Visual Informatics 10 (2026) 100309

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Research article

A study on attention guidance in AR for target localization

Fuqi Xie [a], Luyan Jiang [a], Siqi Xie [a], Qianru Liu [a], Hai-Ning Liang [b], Lingyun Yu [a,*]

[a] *Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China*
[b] *The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China*

ARTICLE INFO

Article history:
Received 29 July 2025
Received in revised form 10 February 2026
Accepted 16 February 2026
Available online 23 February 2026

Keywords:
Attention guidance
Visual cues
Auditory cues
Haptic cues
Augmented reality

ABSTRACT

This work presents a design framework for AR attention-guiding cues in target localization tasks. Localization tasks involve identifying a target, moving toward it, and monitoring changes in the surrounding environment. Maintaining awareness of target-related information is challenging in real-world settings, where clutter, occlusion, and competing perceptual demands are present. AR provides a suitable environment for attention guidance because virtual cues can align with real-world targets and use sensory channels appropriate for the task and context. To understand how attention-guidance cues should express target-related information and remain noticeable in real-world localization settings, we conducted an ideation workshop with experts and novice users. Integrating perceptual and attentional perspectives, we analyzed ideas from the workshop and derived a design framework that structures attention-guidance cue design through four dimensions: What, How, Where, and When. We illustrate the framework through representative localization scenarios, including target finding, route following, and progress tracking. We conducted a two-stage evaluation to examine how well the framework supports real-world design and whether cues created with it help users perform localization tasks effectively.

1. Introduction

Target localization is a fundamental activity in many real-world contexts, where users need to identify a target, move toward it, and monitor how it changes within its surrounding environment. These tasks require continuous awareness of spatial position, spatial relations, and temporal updates. Maintaining such awareness becomes challenging in complex environments: visual clutter, occlusion, competing signals, and divided attention can hinder users' ability to detect or track a target reliably. As a result, localization tasks often require attention guidance (Goldstein, 2001; Wolfe, 2020), which can be provided by external cues that help users notice, interpret, and maintain focus on task-relevant information.

Augmented Reality (AR) offers a particularly suitable environment for attention guidance during localization. Unlike mobile devices and 2D displays, AR can spatially align visual cues with real-world objects, enabling guidance to appear at the location where perception and action occur (Billinghurst and Kato, 2002; Henderson and Feiner, 2011b). This spatial alignment reduces attentional shift costs, mitigates issues caused by clutter or occlusion, and allows cues to remain perceptually connected to real-world context. Importantly, AR supports multiple sensory channels, such as visual, auditory, and haptic displays, enabling cues to be provided through the most appropriate modality for a given context. However, existing research provides limited support for designing attention-guidance cues systematically across different modalities. Although prior work has explored the use of multimodal cues, it lacks a systematic framework for principled reasoning about cue design across visual, auditory, and haptic modalities, while accounting for the demands of different tasks, scenarios, and real-world contexts.

To address this gap and inform the development of a framework, we conducted an ideation workshop involving both visualization experts and novice users. The workshop explored how participants expect target-related information to be expressed and how cues should be designed for different localization tasks and usage contexts to remain noticeable under spatial, temporal, and attentional constraints. These discussions surfaced design considerations that are not explicitly addressed by existing frameworks, such as preferred spatial alignments, timing patterns, and channel usage. Based on these insights, we derived a structured design framework that captures recurring design considerations through four fundamental dimensions: **What** is used to convey target-related information (modality), **How** it is presented (presentation), **Where** it is positioned (spatial alignment), and **When** it updates (temporal behavior). Within the What dimension, we organize modality-specific representational

* Corresponding author.
E-mail addresses: Fuqi.Xie20@student.xjtlu.edu.cn (F. Xie), haining.liang@hkust-gz.edu.cn (H.-N. Liang), Lingyun.Yu@xjtlu.edu.cn (L. Yu).

https://doi.org/10.1016/j.visinf.2026.100309

forms for attention guidance across visual, auditory, and haptic modalities through a design space. We illustrate how our framework supports design reasoning using three representative localization scenarios: target finding, route following, and progress tracking, reflecting core tasks of localization related to detecting, approaching, and monitoring targets. These scenarios demonstrate the potential of the framework to extend to broader, more complex, or multitasking AR contexts.

