From Coarse Search to Precise Alignment: Auditory Guidance for Non-Visual Target Acquisition in VR

Kengo Tanaka
University of Tsukuba
Tsukuba, Japan
Miraikan – The National Museum of Emerging Science and Innovation
Tokyo, Japan
kengo.tanaka@digitalnature.slis.tsukuba.ac.jp

Renato Alexandre Ribeiro
Miraikan – The National Museum of Emerging Science and Innovation
Tokyo, Japan
renato.ribeiro@jst.go.jp

João Guerreiro
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisboa, Portugal
jpguerreiro@fc.ul.pt

Hironobu Takagi
Miraikan – The National Museum of Emerging Science and Innovation
Tokyo, Japan
hironobu.takagi@jst.go.jp

Yoichi Ochiai
University of Tsukuba
Tsukuba, Japan
wizard@slis.tsukuba.ac.jp

Chieko Asakawa
Graduate School of Science and Technology, Keio University
Kanagawa, Japan
casakawa@keio.jp

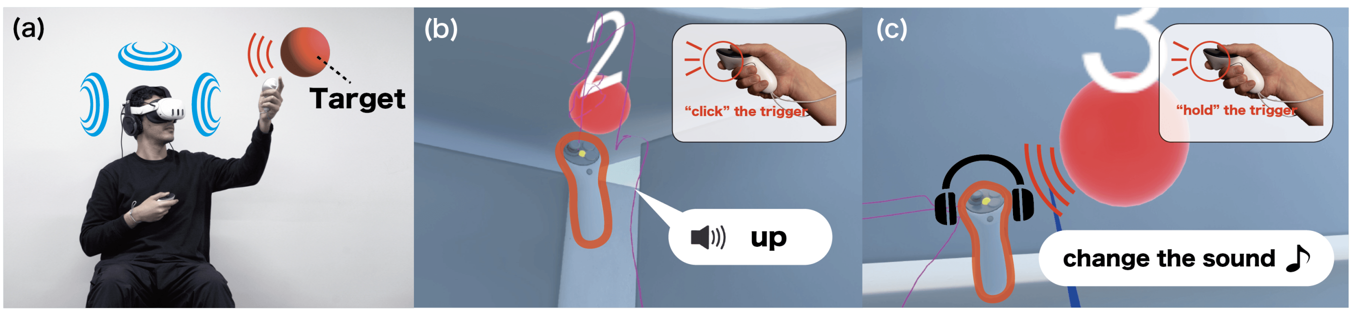

Figure 1: Overview of the target acquisition task and the two supplementary auditory support methods. (a) Task overview: a participant wearing an HMD and headphones reaches toward an arm-reachable target in near 3D space using auditory cues. (b) VERBAL support: when the participant clicks the trigger, the system provides spoken hand-relative directional guidance, such as "up," to support local correction. (c) HAND-REFERENCED support: when the participant holds the trigger, the auditory reference frame shifts to the current controller position, and changes in sound provide hand-referenced cues for adjusting the hand toward the target.

Abstract

Non-visual target acquisition in arm-reachable near 3D space requires both coarse directional search and fine-grained correction, yet it remains unclear how auditory guidance should adapt as precision demands increase. We investigate this question in virtual reality through a comparative study of three auditory guidance techniques: continuous head-referenced spatial audio, on-demand VERBAL guidance that provides spoken hand-relative directions, and on-demand HAND-REFERENCED audio that shifts the auditory reference frame to the hand and uses pitch to indicate vertical offset. Eleven blind participants performed target-reaching tasks across three difficulty levels defined by target size. VERBAL yielded the most robust overall performance, producing lower error rates, faster completion times, lower workload, and the strongest user preference. HAND-REFERENCED helped participants monitor height and lateral alignment, but did not fully resolve depth-related difficulty. Our findings show that non-visual near-space target acquisition is a multi-stage process involving coarse localization and final acquisition, with head-referenced spatial audio supporting initial orientation and on-demand cues supporting local correction near the target.

CCS Concepts

• **Human-centered computing** → **Accessibility systems and tools**; **Accessibility design and evaluation methods**; **Accessibility technologies**.

Keywords

visual impairment, spatial interaction, virtual reality

ACM Reference Format:
Kengo Tanaka, Renato Alexandre Ribeiro, João Guerreiro, Hironobu Takagi, Yoichi Ochiai, and Chieko Asakawa. 2026. From Coarse Search to Precise

ASSETS '26, Vila Nova de Gaia, Portugal

ACM ISBN 979-8-4007-2521-0/2026/10
https://doi.org/10.1145/3797867.3828980

Alignment: Auditory Guidance for Non-Visual Target Acquisition in VR. In *The 28th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '26), October 25–28, 2026, Vila Nova de Gaia, Portugal.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3797867.3828980

1 INTRODUCTION

Locating and accurately reaching nearby targets is a common everyday task. From pressing a button to grabbing an object, or reaching toward items on a desk, these actions rely on the ability to identify a target's position relative to the body and guide the hand toward it with sufficient precision. Similar requirements also arise in activities such as guided movement (e.g., to learn a specific pose in Yoga), physical training, or sports (e.g., positioning the hand on a hold in bouldering), where users must reach specific locations without continuous visual confirmation. This process becomes more demanding when spatial cues are limited or ambiguous. For blind and low-vision (BLV) people, in particular, the absence of visual cues can make it difficult to infer direction, depth, and alignment, requiring continuous adjustment during movement.

Prior work has explored a range of approaches to support target locating and reaching, often relying on audio due to its availability in commodity devices [18, 20, 44]. These approaches commonly use spatial audio to convey direction, sometimes combined with additional cues such as pitch, volume, or tempo to indicate proximity. More recently, work has explored continuous guidance techniques, such as hand steering [1], particularly to avoid dangerous obstacles. However, less is understood about approaches to address varying precision requirements, especially in near-space interactions, where both coarse search and fine-grained correction are needed.

Virtual Reality (VR) has been increasingly used to study target locating, reaching, and aiming in 3D spaces [5, 23, 30]. This is partly because these interactions are common in VR itself, where users select and manipulate elements without relying on traditional input devices [24]. At the same time, VR systems support controlled investigation of these tasks, acting as a proxy for direct physical interaction [21]. Head-Mounted Displays (HMDs) enable precise head and hand tracking while allowing systematic control over spatial layouts and feedback [9, 21]. This makes VR a suitable setting to examine how different forms of guidance support target acquisition under varying conditions. Prior work with sighted users shows that reaching performance varies with object location, viewing direction, and precision demands [26], raising open questions about how these factors affect non-visual target acquisition.

Despite prior research on non-visual target acquisition in VR [20, 23], these studies consider targets with fixed size and distance (e.g., positioned on the same plane). While useful for controlled comparisons, these setups provide limited insight into how different approaches support varying interaction demands. In practice, near-space targets differ not only in direction but also in depth, and tasks vary in required precision. High-precision guidance, such as continuous hand steering to avoid obstacles [1], may be necessary in some scenarios, but can hinder faster interactions when such accuracy is not required. As a result, it remains unclear how factors such as target depth and precision requirements influence the effectiveness of different auditory cues, and how guidance should adapt across stages of interaction, from coarse search to fine-grained correction near the target.

In this work, we investigate target locating and reaching in arm-reachable near 3D space using VR, focusing on how auditory cues support different levels of precision. Rather than assuming a single level of accuracy, we systematically vary the required precision to examine when coarse cues are sufficient and when additional guidance becomes necessary. Lower precision requirements may only demand approximate directional awareness, while higher precision introduces challenges such as front–back and up–down ambiguity, uncertainty in interpreting body-relative space, and the need for local correction near the target.

To study this, we conducted a comparative user study with 11 blind participants, under different precision requirements, using three auditory cue designs: (1) BASELINE, which provided continuous head-referenced spatial audio; (2) VERBAL, an on-demand method providing spoken hand-relative directional cues; and (3) HAND-REFERENCED, an on-demand method providing HAND-REFERENCED auditory feedback.

Our findings suggest that no single cue was equally effective throughout target acquisition. Spatial audio mainly supported coarse search, VERBAL most strongly supported final correction, and HAND-REFERENCED served as a useful but more interpretively demanding aid for local alignment.

Overall, this paper makes the following contributions:

(1) An empirical study of non-visual target locating and reaching in arm-reachable 3D spaces, investigating how layered auditory cues support different precision requirements.
(2) Design implications for auditory guidance in near-space interaction, including when to introduce supplementary cues and how to combine head- and HAND-REFERENCED information.

2 RELATED WORK

This section reviews prior work on non-visual directional guidance and aiming, followed by hand guidance work for locating, reaching, and steering in nearby space.

2.1 Audio Feedback for Non-Visual Directional Guidance and Aiming

Audio feedback has been widely used to convey spatial information in non-visual contexts, particularly for directional guidance. In navigation systems, speech, sonification, and spatialized sound are employed to communicate direction, distance, and environmental structures [3, 35, 39, 45]. Similar approaches have been applied to local directional tasks such as camera aiming, blind photography, and remote assistance, where users have to align a device (e.g., smartphone) or reticle towards a target using audio feedback [4, 8, 33, 42]. Across these tasks, audio primarily supports orientation rather than positioning the hand itself in space.

Studies in 3D pointing further show that spatial audio is effective for directional search, but insufficient for precise alignment. Prior work demonstrates that while stereo sound improves search efficiency, additional cues such as proximity-based feedback are required to achieve higher accuracy [14, 15]. Likewise, in immersive experiences, prior work exploring aiming tasks [30] demonstrated that proximity cues and reticle-centered feedback outperform target-centered spatial audio alone.

Together, prior work showcases that spatial audio supports coarse directional guidance, but requires supplementary cues for fine-grained alignment.

2.2 Non-Visual Hand Guidance for Reaching, Locating, and Steering

Prior work has explored non-visual guidance for hand-based tasks in structured settings, including interactions with shelves [25], domestic environments [12], and outdoors in parcel lockers [17]. These approaches demonstrate that multimodal feedback, including speech, haptics, and audio, can effectively support users in locating and acquiring nearby targets. However, as these systems rely on structured settings, they often assume stable layouts and relatively fixed interaction demands, limiting their applicability to free-space 3D locating tasks.

In accessible VR, research investigates how to represent the position of reachable targets using different modalities. Prior studies compare speech, sonification, and spatial representations, highlighting trade-offs between explicitness, interpretability, and efficiency [20, 23]. While these approaches support hand-to-target interaction, they typically assume a fixed level of precision or rely on simplified spatial layouts, providing limited insight into how guidance should adapt to varying precision requirements.

