuli, and M. Mühlhäuser. Tactileglove: Assistive spatial guidance in 3d space through vibrotactile navigation. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, PETRA '18, p. 273–280. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3197768.3197785

[18] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1063–1072, 2014.

[19] J. Hong, A. Pradhan, J. E. Froehlich, and L. Findlater. Evaluating wrist-based haptic feedback for non-visual target finding and path tracing on a 2d surface. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 210–219, 2017.

[20] R. K. Katzschmann, B. Araki, and D. Rus. Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3):583–593, 2018.

[21] O. B. Kaul and M. Rohs. Haptichead: A spherical vibrotactile grid around the head for 3d guidance in virtual and augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3729–3740, 2017.

[22] J. Kim. Vivr: Presence of immersive interaction for visual impairment virtual reality. *IEEE Access*, 8:196151–196159, 2020.

[23] J. Kreimeier and T. G¨otzelmann. First steps towards walk-in-place locomotion and haptic feedback in virtual reality for visually impaired. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2019.

[24] G. Leplâtre and S. A. Brewster. Designing non-speech sounds to support navigation in mobile phone menus. 2000.

[25] M. Mirzaei, P. Kán, and H. Kaufmann. Multi-modal spatial object localization in virtual reality for deaf and hard-of-hearing people. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 588–596. IEEE, 2021.

[26] K. Nieuwenhuizen, L. Liu, R. van Liere, and J.-B. Martens. Insights from dividing 3d goal-directed movements into meaningful phases. *IEEE Computer Graphics and Applications*, 29(6):44–53, 2009. doi: 10.1109/MCG.2009.121

[27] U. Oh, S. Branham, L. Findlater, and S. K. Kane. Audio-based feedback techniques for teaching touchscreen gestures. *ACM Transactions on Accessible Computing (TACCESS)*, 7(3):1–29, 2015.

[28] U. Oh, S. K. Kane, and L. Findlater. Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–8, 2013.

[29] J. Palmer. Attention in visual search: Distinguishing four causes of a set-size effect. *Current directions in psychological science*, 4(4):118–123, 1995.

[30] S. Rümelin, E. Rukzio, and R. Hardy. Naviradar: a novel tactile information display for pedestrian navigation. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 293–302, 2011.

[31] R. Sigrist, G. Rauter, R. Riener, and P. Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20(1):21–53, 2013.

[32] L. Stearns, R. Du, U. Oh, C. Jou, L. Findlater, D. A. Ross, and J. E. Froehlich. Evaluating haptic and auditory directional guidance to assist blind people in reading printed text using finger-mounted cameras. *ACM Transactions on Accessible Computing (TACCESS)*, 9(1):1–38, 2016.

[33] S. Strachan, P. Eslambolchilar, R. Murray-Smith, S. Hughes, and S. O'Modhrain. Gpstunes: controlling navigation via audio feedback. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pp. 275–278, 2005.

[34] J. Su, A. Rosenzweig, A. Goel, E. de Lara, and K. N. Truong. Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pp. 17–26, 2010.

[35] M. Sun, W. Gu, L. Wang, L. Dong, and Q. Qian. Investigating the effects of vibrotactile feedback on human performance in navigation tasks. *Computers & Electrical Engineering*, 67:608–619, 2018.

[36] L. Thevin, C. Briant, and A. M. Brock. X-road: virtual reality glasses for orientation and mobility training of people with visual impairments. *ACM Transactions on Accessible Computing (TACCESS)*, 13(2):1–47, 2020.

[37] M. Torres-Gil, O. Casanova-Gonzalez, and J. L. González-Mora. Applications of virtual reality for visually impaired people. *WSEAS transactions on computers*, 9(2):184–193, 2010.

[38] M. Vázquez and A. Steinfeld. An assisted photography framework to help visually impaired users properly aim a camera. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(5):1–29, 2014.

[39] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert. Swan: System for wearable audio navigation. In *2007 11th IEEE international symposium on wearable computers*, pp. 91–98. IEEE, 2007.

[40] M. Yavartanoo, E. Y. Kim, and K. M. Lee. Spnet: Deep 3d object classification and retrieval using stereographic projection. In *Asian Conference on Computer Vision*, pp. 691–706. Springer, 2018.

[41] Y. Zhao, C. L. Bennett, H. Benko, E. Cutrell, C. Holz, M. R. Morris, and M. Sinclair. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018.

[42] Y. Zhao, E. Cutrell, C. Holz, M. R. Morris, E. Ofek, and A. D. Wilson. Seeingvr: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[43] Y. Zhao, E. Kupferstein, H. Rojnirun, L. Findlater, and S. Azenkot. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.

[44] Y. Zhao, S. Szpiro, J. Knighten, and S. Azenkot. Cuesee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 73–84, 2016.

[45] S. Zhi, Y. Liu, X. Li, and Y. Guo. Toward real-time 3d object recognition: A lightweight volumetric cnn framework using multitask learning. *Computers & Graphics*, 71:199–207, 2018.