We further demonstrate how the design framework can guide the creation of attention cues across these scenarios to support users in noticing and interpreting target information. In addition, we conducted a user study to examine how users apply the framework in practice, evaluate their performance with the resulting cue designs, and gather their perspectives and recommendations on cue effectiveness. To our knowledge, this is the first framework that integrates modalities, presentations, spatial alignment and updating behavior for attention-guidance cues in AR, offering a unifying structure that goes beyond existing modality-specific taxonomies. In summary, our work contributes:

- a structured design framework that supports principled reasoning about attention-guidance cue design across modalities, presentation, spatial positioning, and temporal updating;
- a design space of attention cues across visual, auditory, and haptic modalities;
- illustrative localization scenarios that show how the design framework can be applied across different task demands and practical AR conditions;
- design guidelines that specify key requirements for individual cues, including task-relevant information and spatial or temporal support.

2. Related work

We examine prior work on localization tasks and their associated perceptual difficulties, attention-guiding cues for localization in AR, and the perceptual mechanisms of attention guidance cues.

2.1. Target localization

Localization involves three foundation tasks: identifying the target (Lin et al., 2021), moving toward the target (Ogawa et al., 2023), and monitoring how the target changes over time (Shi et al., 2013). These three fundamental tasks can be combined to create more advanced localization scenarios. Bowman and Hodges (1999) demonstrated that, in navigation and search contexts, locating a target and then moving toward it typically unfold as two successive stages. Pylyshyn and Storm (1988) showed that identifying multiple targets and tracking the changes operate as concurrent tasks in dynamic visual scenarios. These combinations create different demands on how users access target-related information during localization. Each basic task draws on different aspects of target-related information, which can be affected by the surrounding context in different ways, leading to distinct perceptual challenges for users:

- When **identifying a target**, users rely on perceiving where the target is and distinguishing it from surrounding objects (Seeliger et al., 2021; Sutton et al., 2024; Zhang et al., 2023). The complexity of the surrounding environment has a substantial impact on localization performance. For example, Wolfe (2020) demonstrated that scene clutter increases search time by reducing the separability of relevant and irrelevant features, making it hard for users to visually isolate the target. To address this issue, Lin et al. (2021) introduced three novel designs that encode an object's relative height, angular direction, and attribute values. Elmqvist and Tsigas (2008) reported that occlusion further limits target visibility and complicates the confirmation process. To support visual searching without introducing additional occlusion, Lu et al. (2014) developed subtle AR cues that enhance target detection without masking the surrounding scene. Beyond these in-view challenges, out-of-view cases introduce the additional challenge that the target may not be visible at all. Wieland et al. (2022) leveraged 3D visual cues to encode the spatial direction and distance of out-of-view targets. Trepkowski et al. (2022) compared modality-combination cues, including visual–audio, visual–haptic, and audio–haptic pairs. They found that haptic components enabled the fastest responses to out-of-view AR objects during a distraction task.
- When **moving toward a target**, users estimate direction and distance and maintain a stable sense of the target's position relative to themselves (Lepelley et al., 2009; Weber et al., 2011; Dominic and Robb, 2020). However, users struggle to judge egocentric distance and approach direction when key depth cues are limited, including size, perspective, or motion parallax (Cutting and Vishton, 1995). To supplement depth cues, Lee et al. (2022) designed ten visual cues for the navigation task, finding that users preferred cues with clearer spatial anchoring. Besides using visual modality to complement the depth cue, Hwang et al. (2023) compared auditory and vibration haptic cues to offer directional guidance and support redirected walking under visual overload. As users move, Ellis and Menges (1998) found that changes in viewpoint disrupt spatial updating and introduce inconsistencies in perceived direction, which in turn makes it more difficult to keep a stable sense of heading. To address heading stability, Zhao et al. (2023) compared view-aligned and world-fixed visual cues during movement, finding that world-fixed cues lead to more stable heading.
- When **monitoring a target**, users observe how the state of that target evolves over time (Arbeláez et al., 2019; Henderson and Feiner, 2011a; Trepkowski et al., 2022). Raikwar et al. (2024) developed visual cues to support monitoring target states in a dynamic AR cooking environment, demonstrating that overlays placed directly on the target enabled users to detect status changes more quickly while performing ongoing tasks. In addition to visual cues, Wu et al. (2024) incorporated auditory signals to aid in monitoring targets when the visual channel was constrained, and found that temporal sound patterns enhanced users' ability to notice changes during VR search tasks. In addition, Sohlberg et al. (2007) observed that when the visibility of a target is intermittent, users struggle to maintain an accurate mental representation of its current state and must repeatedly reassess it each time the target reappears after being occluded. Motivated by such challenges, Matviienko et al. (2022) developed visual techniques to visualize the motion of approaching vehicles normally hidden behind obstacles.