Audio guidance has also been applied to physical movement tasks such as climbing and limb exercises, where feedback is directly coupled to body motion [40, 46]. While constrained by surfaces, grids, or predefined motion paths - reducing the need for free-space spatial search and fine-grained correction, these tasks demonstrate that continuous auditory feedback can guide movement effectively. In contrast, free-space 3D hand locating introduces additional challenges. Users must coordinate left–right, up–down, and depth dimensions without physical constraints, and depth remains particularly difficult even with visual feedback [26]. Prior work suggests that representing both the target and the hand can improve performance. For example, dynamic audio feedback encoding both entities improves reaching accuracy [44], and audio-tactile proximity cues enhance spatial awareness and collision avoidance in 3D manipulation tasks [28]. These findings highlight the importance of conveying the relationship between the hand and the target, rather than only the target location.

Recent work further distinguishes between reaching and steering. Reaching involves target acquisition without trajectory constraints, while steering requires maintaining movement along a predefined path [1]. Although auditory techniques have been shown to support both tasks, steering assumes a constrained trajectory, whereas locating requires users to independently search for and approach a target in space.

In addition to differences in guidance tasks, prior research also differs in how guidance is delivered. Some systems provide on-demand guidance, allowing users to explicitly request assistance only when needed, supporting targeted correction while preserving autonomy and independence during exploration [34]. Others provide continuous guidance, offering ongoing auditory feedback throughout task execution, enabling users to adjust their movement toward a target [16]. These approaches reflect different design trade-offs between user control and continuous assistance. Although discussed in the context of navigation, prior work has highlighted that continuous auditory guidance should balance providing informative feedback while remaining unobtrusive, as excessive auditory information may interfere with environmental sounds relied upon by blind uses [2, 7].

Across these lines of work, most approaches provide a single form of guidance and assume relatively stable precision requirements. As a result, it remains unclear how auditory guidance should adapt across different interaction stages, such as coarse directional search, approach, and near-target correction, particularly in free, arm-reachable 3D space. Our work addresses this gap by investigating how auditory cues support target locating and reaching under varying precision requirements. We focus on how different forms of guidance can be combined to support multiple stages of interaction, enabling both efficient coarse search and accurate fine-grained correction in near-space 3D environments.

3 DESIGN PROCESS

To design auditory guidance methods for near-space target acquisition, we conducted a participatory design process with two contributors with visual impairments (one blind co-author and one collaborator with residual light perception). Across three 60-minute design sessions and a short follow-up, we defined a baseline grounded in prior work and iteratively developed additional methods to better support higher precision requirements.

3.1 Apparatus

The prototype system consisted of a HMD (Meta Quest 3), a handheld controller, and a Unity-based application running on a Windows PC. Users held handheld controllers and searched for targets in near 3D space. The system presented spatial audio to indicate the target direction, with additional support available via a trigger input when needed. Vibrotactile feedback was also used as a confirmation cue when the controller contacted the target.

3.2 Participatory Design Process

In the first session, we evaluated a baseline prototype inspired by prior work [1, 30] that used head-referenced spatial audio to indicate target direction and tempo changes to indicate proximity (a faster tempo indicating a closer distance). We also explored proximity-dependent changes in vibration; however, both contributors reported that adding too many simultaneous cues made it difficult to determine which information to attend to during exploration. Based on this feedback, we simplified the baseline by retaining head-referenced spatial audio and tempo change as the main cues, while removing continuous vibration change from the default interaction. Vibration itself was not removed entirely, but was retained later only as a confirmation cue upon target contact.

The first session showed that contributors could approach the target but often struggled to make final contact, especially for smaller targets, because the baseline method supported coarse localization but not precise adjustment. This difficulty was further evident by perceptual limitations of spatial audio, as prior work has shown that left–right discrimination is generally easier than front–back or up–down discrimination, which can remain ambiguous in auditory guidance [43].

Based on this, the second session focused on designing two supplementary methods for near-target support - (1) a verbal guidance method that explicitly stated the direction in which the hand should move (inspired by prior research on hand steering and on aiming [1, 4, 17, 23, 41]); and (2) a HAND-REFERENCED spatial-audio method that shifted the auditory reference frame from the head to the hand (drawing inspiration from [30]). In the third session, we refined these two methods, especially the HAND-REFERENCED method, by introducing a three-level pitch representation to better convey vertical information (figure 3. A final follow-up session was then used to confirm the resulting designs and prepare them for a formal study.

3.3 VERBAL

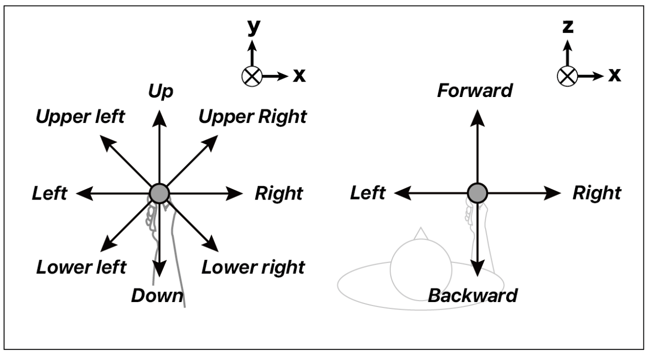

Figure 2: Directional vocabulary used in VERBAL. Spoken guidance was provided using 12 discrete directions defined relative to the current hand position.

The first supplementary method was an on-demand spoken guidance mode - VERBAL. When activated (pressing the trigger button), the system announced the direction in which the user should move the hand relative to its current position, rather than describing the absolute target location. This design was intended to compensate for directional ambiguity that remained difficult to resolve through spatial audio and tempo alone, especially for front–back and up–down adjustments near the target.

To keep the guidance concise and consistent, VERBAL used a predefined set of 12 discrete directions (Fig. 2), following prior work on language-based directional guidance [1, 4, 17, 23, 41]. By limiting the guidance to a fixed set of short direction labels, we aimed to reduce ambiguity while preserving interpretability during action. VERBAL was designed as an explicit supplementary aid for situations in which users knew they were close to the target but still needed a clearer indication of where to move next.

3.4 HAND-REFERENCED

The second supplementary method was the HAND-REFERENCED spatial-audio mode. This method was active only while the user pressed and held the trigger button. During the trigger hold, the auditory reference shifted from the head to the current hand position, allowing the user to perceive the target direction relative

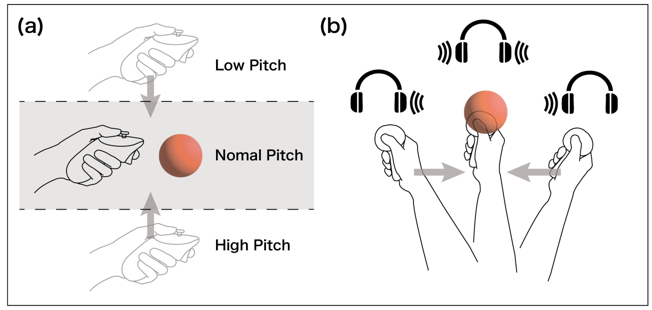

Figure 3: Overview of HAND-REFERENCED. (a) Vertical guidance was conveyed using three discrete pitch levels: high, normal, and low, depending on whether the target was above, at approximately the same height as, or below the hand. (b) While the trigger was pressed and held, the auditory reference frame shifted from the head to the current hand position, allowing the user to perceive the target direction relative to the hand.

to the hand. During early prototyping, we considered providing only directional information relative to the hand. Although this simpler design helped the contributors orient toward the target, they continued to experience difficulty with the final approach, particularly when adjusting depth. We therefore incorporated distance-dependent feedback to provide additional support for precise alignment near the target. Specifically, loudness was modulated by the hand–target distance: the sound became louder as the hand approached the target and quieter as the hand moved away, and no sound was presented when the hand–target distance exceeded the arm-reach threshold. The inclusion of distance-dependent feedback reflected a trade-off between simplifying the auditory presentation and providing sufficient support for final target alignment. When the trigger button was released, the system returned to the continuous head-referenced spatial audio used in the BASELINE condition. HAND-REFERENCED was intended to support local directional correction near the target while maintaining a non-verbal interaction style.

In addition to horizontal direction, HAND-REFERENCED also supported vertical adjustment through pitch (as in [23, 27, 30, 32]). When the target was above the hand, the system presented a higher-pitched sound, whereas a lower-pitched sound was used when the target was below the hand. This design was introduced because up–down discrimination is often difficult with spatial audio alone, and contributors similarly reported that vertical adjustment remained challenging near the target.

In an early implementation, the height cue was mapped continuously, with pitch changing linearly between a minimum and maximum value according to the vertical difference between the hand and the target. However, contributors reported that this made it difficult to identify the reference pitch corresponding to the same height, and therefore difficult to judge whether the target was slightly above, aligned with, or slightly below the hand.

In the final design, three discrete pitch levels were used: a low pitch of 250 Hz when the target was below the hand, a normal pitch of 800 Hz when the target was approximately at the same height as the hand, and a high pitch of 1350 Hz when the target was above the hand. As we did not find specific reference values in prior work, we refined these frequencies through iterative prototyping with the contributors, selecting levels that were readily distinguishable from one another and remained comfortable during repeated listening.

As shown in Fig. 3, this refinement preserved the meaning of vertical information while making the cue easier to interpret during interaction.

Overall, the participatory design process began with a baseline method that was effective for coarse localization but insufficient for near-target adjustment. We therefore developed two supplementary methods to address this limitation: VERBAL, which provided explicit spoken directional correction, and HAND-REFERENCED, which provided HAND-REFERENCED spatial-audio guidance with an additional pitch-based height cue. These resulting methods were then used in the formal user study described in the next sections.

4 USER STUDY

We conducted a user study to investigate three auditory guidance methods for supporting target locating and reaching in arm-reachable 3D space. We wanted to identify when coarse cues are sufficient and how different forms of spatial-audio guidance and supplementary directional support affected participants' performance and subjective experiences during target search.

4.1 Participants

We recruited eleven blind participants (9 males and 2 females; mean age = 50.27, SD = 15.50) through an e-newsletter for people with visual impairments (Table 1). Among them, six participants were totally blind, and five had light perception. No participant could visually perceive elements of the virtual environment during the study. Most participants rated themselves as experienced with technology but less experienced with virtual environments. Three participants never tried VR, while seven tried it once or twice, some having experience in prior research studies. All participants provided informed consent before participation. The study received approval from the Ethics Review Committee.

4.2 Apparatus

We used the Meta Quest 3, its handheld controllers, and a Unity application running on a Windows PC, with a wired connection, enabling the researchers to set up the study sessions and have more control over the experience. Participants were instructed to wear headphones, enhancing the spatialized audio.

4.3 Methodology

The study had a within-subject design, where all participants performed a target-search task in a near 3D space with the three methods in a counterbalanced order. We compared the BASELINE condition with two supplementary methods that differed in how directional information was presented during exploration.

4.3.1 Search task. Participants completed a total of nine target-search tasks, covering all combinations of the three methods (BASELINE, VERBAL, and HAND-REFERENCED) and three difficulty levels. In each task, participants were asked to reach five targets and could use both hands simultaneously. Targets were presented using spatialized audio, enabling participants to perceive their location. To systematically vary target position while maintaining balanced spatial difficulty, targets were arranged in near 3D space by combining four plane-angle conditions ($90°$, $45°$, $0°$, and $-45°$) with two distance conditions relative to each participant's calibrated reach (outer: 100%, inner: 50%), as shown in Fig. 4. These configurations were grouped into nine sets (G1–G9), each containing five targets distributed across angles, axes, and distances. Each participant completed all nine sets, and their assignment to the nine method $\times$ difficulty-level conditions was counterbalanced across participants. The exact target positions and presentation order for each set are provided in Appendix A.

Within each group, targets appeared sequentially, with only one active at a time. Each trial ended when the participant successfully reached the target or when the two-minute timeout was reached. The timeout was selected as a practical upper bound considering the overall session duration, the repeated trials, and the need to provide sufficient opportunities for rest. It was intended to give participants sufficient time to complete each trial without causing excessive fatigue or frustration. The same timeout was used across all difficulty levels to maintain a consistent maximum trial duration.

When a target was not reached within the time limit, a timeout sound was played. After either a successful reach or a timeout, participants returned the controllers to a predefined resting zone near the chest and remained there for two seconds before the next target appeared, as shown in Fig. 5 (a). The starting position was defined as a spatial region rather than a single exact coordinate, as it would be difficult to always ensure the exact same position.

Haptic feedback indicated target contact across all methods. To encourage precision and avoid sweep-like strategies, haptic feedback was triggered only when the controller remained on the target for 0.5 s, and a successful reach required maintaining contact for 2 s. Task difficulty increased across levels by reducing target size, requiring more precise alignment. Specifically, the target diameter was set to 15 cm at level 1, 10 cm at level 2, and 7.5 cm at Level 3.

We recorded task success, completion time (per target and overall), controller position, and trigger interactions for the two supplementary methods. After each difficulty level, participants rated task ease using the Single Ease Question (SEQ) [36] on a 7-point scale. All interviews were audio-recorded and transcribed using Whisper by OpenAI.

4.3.2 Procedure. All sessions were conducted by two researchers, and each session lasted approximately 150 minutes. After providing informed consent and completing a pre-study questionnaire, participants underwent a calibration procedure to register each participant's arm length (used to define individualized target distances), completed a tutorials related to the methods used, followed by the main task and post-task evaluation.

For each method, participants first completed a tutorial in which the researcher introduced the system and explained its feedback interactions. Participants were encouraged to explore the method hands-on and ask questions during this process. Once any doubts were clarified, the task began with a system-provided countdown.

Each trial began with both hands placed at a predefined start position near the chest, as shown in Figure 5 (a).

Table 1: Participant demographics and self-reported experience. Technology refers to participants' general experience using digital technologies, such as computers and smartphones, whereas Virtual Environments refers to their experience interacting with computer-generated environments, such as digital games, 3D environments, and VR applications. Both were rated from 1 (Not Experienced) to 7 (Very Experienced). VR Experience indicates participants' prior experience using VR.

ID	Gender	Dominant Hand	Age	Blindness Onset	Visual Condition	Technology	Virtual Environments	VR Experience
P1	Male	Right	70	Late	Totally blind	5	1	None
P2	Male	Right	64	Late	Light perception	2	1	None
P3	Male	Right	34	Early	Light perception	4	3	>5 times
P4	Male	Right	22	Early	Totally blind	5	3	1-2 times
P5	Male	Right	60	Early	Totally blind	5	2	1-2 times
P6	Male	Right	54	Late	Totally blind	3	3	1-2 times
P7	Female	Right	62	Late	Light perception	3	1	1-2 times
P8	Male	Right	29	Early	Totally blind	4	1	1-2 times
P9	Male	Right	49	Late	Light perception	3	1	1-2 times
P10	Female	Right	58	Late	Light perception	5	1	None
P11	Male	Left	51	Late	Totally blind	7	1	1-2 times

After completing all three levels of a given method, participants took part in a short semi-structured interview. The interview asked about the ease of understanding the target direction, ease of hand movement and fine adjustment, search strategies, and the strengths and difficulties of the method. After completing all three methods, participants answered a post-study questionnaire. We measured workload using the NASA Task Load Index (Raw-TLX) [11] and also collected additional 7-point Likert-scale ratings, including easiness, usefulness, enjoyment, fine adjustment, and confidence. Finally, a semi-structured interview was conducted to gather comparative feedback across the three methods and suggestions for future improvements. After the study ended, we thanked participants for their time and insights. All participants received approximately 50 USD for their participation.

4.3.3 Data Analysis. We evaluated task performance using quantitative and subjective measures. Performance metrics were computed at the target level, where each target was treated as a single trial.

Error rate. Error rate was defined as the proportion of targets that were not successfully reached within the time limit. For each condition (method × difficulty level), participants attempted five targets, resulting in 55 trials per condition (11 participants × 5 targets). To analyze error rate, each trial was coded as either success (1) or timeout (0). The binary outcomes were analyzed using a binomial generalized linear mixed-effects model with a logit link function. Method and difficulty level were included as fixed effects without an interaction term, while participant was included as a random intercept to account for repeated trials.

Completion time. Completion time was defined as the time required to successfully reach a target. This metric was computed for successful trials only (i.e., trials in which the target was reached before the time limit). Completion times were log-transformed and analyzed using a linear mixed-effects model. Method and difficulty level were included as fixed effects, while participant was included as a random intercept to account for repeated trials.

Supplementary cue usage. We also analyzed how participants invoked the two supplementary cues. For VERBAL, each trigger press that produced spoken guidance was counted as one VERBAL request. For HAND-REFERENCED, we calculated the trigger-hold proportion for each five-target task by dividing the total duration for which the trigger was held by the total task duration. We then averaged these task-level values within each difficulty level.

Subjective measures. Given the small sample size and the ordinal nature of the subjective measures, non-parametric tests were used throughout the analysis. Raw-TLX scores, each of the five 7-point Likert items, and SEQ scores were analyzed separately using Friedman tests to assess overall differences among the three methods. When a Friedman test was significant, pairwise comparisons were conducted using Wilcoxon signed-rank tests. Bonferroni correction was applied to the Raw-TLX comparisons, whereas Holm correction was used for the Likert items and SEQ scores. Preference rankings were summarized by counting the number of participants who ranked each method first, second, or third.

We analyzed the interview transcripts following what Braun and Clarke [10] refer to as codebook thematic analysis. Two researchers first reviewed the transcripts and then combined deductive codes derived from our research focus with inductive codes that emerged from the data. They iteratively compared and refined the codes, discussed interpretations, and resolved disagreements through discussion to finalize the themes reported in the qualitative findings.

5 RESULTS

We present our findings supported by quantitative analysis of performance and participants self-reported ratings, and a qualitative analysis based on participants feedback and observations of their navigation tasks.

5.1 Performance Analysis

5.1.1 Error Rate. Figure 6-Left) shows error rates across methods and difficulty levels. VERBAL consistently resulted in the lowest error rates across all levels, whereas BASELINE exhibited a sharp increase in error rate at level 3. HAND-REFERENCED showed relatively stable error rates across levels. These results suggest that VERBAL provides more reliable target acquisition, particularly under increased task difficulty. We analyzed error rates using a binomial generalized linear mixed-effects model with a logit link function, with method and level as fixed effects and participant as a

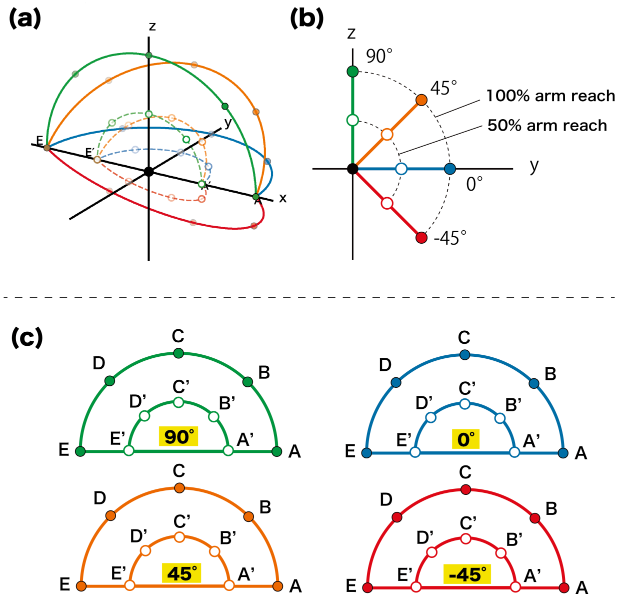

Figure 4: Simplified target layout used in the search task. (a) Combined 3D overview of the target locations in arm-reachable near 3D space. (b) Construction of the target layout. A five-position semicircular arc was rotated around the x-axis to create four plane-angle conditions ($90°$, $45°$, $0°$, and $-45°$). For each plane, targets were placed at two radial distances from the origin based on each participant's calibrated arm reach: outer targets at 100% arm reach and inner targets at 50% arm reach. Filled and open circles indicate outer and inner targets, respectively. (c) The four plane-angle conditions are shown separately for clarity. Each plane contained five outer positions (A–E) and five corresponding inner positions (A'–E'). A and E were shared endpoints across plane conditions, whereas B–D changed according to the plane rotation. The experimental five-target groups were organized using these layouts and counterbalanced across participants, methods, and difficulty levels. Difficulty level changed only the target diameter, not the target location.

random effect. The analysis revealed a significant effect of method for VERBAL ($\beta = 2.02$, $p < .001$), indicating that VERBAL significantly reduced error rates compared to BASELINE. In contrast, HAND-REFERENCED did not show a significant difference from BASELINE ($p = .20$). There was also a significant effect of difficulty level, with level 3 resulting in higher error rates compared to level 1 ($\beta = -1.34$, $p < .01$), indicating that task difficulty significantly impacted performance.

5.1.2 Completion Time. Figure 6-Right) shows completion time distributions for successful trials only, presented as boxplots by method and difficulty level. VERBAL consistently exhibited shorter completion times across all levels, with lower medians and a more compact distribution compared to BASELINE. BASELINE showed longer completion times and greater variability, particularly at level 3. HAND-REFERENCED showed intermediate performance, with broader distributions and generally higher completion times than VERBAL. These results suggest that VERBAL enables more efficient and consistent target acquisition, particularly under increased task difficulty. We analyzed completion time using a linear mixed-effects model on log-transformed time, with method and level as fixed effects and participant as a random effect. The analysis revealed a significant effect of method for VERBAL ($\beta = -0.33$, $p < .001$), indicating that VERBAL significantly reduced completion time compared to BASELINE, whereas HAND-REFERENCED did not show a significant difference from BASELINE ($\beta = -0.04$, $p = .598$). There was also a significant effect of difficulty level, with completion time increasing significantly for level 2 ($\beta = 0.26$, $p < .001$) and level 3 ($\beta = 0.50$, $p < .001$), indicating that task difficulty impacted performance.