2.2. Attention guidance cues in AR localization

Extensive work on attention-guidance cue design in AR environments has examined a wide range of sensory modalities, including visual (Lu et al., 2014; Sutton et al., 2022), auditory (Parseihian et al., 2014; Nogalski and Fohl, 2016), and haptic (Tan et al., 2003; Lehtinen et al., 2012) cues, as well as comparative studies that evaluate differences between these modalities (Marquardt et al., 2020; Zhang et al., 2023). However, most cue designs are closely tied to the specific tasks and contextual assumptions under which they were developed, which limits their transferability to new situations. This challenge becomes evident when we look at how the same cue format is reused in different AR tasks. Even when the cue element itself is identical, its placement, alignment, and how it updates can differ between tasks and contexts. For instance, arrows may appear as screen-edge indicators in out-of-view search (Wieland et al., 2022), as screen-fixed or world-fixed annotations in navigation (Dominic and Robb, 2020), or as anchored path guides in tracing tasks (Trepkowski et al., 2022). As a result, such cues often do not generalize well to new tasks or environments, where the underlying task demands or spatial configurations may differ substantially. This limited transferability hinders designers from recognizing or developing effective cues in new AR situations and from extending their findings across different tasks, contexts, or output channels.

Compared with a single-modal cue, multimodal cues demand not only the design of each individual modality but also careful planning of how these modalities interact, complement, and reinforce one another within the specific task context. Prior work has explored visual–auditory (Zhang et al., 2023; Sutton et al., 2024; Gröhn et al., 2005), visual–haptic (Lepelley et al., 2009; Arbeláez et al., 2019; Weber et al., 2011), and auditory–haptic (Trepkowski et al., 2022) combinations to support cases where a single modality cannot provide sufficient spatial, directional, or state information. These studies demonstrate that multimodal cues can be beneficial in specific task settings; however, their reliance on task-specific assumptions, such as how modalities should be aligned or temporally updated, further complicates their transferability to new contexts.

Taken together, prior work shows that both single-modality and multimodal cue designs face transferability challenges. Because most cues are developed and evaluated within narrow task and context boundaries, designers have little guidance on how to adapt existing cues or design insights derived from previous work to new tasks or environments. In this paper, we address this gap by introducing a framework that organizes key design factors underlying the effectiveness of attention-guidance cues, enabling systematic reasoning and adaptation across tasks and contexts.