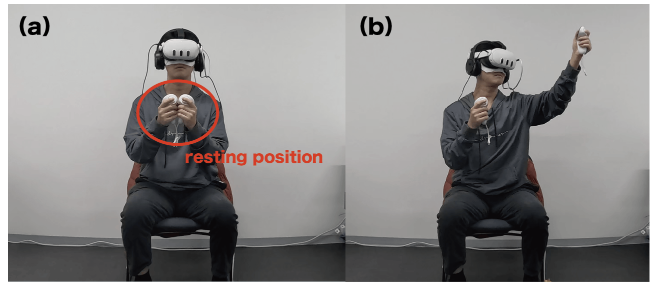

Figure 5: Overview of the participant posture and target-search task. (a) Resting position used as the standardized start/restart posture. Participants held the controllers near the chest before starting or continuing the task, including after a timeout. (b) Example of the actual target-search task, in which a participant wearing an HMD and headphones reached toward an arm-reachable target in near 3D space while using auditory cues and handheld controllers.

Taken together with the error-rate results, these findings indicate that VERBAL improves both accuracy and efficiency, particularly under higher task difficulty.

5.1.3 Supplementary Cue Usage. We also examined how participants used the supplementary cues. The mean number of VERBAL requests per five-target task increased with difficulty, from 3.8 requests at level 1 to 4.9 at level 2 and 7.3 at level 3. For HAND-REFERENCED, participants held the trigger for approximately one-third of the task duration. The mean trigger-hold proportion was 31.1% at level 1, 34.4% at level 2, and 35.4% at level 3. Overall, VERBAL requests increased with difficulty, whereas the HAND-REFERENCED trigger-hold proportion showed only a modest increase across levels.

5.2 Self-Reported Ratings

5.2.1 Workload Score. Raw-TLX results [11] showed that the mean workload score was 71 for BASELINE, 43 for VERBAL, and 58 for HAND-REFERENCED. Prior work reports typical workload scores for people with special needs in the range of 26–48 [19]. Among the three methods, VERBAL fell within this range, whereas BASELINE and HAND-REFERENCED were above it. This pattern suggests that

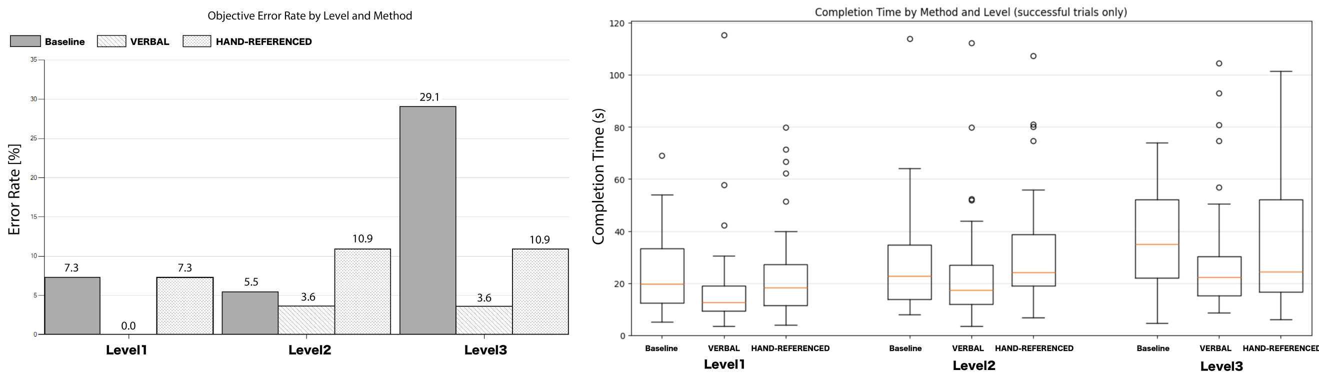

Figure 6: Performance results across methods and difficulty levels. Left: error rates for BASELINE, VERBAL, and HAND-REFERENCED in each level. Right: completion time distributions for successful trials only, shown as boxplots by method and level.

Table 2: Preference ranking across methods.

Method	1st	2nd	3rd
BASELINE	2	2	7
VERBAL	8	2	1
HAND-REFERENCED	1	7	3

VERBAL imposed the lowest workload, while BASELINE was the most demanding. HAND-REFERENCED showed an intermediate workload, indicating that hand-referenced support reduced some of the burden compared with BASELINE, but still required more effort than VERBAL.

We analyzed workload scores using a Friedman test, which revealed a significant effect of Method ($\chi^2(2) = 8.19$, $p = .017$). Post-hoc pairwise comparisons using Wilcoxon signed-rank tests with Bonferroni correction showed that VERBAL significantly reduced workload compared to BASELINE ($p = .043$), whereas HAND-REFERENCED did not show a significant difference from BASELINE ($p = .708$). No significant difference was observed between VERBAL and HAND-REFERENCED ($p = .151$). These results suggest that VERBAL imposed the lowest perceived workload, while BASELINE was the most demanding.

5.2.2 Preference Ranking. After the debriefing, participants ranked the three methods according to overall preference (Table 2). VERBAL was the most preferred method, ranked first by eight participants, compared with two for BASELINE and one for HAND-REFERENCED. HAND-REFERENCED was most often ranked second (seven participants), suggesting that it was generally viewed favorably even when it was not selected as the top choice. In contrast, BASELINE was most often ranked last (seven participants), indicating that participants generally preferred the supplementary methods over spatial audio alone.

5.2.3 7-point Likert ratings. Participants also rated each method on five 7-point Likert items: *Easiness*, *Usefulness*, *Enjoyment*, *Fine Adjustment*, and *Confidence*, referring to the ease of understanding the sound-source location, the helpfulness of the method for finding the target, the enjoyment of the target-search experience, the ease of making directional fine adjustments, and confidence in moving the hand, respectively (Fig. 7).

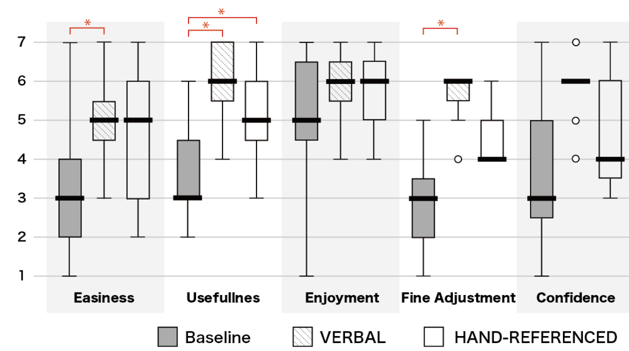

Figure 7: Boxplots of the 7-point Likert ratings for the three methods across five subjective measures: *Easiness, Usefulness, Enjoyment, Fine Adjustment*, and *Confidence*. Red brackets indicate significant pairwise differences.

Easiness. BASELINE, VERBAL, and HAND-REFERENCED received mean ratings of 3.18 (SD = 1.59), 5.09 (SD = 1.16), and 4.55 (SD = 1.56), respectively. A Friedman test showed a significant difference among methods ($p = .007$), and post-hoc Wilcoxon signed-rank tests with Holm correction showed that VERBAL was rated significantly higher than BASELINE ($p = .023$).

Usefulness. The mean ratings were 3.64 (SD = 1.23) for BASELINE, 6.00 (SD = 0.95) for VERBAL, and 5.27 (SD = 1.21) for HAND-REFERENCED. The overall difference was significant ($p < .001$). Post-hoc comparisons showed that both VERBAL ($p = .003$) and HAND-REFERENCED ($p = .008$) were rated significantly higher than BASELINE.

Enjoyment. The mean ratings were 5.18 (SD = 1.59), 5.91 (SD = 0.90), and 5.82 (SD = 0.94) for BASELINE, VERBAL, and HAND-REFERENCED, respectively. No significant difference was found among methods ($p = .112$).

Fine Adjustment. The mean ratings were 3.00 (SD = 1.71) for BASELINE, 5.82 (SD = 0.83) for VERBAL, and 4.73 (SD = 0.96) for HAND-REFERENCED. The overall difference was significant (p = .004). Post-hoc analysis showed that VERBAL was rated significantly higher than BASELINE (p = .023).

Confidence. The mean ratings were 3.64 (SD = 1.67) for BASELINE, 5.91 (SD = 0.79) for VERBAL, and 4.64 (SD = 1.37) for HAND-REFERENCED. A Friedman test showed a significant overall difference (p = .005), although no pairwise comparison remained significant after Holm correction. Overall, VERBAL received the highest ratings on most items, HAND-REFERENCED generally showed intermediate ratings, and BASELINE received the lowest ratings.

5.2.4 Perceived Ease (SEQ). Participants rated the overall ease of each method using the Single Ease Question (SEQ). The mean ratings were 4.03 (SD = 1.44) for BASELINE, 4.48 (SD = 1.45) for VERBAL, and 4.67 (SD = 1.47) for HAND-REFERENCED. A Friedman test revealed no significant difference among methods ($\chi^2(2) = 3.90$, $p = .143$). Post-hoc pairwise comparisons using Wilcoxon signed-rank tests with Holm correction also showed no significant differences between any pair of methods ($p > .12$). These results suggest that perceived ease did not differ significantly across methods.

5.3 Qualitative Findings

In addition to the quantitative results, we analyzed observation notes and debriefing comments to understand which stages of the search process were supported by each method and what difficulties remained. Participants highlighted different strengths depending on the phase of exploration and target location BASELINE, VERBAL, and HAND-REFERENCED served distinct roles and were used non-uniformly. Embodied search strategies, such as head orientation, arm extension, torso leaning, and controller posture, strongly influenced performance and method use.

5.3.1 BASELINE. BASELINE supported target acquisition in Level 1 and Level 2, but error rates increased substantially in Level 3, indicating that spatial audio and tempo cues were sufficient for coarse localization but not for fine adjustment.

All participants described BASELINE as useful for rough directional estimates, especially horizontally. P1 explained that "left and right were easier to understand when I turned my head, but up/down and front/back were difficult", while P2 and P9 further emphasized the difficulties in understanding distance. P7 described a strategy of first identifying the rough left–right direction, then extending the arm and searching for the region where the tempo became faster.

BASELINE was also appreciated for its simplicity. Because it required no additional operation, Three participants (P4, P5, and P6) noted that the absence of additional operations or support helped them concentrate on the sound or their bodily movement. For instance, P4 described it as "easy in the sense that I did not have to do anything extra, and I could concentrate on moving my body." Participants emphasized that reduced information supported more natural exploration. For instance, P10 described it as "similar to searching for an object," adding that "I liked being able to think and move based on my own sense."

At the same time, all participants consistently reported difficulty with precise alignment once near the target, especially for smaller targets. P3, for instance, explained that "for the first level, sound alone was enough to find the target, but once the targets became smaller, it became difficult to search using sound alone." P6 expanded this thought indicating that moving the face helped distinguish left–right differences, but that facial movements toward diagonal-up or upward directions were themselves difficult. Difficulties with verticality and depth were especially common. P5 said that "I wanted more information about the vertical direction", while P7 reported that "up/down and near/far adjustments were difficult."

Observation notes showed that failures often occurred during the final adjustment phase rather than initial localization. Participants frequently reached the correct region but failed to complete the final movement. For example, in Level 3, P2 often reached the approximate area but then continued circling around it, while P7 often stayed in a region where the tempo was already fast, but could not determine how much further to move, especially upward. Taken together, these patterns suggest that BASELINE was useful for coarse localization, but offered limited support for precise acquisition near the target.

5.3.2 VERBAL. VERBAL was perceived as clear and reassuring, supporting confident movement. All participants noted that speech reduced uncertainty and hesitation; as P10 explained, "words are reassuring" and that "without words, I feel anxious." P2 felt that the combination of sound and speech was especially understandable because spoken guidance resembled forms of assistance already familiar in daily life.

VERBAL was generally used selectively, although the timing and manner of use varied across participants. Ten participants (except P6) explicitly described using spatial audio to establish a coarse estimate of the target location before using speech mainly for confirmation or fine adjustment. P6, by contrast, emphasized frequently requesting verbal guidance during both the approach and adjustment phases. P4, for example, preferred to rely on sound first and only use speech when the position remained unclear. Similarly, P9 described first identifying left–right direction, approaching the region, and then using speech for minor corrections. P8 valued having both sound and speech cues available, but stressed that "it is definitely better if I can press it at my own timing."

Using speech from the start was not always effective. Participants initially relied heavily on it, but later shifted strategies. P3 initially relied heavily on verbal guidance, but later shifted to first placing the controller near the perceived target region and then using speech as a secondary tool. P7 noted that trusting speech too early could cause confusion, and instead emphasized that "it is important to rely on the direction of the sound first." This suggests that VERBAL was most effective when layered on top of an initial spatial estimate, serving as a fallback or "last resort", as posed by P7, to reduce uncertainty near the target.

A key limitation was that VERBAL conveyed direction but not magnitude. Seven participants (P1, P3, P4, P5, P6, P10, and P11) reported that verbal guidance indicated where to move but not how far. As P1 noted, "it was difficult to know how far I should move." P10 also explained that they sometimes moved too far in the near–far dimension and therefore still had to make gradual self-corrections.

The findings also suggest that simply increasing the number of verbal expressions would not necessarily solve this limitation. P10 felt that hearing exact distances such as centimeters might be less useful than making small self-adjustments, while P11 mentioned that adding too many directional variants could itself increase difficulty. Overall, VERBAL provided clear and reassuring guidance, particularly near the target, but fine adjustment still relied on iterative movement.

5.3.3 HAND-REFERENCED. HAND-REFERENCED was less often preferred, but frequently ranked second. It provided richer continuous information, particularly for height and proximity, but required more interpretation.

Its strongest advantage was vertical guidance. All participants consistently reported that pitch supported up–down discrimination. For example, both P1 and P11 referred to using pitch for up–down, stereo balance for left–right, and tempo for depth. P4 reported that "matching height was easy" and that changes in sound felt more intuitive than verbal expressions. Similarly, some participants used volume changes to estimate distance.

The use of HAND-REFERENCED varied not only across participants, but also within participants depending on task difficulty and the stage of the search. Participants sometimes used prolonged holds when the target direction remained uncertain or when broader exploration was needed, but shifted to brief or selective activations after establishing a coarse target location, using the cues to confirm or refine specific dimensions. P2 described HAND-REFERENCED as helpful for both left–right and up–down, but also cognitively demanding because it required simultaneous processing of horizontal and vertical information. P6 tended to use the trigger only briefly, mainly to check pitch. P11 also oriented the body and head first, then used short clicks to confirm and refine up–down, left–right, and front–back adjustments. Thus, HAND-REFERENCED did not function exclusively as either a primary exploration method or a confirmation tool. Instead, its role shifted between broader exploration and targeted refinement according to the participant's uncertainty, the target difficulty, and the stage of acquisition.

The multidimensional feedback provided by HAND-REFERENCED also introduced interpretive demands for some participants. Three participants (P2, P6, and P10) reported difficulty interpreting multiple auditory cues simultaneously. P2 described the combination of left–right and height information as complex because of the amount of information presented. P6 similarly reported difficulty coordinating trigger use while processing multiple cues. P10 found it difficult to determine whether the pitch was high, low, or neutral while also attending to stereo balance and tempo. P10 also noted that prolonged trigger use made it difficult to keep track of the current mode and that speech was easier to process in this respect. These comments highlight a trade-off between information richness and interpretability: although HAND-REFERENCED made multiple spatial dimensions continuously available, interpreting and integrating these cues was challenging for some participants.

It also did not fully resolve depth-related difficulty. Participants still relied primarily on tempo for front–back adjustment, and five participants (P1, P4, P7, P8, and P9) reported that depth or front–back adjustment remained challenging even near the target. Thus, HAND-REFERENCED provided continuous cues for some participants, especially for height and proximity, but did not eliminate the difficulty of coordinating multiple dimensions near the target. At the same time, performance across difficulty levels was relatively stable, suggesting that participants began to adapt to the mapping with use. While this study does not explicitly evaluate learning effects, this pattern indicates that familiarity may reduce interpretation effort over time and improve the effective use of the available cues.

5.3.4 Embodied search strategies. Across all three methods, embodied search strategies strongly influenced performance and method use. During the tutorial, participants were instructed to move the head or face to better interpret spatial audio. In the actual tasks, however, some participants actively oriented the head and body toward the target, whereas others kept the head relatively fixed and relied mainly on arm movement.

Three participants (P9, P10, and P11) actively oriented both the head and body toward the target and showed relatively stable search strategies. P9 actively rotated both head and body even with BASELINE, and looked clearly upward for upper targets. P10 also used explicit head and torso rotation, first using head direction to establish the rough target direction and then adjusting up–down and front–back. P11 similarly aligned the body and head toward the target before moving the controller in different directions and using the trigger for confirmation. These participants often showed more direct movement patterns.

In contrast, orienting the face and head toward the target was not natural for everyone. P6 reported that "moving the face itself was difficult," even though moving the face helped distinguish left–right differences. P8 also tended not to orient the face toward the target even after practice, relying primarily on arm movement during search.

For two participants (P5 and P6), reaching toward the upper targets showed limitations of the controller tracking, making these targets more difficult to reach (*e.g.*, C 90° - Fig 4). P5 often tilted the head downward while raising the controller, which may have made it more difficult for the HMD camera to track the controller reliably. A similar issue was noted for P6, where unstable controller tracking during upper-target trials may have affected successful registration. Thus, the difficulty of upper targets did not seem to arise solely from auditory perception, but also from the bodily action required to raise the arm and from technical constraints of the HMD-based tracking setup.

Participants also used torso movement or a restricted reach range instead of fully extending the arm. P5 tended to lean the torso forward instead of fully stretching the arm, even when the target remained within reachable distance. P7 searched within a limited reach range rather than fully extending the arm, making it difficult to close the remaining depth gap for near targets. P8 reported that, when the arm became tired, they sometimes switched to the other hand, but then lost the previously explored position and had to restart the search from the beginning.

Three participants (P7, P8, and P11) described difficulty maintaining one movement dimension while adjusting another. P7 said that under HAND-REFERENCED, "even when I thought I was moving in parallel, my hand moved upward instead," making directional correction difficult without vision. P8 likewise reported that "it was difficult to move sideways while keeping the same height." P11 also described moments in which "the arm position I imagined and the actual position of the arm did not match," making fine adjustment difficult. These comments suggest that 3D target search involved not only hearing and interpreting auditory cues, but also stabilizing the hand trajectory without visual confirmation.

Overall, these observations indicate that target search was shaped not only by the feedback method itself, but also by the way participants coordinated the head, torso, arm, wrist, and controller. In particular, upper targets, targets near the body, and targets requiring depth adjustment were influenced not only by the difficulty of interpreting auditory cues, but also by posture and embodied movement strategies.

6 DISCUSSION

We investigated if and how three auditory feedback methods support blind people in locating and reaching targets in arm-reachable 3D spaces. VERBAL achieved the strongest overall performance. However, the qualitative findings indicate that the three methods supported different stages of target acquisition and were used in distinct ways. Since VERBAL relied on discrete spoken requests, whereas HAND-REFERENCED provided continuous feedback while the trigger was held, the results should be interpreted as differences between the overall interaction methods rather than as a direct comparison of equivalent auditory feedback. The following sections discuss coarse and fine target acquisition, the roles of the two supplementary methods, failure trajectories, and body-relative constraints.

6.1 Coarse Search vs. Final Acquisition

Our results show that identifying the approximate direction of a target and precisely acquiring it at the end of the movement are different challenges. BASELINE worked reasonably well in Level 1 and Level 2, but its error rate increased substantially in Level 3. In the qualitative findings, many participants reported that they could understand the left–right direction or the approximate target region, while vertical direction, depth, and near-target fine adjustment remained difficult. Figure 8 shows representative trajectories: under BASELINE, participants often approached the target region but continued local exploratory movements near the target. These results suggest that spatial audio and tempo changes supported coarse localization, but became insufficient for final acquisition when the acquisition tolerance became tighter.

These findings suggest that nonvisual 3D target acquisition comprises two distinct stages: coarse localization and final acquisition. BASELINE primarily supported directional approach toward the target region, whereas VERBAL and HAND-REFERENCED became more useful during the final adjustment phase, when participants required confirmation or more localized feedback. This indicates that the information needed during acquisition changes as users approach the target.

This staged interpretation aligns with prior work on nonvisual 3D guidance. Chung et al.'s Two-Step Nonvisual Omnidirectional Guidance showed that separating horizontal and vertical guidance, rather than presenting all 3D directional information at once, improved completion time and travel distance [15]. Our findings extend this perspective to BLV users' target acquisition in arm-reachable near 3D space. Spatial audio was useful for establishing the initial direction, but as participants approached the target and the remaining error became smaller, supplementary cues such as verbal confirmation or hand-referenced auditory feedback became more important.