2.3. Perception mechanisms of attention guidance cue

Prior work on visual attention (Itti and Koch, 2001; Desimone and Duncan, 1995) showed that how easily a visual signal is noticed is determined both by its distinctiveness and by the competing signals in the surrounding environment. Signals that exhibit greater contrast with their surroundings are more likely to capture attention (Treisman and Gelade, 1980; Wolfe, 1994), whereas signals that resemble nearby elements are more easily missed (Duncan and Humphreys, 1989). Sawaki and Luck (2010) and Gaspelin and Luck (2018)'s research on distractor suppression further demonstrated that reducing the influence of competing signals increases the likelihood that a task-relevant visual stimulus will be selected. In addition, Müller and Rabbitt (1989) showed that certain attention-guidance cues can redirect attention toward a location even before the corresponding visual stimulus becomes fully perceptible. These findings clarify how visual distinctiveness, reduced competition, and attention-guidance cues jointly influence users' ability to notice task-relevant visual stimuli in complex scenes.

Beyond feature-level properties, spatial attention research (Posner, 1980) shows that the spatial correspondence between a cue and the location it indicates strongly influences the efficiency of attentional orienting. When a cue's spatial position or structural layout aligns well with the referenced location, Eriksen and St. James (1986) found attention can be redirected with less effort and at a lower shift cost. In contrast, Klein (2000) reported that cues that are spatially distant, misaligned, or weakly related to the indicated location increase spatial uncertainty and make attentional orienting slower and more effortful. Posner et al. (1980) further demonstrated that clearer cue–location correspondence supports more efficient orienting, whereas ambiguous or conflicting spatial signals slow responses and reduce the reliability of spatial judgments. Together, these findings clarify how correspondence, alignment, and spatial proximity shape the efficiency of spatial attentional allocation in complex environments.

A further tendency involves detecting changes over time, with temporal attention research showing that users' ability to notice such updates depends on how attention is prepared and sustained. Nobre (2001) found that when the timing of an update is predictable, attention can be allocated in advance, enabling faster and more reliable detection. When updates occur unexpectedly, however, Coull and Nobre (1998) found that attention needs to shift without preparation, slowing detection and increasing uncertainty about whether a change has occurred. Warm et al. (2008) further showed that sensitivity to ongoing updates decreases as attention is maintained over prolonged periods, making gradual or subtle changes easier to miss. Together, these findings clarify how temporal predictability and sustained attention influence effectiveness in perceiving time-varying information.

In conclusion, these visual, spatial, and temporal attentional tendencies encourage us to explore additional ways in which attention-guiding cues can be perceived and processed beyond their modality. These viewpoints likewise underscore the need for a more systematic approach to reasoning about cue design across varying task goals and contextual conditions.

3. Ideation workshop

We conducted an ideation workshop to capture how participants intuitively think about and design attention-guidance cues across different localization situations. These insights, interpreted alongside relevant perceptual and attentional theories, informed our design framework for attention-guidance cues in AR.

3.1. Ideation workshop

Participants. Following prior ideation workshops in user-inspired visualization design (He et al., 2024), we invited 18 participants (8 males and 10 females) from a local university. The group comprised 15 master students in Human-Computer Interaction Program with prior coursework related to AR and interaction design. In addition, one professor and two senior Ph.D. students with research backgrounds in visualization and HCI joined the workshop as participants and contributed domain expertise on spatial tasks and interaction design. All participants were briefed on the localization scenarios and demonstrated understanding of the task requirements.

Setup and Materials. We selected three localization tasks that require varying levels of attention guidance:

- *Identification (identifying targets)*: recognizing specific targets within a given context.
- *Navigation (following routes)*: identifying target locations, following the path, and making necessary adjustments.
- *Temporal Recognition (tracking progress)*: monitoring progress to ensure actions align with task requirements.

These tasks differ in their attention guidance requirements. For instance, navigation tasks typically require continuous, real-time guidance, whereas identification and temporal recognition tasks often rely on discrete or event-triggered cues. To support more focused ideation and enable structured analysis, we organized these task types into seven example scenarios, each designed to highlight variations in target position, route properties, or contextual information (e.g., time remaining to reach a target).

- Target is near the user and directly visible.
- Target is distant but still directly visible.