6.2 On-Demand Speech Support

VERBAL showed the strongest performance and subjective ratings in our study. However, this result should not be interpreted simply as evidence that speech is always the best form of support. A key aspect of our design was that VERBAL was not a continuously delivered instruction stream. Instead, participants could request verbal guidance at their own timing. In the qualitative findings, many participants first used spatial audio to form a rough estimate of the target direction and then invoked VERBAL for confirmation or fine adjustment. Thus, VERBAL did not replace spatial audio; rather, it helped participants confirm the spatial estimate they had already formed and reduce uncertainty near the target.

This differs from prior auditory hand-steering work by Abe et al.; in their study, verbal guidance was clear and familiar, but when presented as continuous verbal instructions, it could increase cognitive load, reduce users' sense of agency, and make movements feel robotic [1]. In contrast, VERBAL in our study was user-initiated. This may explain why participants experienced it as reassurance that they could invoke when needed, rather than as a continuous command.

At the same time, VERBAL provided directional correction without prescribing the full movement, leaving participants to determine movement extent themselves. Increasing the specificity of verbal instructions could reduce ambiguity, but it may also increase the effort required to translate speech into movement.

These findings align with prior work showing that users differ in how they process verbal-spatial information [6] and that speech can become burdensome in time-sensitive interactions [23]. Together with our results, this suggests that speech is most effective as concise, user-initiated support that complements continuous spatial feedback, rather than replacing it. While this approach appears well suited to static target acquisition, longer or dynamic tasks may benefit from combining intermittent speech with continuous auditory or haptic guidance [22, 38].

6.3 HAND-REFERENCED Depends on Task Structure

HAND-REFERENCED did not outperform VERBAL overall, but it was evaluated more favorably than BASELINE in several aspects, and many participants ranked it second. This suggests that HAND-REFERENCED was not simply an unsuccessful method, but rather a form of support with a different role from VERBAL. In the qualitative findings, HAND-REFERENCED functioned as a useful cue for some participants when checking height or lateral direction. In particular, pitch-based vertical information and hand-referenced left–right cues were reported as advantages not available in BASELINE. However, precise depth adjustment remained difficult even

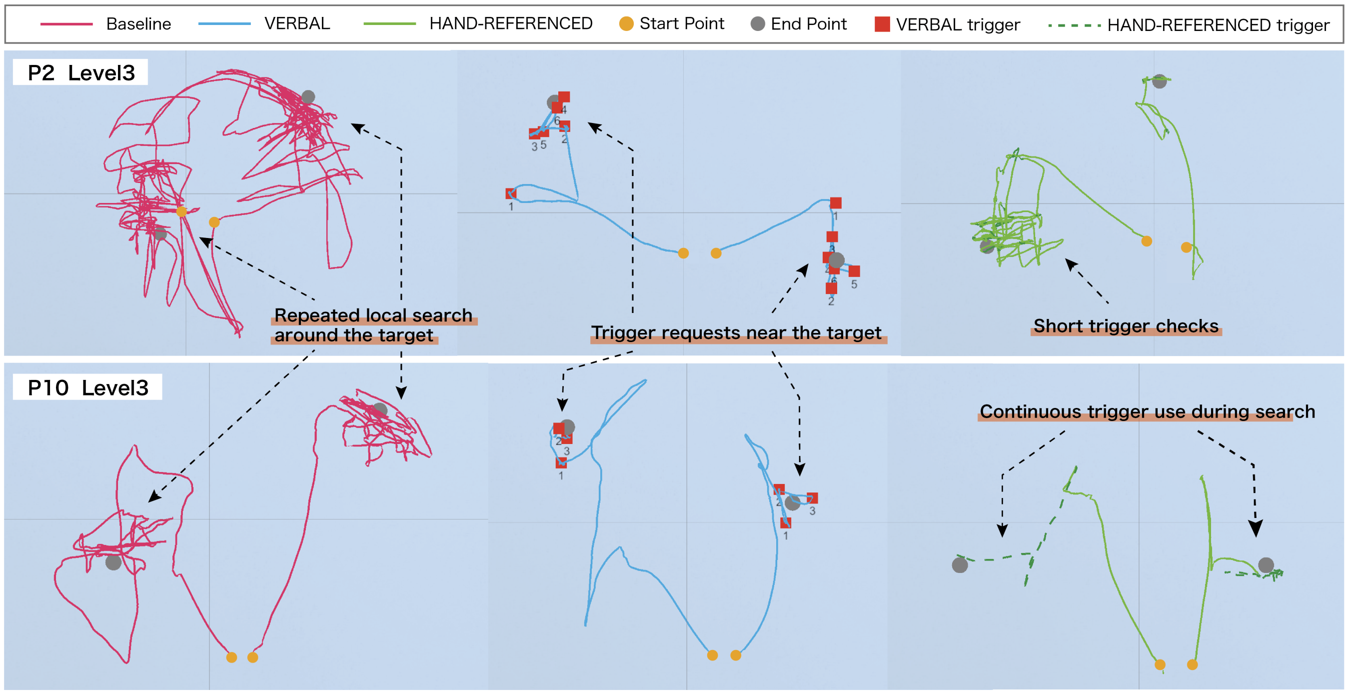

Figure 8: Representative Level 3 trajectories illustrating how the three methods shaped participants' search behavior. For each participant and method, two target-search trials were selected from the five Level 3 targets. Magenta, blue, and green lines indicate BASELINE, VERBAL, and HAND-REFERENCED trajectories, respectively. Yellow circles indicate start positions, gray circles indicate trial end positions, red squares indicate moments when participants requested VERBAL guidance, and dashed green segments indicate periods when the HAND-REFERENCED trigger was held. BASELINE trajectories show that participants could often reach the approximate target region but continued exploratory movements during final adjustment. VERBAL trajectories show that participants requested speech during local correction near the target. HAND-REFERENCED trajectories highlight different usage patterns: P2 used HAND-REFERENCED intermittently as a check-in cue, whereas P10 held the trigger for longer periods and used HAND-REFERENCED more continuously during the search.

with HAND-REFERENCED. Volume and tempo changes could provide a coarse sense of distance, but they did not always resolve front–back error during near-target correction.

This does not mean that hand-referenced auditory feedback is ineffective. Rather, its usefulness depends on the structure of the task. In aiming tasks such as Mendes et al. [30], users do not need to physically place the hand at the 3D position of the target. Instead, the task primarily involves correcting the aiming direction or the relationship between the reticle and the target. In that context, hand- or reticle-referenced auditory feedback can provide a relatively direct cue for correcting directional error. Our task required participants to place and stabilize the controller at the target while simultaneously coordinating lateral position, height, depth, arm extension, and controller posture. Although HAND-REFERENCED helped participants check vertical and lateral alignment, interpreting multiple auditory dimensions simultaneously remained demanding, particularly for depth adjustment during final acquisition.

This distinction suggests an important boundary condition for the use of hand-referenced feedback. Such cues appear effective for correcting directional misalignment or checking height, but seem less suited to tasks requiring precise 3D hand placement and depth adjustment. Rather than contradicting previous work on nonvisual aiming, our results clarify the task conditions under which hand-referenced feedback is likely to be most effective.

6.4 Interpreting Error Trajectories

To better understand the error patterns, we analyzed error-trial trajectories using normalized remaining distance over time. We focused on BASELINE and HAND-REFERENCED because VERBAL produced very few errors, whereas BASELINE and HAND-REFERENCED showed a more informative contrast. In particular, HAND-REFERENCED had a higher error rate than BASELINE in Level 2, even though HAND-REFERENCED was generally rated more favorably than BASELINE in the subjective measures. Across 495 objective-level trials, there were 43 errors in total: 23 in BASELINE, 4 in VERBAL, and 16 in HAND-REFERENCED. BASELINE errors were concentrated in Level 3, with 16 of the 23 BASELINE errors occurring at the tightest tolerance. In contrast, errors were distributed across all three levels in the HAND-REFERENCED condition, with 4, 6, and 6 errors in Levels 1–3, respectively.

Figure 9 shows the normalized remaining distance to the target for error trials in BASELINE and HAND-REFERENCED. The x-axis is normalized from trial start to trial end, and the y-axis

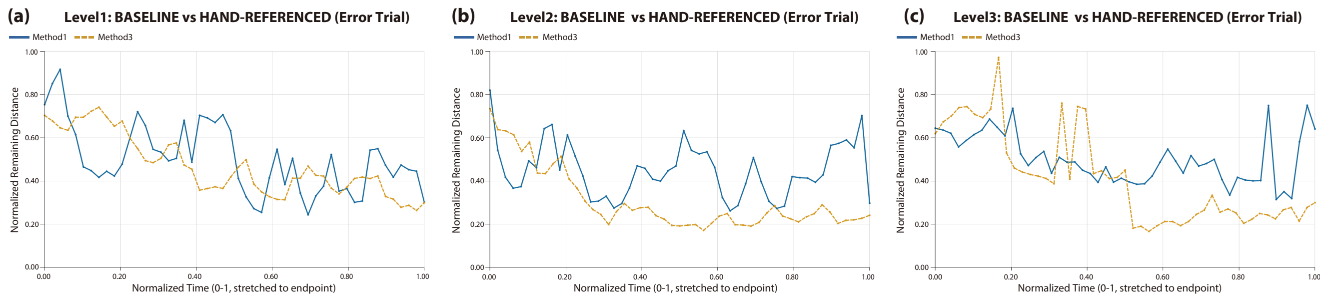

Figure 9: Normalized remaining distance over time for error trials in BASELINE and HAND-REFERENCED. Each panel shows one difficulty level: (a) Level 1, (b) Level 2, and (c) Level 3. Only error trials are included. The x-axis is normalized from trial start to trial end, and the y-axis shows the remaining distance to the target normalized by the initial distance at trial start. Red and green lines indicate BASELINE and HAND-REFERENCED, respectively. The trajectories provide an exploratory view of how participants approached the target before the trial ended without success.

shows the remaining distance to the target normalized by the initial distance at trial start. Because some method-level combinations included only a small number of error trials, we use this analysis as an exploratory inspection of failure tendencies rather than as a confirmatory statistical comparison.

The trajectories suggest distinct failure patterns for BASELINE and HAND-REFERENCED. In Level 2, where HAND-REFERENCED had a higher error rate than BASELINE, HAND-REFERENCED error trials nevertheless tended to approach and remain closer to the target toward the later part of the trial. A similar tendency appears in Level 3. In contrast, BASELINE error trials showed more fluctuating distance patterns and sometimes moved away from the target later in the trial. This suggests that some HAND-REFERENCED errors were not failures to approach the target region. Rather, participants appeared to get close to the target but did not complete the final acquisition.

These trajectories suggest that HAND-REFERENCED often supported a successful approach toward the target without consistently enabling final acquisition. Consequently, error rate alone may underestimate its usefulness during the earlier stages of target acquisition.

6.5 Body-Relative Target Constraints

The qualitative findings suggest that acquisition difficulty depended not only on the feedback method but also on the target's location relative to the body. Participants repeatedly reported difficulty with vertical direction, depth adjustment, upper targets, and fine correction near the body. These difficulties indicate that near-space 3D acquisition was not merely an auditory localization problem, but also involved bodily movement and tracking constraints. The target-location analysis supports this interpretation. Accuracy varied across the workspace, with C′90° (68.2%) and D45° (77.3%) showing the lowest performance across Levels 1–3. These positions correspond to the upper, depth-dependent, and near-body regions that participants also identified as particularly challenging.

Part of this difficulty can be explained by the perceptual characteristics of spatial audio. Vertical localization is generally less precise than horizontal localization. Perrott and Saberi showed that minimum audible angle thresholds are larger in elevation than in azimuth, indicating that vertical discrimination is coarser than horizontal discrimination [31]. This aligns with our findings: participants often reported that left–right direction was relatively understandable, whereas up–down and depth-related adjustments remained difficult.

Depth-related difficulty may reflect a broader challenge of target selection in 3D space. Prior work has shown that target depth affects direct-selection performance in HMD environments even when visual feedback is available [26]. In our study, participants similarly reported difficulty with depth adjustment and fine-grained correction while relying on auditory and proprioceptive cues. Together, these findings suggest that depth can be challenging in both visually guided and non-visual 3D target acquisition, although the magnitude and causes of the difficulty cannot be directly compared across the two settings.

Previous work has shown that allowing listeners to move their heads improves spatial-audio localization by reducing elevation errors and front–back confusions [29]. Accordingly, participants were encouraged during the tutorial to move their head to better understand sound direction. However, they adopted different strategies during the task: some oriented their head and body toward the target, whereas others primarily searched with the arm. For elevated targets, some participants also lowered their face while raising the controller, making it difficult to maintain orientation toward the target. This variability is consistent with prior work suggesting that face and gaze orientation do not necessarily play the same role across blind and sighted interaction contexts [13, 37]. Together, these findings suggest that HMD-based spatial-audio systems should explicitly consider head orientation as part of the interaction design, rather than assuming users will naturally maintain a consistent listening posture.

Participants also struggled to maintain one movement dimension while adjusting another. For example, keeping the hand at the same height while moving laterally, or adjusting depth without unintentionally changing height, was difficult without vision. This difficulty was not simply a matter of knowing which direction to move; it was also a matter of translating auditory information into stable embodied movement. These embodied strategies may be

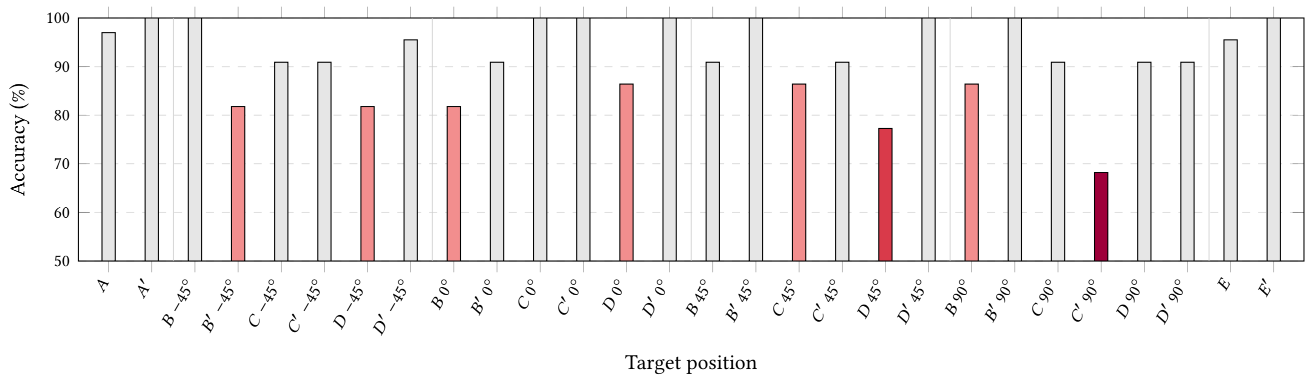

Figure 10: Accuracy at each target position. Unprimed labels denote outer targets (100% arm reach), and primed labels denote inner targets (50% arm reach). Lower-accuracy positions are highlighted with progressively stronger red. Thin vertical lines indicate boundaries between angle groups. The y-axis is truncated to 50–100% to emphasize differences near ceiling performance.

shaped by multiple interacting factors, including blindness onset, residual vision, motor experience, VR experience, and age. For upper targets, this embodied difficulty may have further interacted with HMD controller tracking.

6.6 Limitations

This study used VR as a controlled testbed to isolate the auditory guidance methods and systematically vary spatial conditions and precision requirements. However, this experimental setup does not capture all cues and constraints that may arise in less controlled settings. Future VR studies could, for instance, incorporate force feedback to approximate contact with a physical target or systematically examine the effects of the background noise of a realistic environment. Further work should also investigate broader real-world constraints, such as competing sounds, concurrent activities, and attentional demands. Accordingly, our findings should be interpreted in the context of static, near-space target acquisition under controlled VR and auditory conditions.

The participant sample included diverse backgrounds in terms of visual condition, experience, and interaction strategies. While this diversity is important, the relatively small sample size prevents strong claims about how individual factors such as blindness onset or prior VR experience influence behavior. In addition, the study was conducted within a single session, capturing initial use rather than longer-term adaptation. Participants may refine how and when they invoke supplementary cues with practice, particularly for methods such as VERBAL and HAND-REFERENCED.

The experimental setup relied on an HMD and handheld controller to ensure controlled tracking, but this may not reflect how such techniques would be used in practice. Holding a controller can constrain movement, and tracking accuracy may vary depending on posture and visibility (e.g., consider conveying the location of holds during activities such as bouldering). Future systems should explore alternative sensing approaches, such as hand tracking or wearable sensors, to better support natural interaction.

Finally, we evaluated fixed auditory cue designs to enable comparison across methods. While this enabled systematic analysis, the designs may not suit all users equally, and combining multiple auditory dimensions can increase the effort required for interpretation. Future work should explore adaptive or personalized feedback, while balancing the amount of information with usability. Beyond object acquisition, these findings may also inform broader movement guidance scenarios, where feedback supports coordination across multiple body parts and incorporates not only spatial guidance but also movement timing.

7 CONCLUSION

In this paper, we investigated auditory guidance for non-visual target acquisition in arm-reachable 3D space. A study with 11 blind participants showed that VERBAL guidance provided the strongest overall support, with lower error rates, faster completion times, lower workload, and the highest preference. Spatial audio in the BASELINE condition was useful for coarse localization but less effective for vertical, depth-related, and near-target adjustment. HAND-REFERENCED guidance supported height and lateral correction but required participants to interpret distinct auditory cues and did not fully resolve depth-related difficulty.

These findings suggest that non-visual target acquisition involves both coarse localization and fine correction. Auditory guidance should therefore combine head-referenced spatial audio for initial orientation with on-demand supplementary cues for precise alignment. Future work should examine dynamic environments, longer-term use, controller-free tracking, and personalized combinations of auditory and haptic feedback.

Acknowledgments

This work was supported by JSPS KAKENHI (JP26KJ0629). João Guerreiro was partially funded by FCT – Fundação para a Ciência e a Tecnologia, through the LASIGE Research Unit, DOI: https://doi.org/10.54499/UID/00408/2025, and the FCT Mobility program, ref. FCT/Mobility/1379258621/2024-25.

References

[1] Yuki Abe, Rose Xin Lin, Kotaro Hara, and Daisuke Sakamoto. 2026. Understanding the Feasibility of Auditory Hand-Steering Guidance for Blind and Low-Vision People. In Proceedings of the 2026 CHI Conference on Human Factors in Computing

Systems (CHI '26). Association for Computing Machinery, New York, NY, USA, Article 467, 16 pages. doi:10.1145/3772318.3790782

[2] Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Marco Ciardullo, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, and Giorgio Presti. 2023. Sonification of navigation instructions for people with visual impairment. International Journal of Human-Computer Studies 177 (2023), 103057. doi:10.1016/j.ijhcs.2023.103057

[3] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: a navigational cognitive assistant for the blind. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (Florence, Italy) (MobileHCI '16). Association for Computing Machinery, New York, NY, USA, 90–99. doi:10.1145/2935334.2935361

[4] Dragan Ahmetovic, Daisuke Sato, Uran Oh, Tatsuya Ishihara, Kris Kitani, and Chieko Asakawa. 2020. ReCog: Supporting Blind People in Recognizing Personal Objects. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376143

[5] Florian Apavou, Tifanie Bouchara, and Patrick Bourdot. 2025. AVA: An Audio-based Virtual Aiming System for Accessible VR Shooting Games. In 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 1172–1173. doi:10.1109/VRW66409.2025.00232

[6] Jane H. Barrow and Carryl L. Baldwin. 2015. Individual Differences in Verbal-Spatial Conflict in Rapid Spatial-Orientation Tasks. Human Factors 57, 3 (2015). doi:10.1177/0018720814553792

[7] Arnav Bharadwaj, Saurabh Bhaskar Shaw, and Daniel Goldreich. 2019. Comparing Tactile to Auditory Guidance for Blind Individuals. Frontiers in Human Neuroscience Volume 13 - 2019 (2019). doi:10.3389/fnhum.2019.00443

[8] Jeffrey P. Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010. VizWiz::LocateIt: Enabling Blind People to Locate Objects in Their Environment. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 65–72. doi:10.1109/CVPRW.2010.5543821

[9] Doug A. Bowman, Joseph L. Gabbard, and Deborah Hix. 2002. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. Presence: Teleoperators and Virtual Environments 11, 4 (08 2002), 404–424. doi:10.1162/105474602760204309 arXiv:https://direct.mit.edu/pvar/article-pdf/11/4/404/1623807/105474602760204309.pdf

[10] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology 18, 3 (2021), 328–352. doi:10.1080/14780887.2020.1769238 arXiv:https://doi.org/10.1080/14780887.2020.1769238

[11] James C. Byers, Alvah C. Bittner, and Susan G. Hill. 1989. Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. https://api.semanticscholar.org/CorpusID:58504578

[12] Kaixin Chen, Yongzhi Huang, Yicong Chen, Haobin Zhong, Lihua Lin, Lu Wang, and Kaishun Wu. 2022. LiSee: A Headphone that Provides All-day Assistance for Blind and Low-vision Users to Reach Surrounding Objects. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6, 3, Article 104 (2022). doi:10.1145/3550282

[13] Silvia Chiesa, Dario Galati, and Susanna Schmidt. 2015. Communicative interactions between visually impaired mothers and their sighted children: Analysis of gaze, facial expressions, voice and physical contacts. Child: Care, Health and Development 41, 6 (2015), 1040–1046. doi:10.1111/cch.12274

[14] SeungA Chung, Kyungyeon Lee, and Uran Oh. 2020. 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). 206–210. doi:10.1109/ISMAR-Adjunct51615.2020.00061

[15] SeungA Chung, Kyungyeon Lee, and Uran Oh. 2021. Understanding the Two-Step Nonvisual Omnidirectional Guidance for Target Acquisition in 3D Spaces. In 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 339–346. doi:10.1109/ISMAR52148.2021.00050

[16] James M. Coughlan, Brandon Biggs, Marc-Aurèle Rivière, and Huiying Shen. 2020. An Audio-Based 3D Spatial Guidance AR System for Blind Users. In Computers Helping People with Special Needs, Klaus Miesenberger, Roberto Manduchi, Mario Covarrubias Rodriguez, and Petr Peňáz (Eds.). Springer International Publishing, Cham, 475–484.

[17] Zhitong Guan, Zeyu Xiong, and Mingming Fan. 2024. FetchAid: Making Parcel Lockers More Accessible to Blind and Low Vision People With Deep-learning Enhanced Touchscreen Guidance, Error-Recovery Mechanism, and AR-based Search Support. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. doi:10.1145/3613904.3642213

[18] Renan Guarese, Fabio Zambetta, and Ron van Schyndel. 2023. Evaluating micro-guidance sonification methods in manual tasks for Blind and Visually Impaired people. In Proceedings of the 34th Australian Conference on Human-Computer Interaction (Canberra, ACT, Australia) (OzCHI '22). Association for Computing Machinery, New York, NY, USA, 260–271. doi:10.1145/3572921.3572929

[19] Morten Hertzum. 2021. Reference Values and Subscale Patterns for the Task Load Index (TLX): A Meta-Analytic Review. Ergonomics 64 (01 2021), 1–14. doi:10.1080/00140139.2021.1876927

[20] Xuhui Hu, Aiguo Song, Zhikai Wei, and Hong Zeng. 2022. StereoPilot: A Wearable Target Location System for Blind and Visually Impaired Using Spatial Audio Rendering. IEEE Transactions on Neural Systems and Rehabilitation Engineering 30 (2022), 1621–1630. doi:10.1109/TNSRE.2022.3182661

[21] Jason Jerald. 2015. The VR book. Morgan & Claypool, San Rafael, CA.

[22] Roberta L. Klatzky, James R. Marston, Nicholas A. Giudice, Reginald G. Golledge, and Jack M. Loomis. 2006. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. Journal of Experimental Psychology: Applied 12, 4 (2006), 223–232. doi:10.1037/1076-898X.12.4.223

[23] Diogo Lança, Manuel Piçarra, Inês Gonçalves, Uran Oh, André Rodrigues, and João Guerreiro. 2024. Speed-of-Light VR for Blind People: Conveying the Location of Arm-Reach Targets. In Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (St. John's, NL, Canada) (ASSETS '24). Association for Computing Machinery, New York, NY, USA, Article 110, 5 pages. doi:10.1145/3663548.3688533

[24] J.J. LaViola, E. Kruijff, R.P. McMahan, D. Bowman, and I.P. Poupyrev. 2017. 3D User Interfaces: Theory and Practice. Pearson Education. https://books.google.co.jp/books?id=fxWSDgAAQBAJ

[25] Sooyeon Lee, Chien Wen Yuan, Benjamin V. Hanrahan, Mary Beth Rosson, and John M. Carroll. 2017. Reaching Out: Investigating Different Modalities to Help People with Visual Impairments Acquire Items. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility. Association for Computing Machinery, New York, NY, USA, 389–390. doi:10.1145/3132525.3134817

[26] Paul Lubos, Gerd Bruder, and Frank Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In 2014 IEEE Symposium on 3D User Interfaces (3DUI). 11–18. doi:10.1109/3DUI.2014.6798834

[27] Shan Luo, Jianan Johanna Liu, and Botao Amber Hu. 2024. Designing a Safe Auditory-Cued Archery Exertion Game for the Visually Impaired and Sighted to Enjoy Together. In Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3663548.3688510

[28] Alexander Marquardt, Ernst Kruijff, Christina Trepkowski, Jens Maiero, Andrea Schwandt, André Hinkenjann, Wolfgang Stuerzlinger, and Johannes Schöning. 2018. Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3281505.3281525

[29] Ken I. McAnally and Russell L. Martin. 2014. Sound localization with head movement: implications for 3-D audio displays. Frontiers in Neuroscience 8 (2014), 210. doi:10.3389/fnins.2014.00210

[30] João Mendes, Manuel Piçarra, Inês Gonçalves, André Rodrigues, and João Guerreiro. 2025. Exploring Aiming Techniques for Blind People in Virtual Reality. IEEE Transactions on Visualization and Computer Graphics 31, 5 (2025), 3267–3274. doi:10.1109/TVCG.2025.3549847

[31] David R. Perrott and Kourosh Saberi. 1990. Minimum audible angle thresholds for sources varying in both elevation and azimuth. Journal of the Acoustical Society of America 87, 4 (1990), 1728–1731. doi:10.1121/1.399421

[32] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2019. WatchOut: Obstacle Sonification for People with Visual Impairment or Blindness. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (Pittsburgh, PA, USA) (ASSETS '19). Association for Computing Machinery, New York, NY, USA, 402–413. doi:10.1145/3308561.3353779

[33] Abhijeet Raina, Ronak Mohanty, Abhirath Bhuvanesh, Divya J, Manohar Swaminathan, and Vinayak Krishnamurthy. 2023. Pointing Tasks Using Spatial Audio on Smartphones for People With Vision Impairments. Journal of Computing and Information Science in Engineering 23 (05 2023). doi:10.1115/1.4062426

[34] Andreas Reichinger, Anton Fuhrmann, Stefan Maierhofer, and Werner Purgathofer. 2016. Gesture-Based Interactive Audio Guide on Tactile Reliefs. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (Reno, Nevada, USA) (ASSETS '16). Association for Computing Machinery, New York, NY, USA, 91–100. doi:10.1145/2982142.2982176

[35] Microsoft Research. 2023. Microsoft Soundscape. https://www.microsoft.com/en-us/research/product/soundscape/

[36] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of three one-question, post-task usability questionnaires. 10 pages. doi:10.1145/1518701.1518946

[37] William F. Sharkey and Laura Stafford. 1990. Turn-taking resources employed by congenitally blind conversers. Communication Studies 41, 2 (1990), 161–182. doi:10.1080/10510979009368299

[38] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Terminal Feedback Outperforms Concurrent Visual, Auditory, and Haptic Feedback in Learning a Complex Rowing-Type Task. Journal of Motor Behavior 45, 6 (2013),

455–472. doi:10.1080/00222895.2013.826169

[39] Steven Strachan, Parisa Eslambolchilar, Roderick Murray-Smith, Stephen Hughes, and Sile O'Modhrain. 2005. GpsTunes: Controlling Navigation via Audio Feedback. In Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services. Association for Computing Machinery, New York, NY, USA, 275–278. doi:10.1145/1085777.1085831

[40] Liyuan Sun, Yue Zheng, Qinchuan Jiang, Xiangyu Shi, Mario Romero, and Renan Guarese. 2025. Sound-Assisted Climbing: Exploring Bouldering Performance through Auditory Guidance. In Proceedings of the First Annual Conference on Human-Computer Interaction and Sports. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3749385.3749402

[41] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (Boulder, Colorado, USA) (ASSETS '12). Association for Computing Machinery, New York, NY, USA, 95–102. doi:10.1145/2384916.2384934

[42] Samuel White, Hanjie Ji, and Jeffrey P. Bigham. 2010. EasySnap: Real-Time Audio Feedback for Blind Photography. In Adjunct Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, NY, USA, 409–410. doi:10.1145/1866218.1866244

[43] F L Wightman and D J Kistler. 1999. Resolution of front-back ambiguity in spatial hearing by listener and source movement. J. Acoust. Soc. Am. 105, 5 (May 1999), 2841–2853.

[44] Graham Wilson and Stephen A. Brewster. 2016. Using Dynamic Audio Feedback to Support Peripersonal Reaching in Young Visually Impaired People. Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (2016), 209–218. doi:10.1145/2982142.2982160

[45] Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. 2007. SWAN: System for Wearable Audio Navigation. In 2007 11th IEEE International Symposium on Wearable Computers. 91–98. doi:10.1109/ISWC.2007.4373786

[46] Chengshuo Xia, Tian Min, and Yuta Sugiura. 2024. AudioMove: Applying the Spatial Audio to Multi-Directional Limb Exercise Guidance. Proceedings of the ACM on Human-Computer Interaction 8, MHCI, Article 244 (2024). doi:10.1145/3676489

A Target-Set Composition

Each participant completed all nine target sets. The assignment of G1–G9 to the nine method-by-difficulty conditions (3 methods × 3 levels) was counterbalanced across participants. Table 3 lists the five target positions in their presentation order for each set.

Table 3: Target positions and presentation order for the nine target sets. Primed labels denote inner targets positioned at 50% of the participant's calibrated arm reach, whereas unprimed labels denote outer targets positioned at 100% of the calibrated arm reach. Targets A and E are shared endpoints across the plane-angle conditions.

Set	Target sequence in presentation order								
G1	A	→	$B_{90°}$	→	$C'_{45°}$	→	$D_{0°}$	→	$B'_{-45°}$
G2	$C_{90°}$	→	$D'_{45°}$	→	$B_{0°}$	→	$C'_{-45°}$	→	E
G3	$B'_{45°}$	→	$C_{0°}$	→	$D'_{-45°}$	→	A'	→	$D_{90°}$
G4	$D'_{0°}$	→	$B_{-45°}$	→	E'	→	$B'_{90°}$	→	$C_{45°}$
G5	$C_{-45°}$	→	A	→	$C'_{90°}$	→	$D_{45°}$	→	$B'_{0°}$
G6	E	→	$B_{45°}$	→	$D'_{90°}$	→	$D_{-45°}$	→	$C'_{0°}$
G7	$B_{90°}$	→	$C'_{0°}$	→	A'	→	$B'_{-45°}$	→	$D_{45°}$
G8	$B'_{45°}$	→	E'	→	$C_{-45°}$	→	$D_{0°}$	→	$C'_{90°}$
G9	$B'_{0°}$	→	$D_{90°}$	→	$D'_{-45°}$	→	A	→	$C_{45°}$