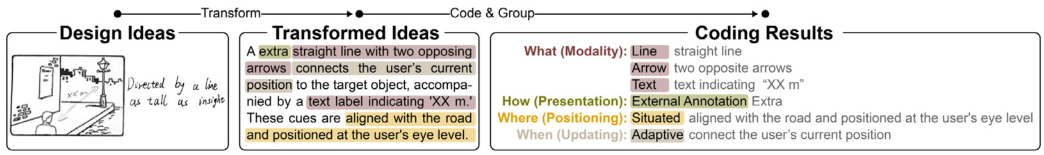

Fig. 1. These panels illustrate three key stages in the workshop analysis process. The first shows an example of an original idea, the second shows a text-based version with highlighted design choices, and the third shows how the extracted elements were organized under the four design dimensions.

- Target is near but not immediately visible, requiring the user to adjust their viewpoint.
- The object is distant and not immediately visible, necessitating significant navigation to locate it.
- Target location and route are straight ahead: the path to the object is direct and without turns.
- Target location is not straight ahead, requiring navigation.
- Remaining time for task completion needs to be displayed.

The workshop was conducted in a classroom setting. Each scenario was detailed on a single A4 sheet, with three columns provided for participants to provide their suggestions for visual cues, auditory cues, and haptic cues to attract attention to target objects, paths, and time. Participants were also encouraged to suggest alternative cues if they had additional ideas.

Procedure. The workshop consisted of two parts: an introduction to attention cues (15 min) and individual brainstorming (65 min across three tasks). At the beginning, we explained the workshop goals and task requirements, ensuring that all participants understood the localization scenarios. Participants were then asked to independently brainstorm and sketch suitable attention cues based on the scenarios and task prompts provided in the task sheets. The scenarios included, for example, designing cues to guide attention toward out-of-view indoor objects at varying distances. To help participants understand the setup and the expected response format, we provided example sketches (shown in Fig. A.8 in Appendix A). These examples served only to illustrate the submission format and were deliberately neutral to avoid suggesting any specific design direction. During the brainstorming stage, we interacted with participants, answered clarification questions, and noted emerging points of interest relevant to future design considerations. After all tasks were completed, we collected the completed sheets, which contained participants' cue designs, written explanations, and demographic information. The study protocol was reviewed and approved by the University Research Ethics Review Panel (No. ER-SAT-0010000120520230407163156).

3.2. Cue elements extraction

Inspired by Tong et al. (2023), we analyzed the ideas collected from the workshop (Fig. 1) to identify the key concerns that guided how participants designed attention-guidance cues for AR localization tasks under contextual constraints. Three co-authors independently translated all 342 written and sketched designs into text descriptions and extracted the cue-related elements referenced in each idea. These elements were then compared, merged, and refined through collaborative discussion. During this process, we observed that participants consistently emphasized several categories of design considerations, including the choice of modality and form, the manner in which cues should be presented, the spatial positioning of cues relative to targets, and how cues should dynamically adapt as users move. These recurring concerns informed the four initial dimensions in our framework: What (Modality), How (Presentation), Where (Positioning), and When (Updating). In the next section, we introduce these four dimensions in detail and show how the identified design elements are organized within them.

4. Design framework

We proposed a design framework for attention-guidance cues in AR localization tasks. In this section, we detail the framework with four foundational dimensions: What (Modality), How (Presentation), Where (Positioning), and When (Updating). Each dimension captures a key aspect of attention cue design.

- **Modality** specifies the type of cue and representational form used to encode target-related information.
- **Presentation** describes how cues are visually integrated into the environment.
- **Positioning** defines the spatial relationship between a cue and its target.
- **Updating** indicates how cues adapt over time in response to user movement or task progress.

4.1. What (modality)

This dimension organizes the choice of modalities and representational forms for encoding target-related information with a design space (shown in Fig. 2A). Given the practical constraints and current device capabilities of commercial AR platforms, we focus on three main modalities: visual, auditory, and vibration-based haptic cues. Design choices in this dimension help determine what information the cue conveys and influence how noticeable the cue remains in real-world AR environments.

Forms that are anchored in space and exhibit a clear directional tendency can convey spatial information, which typically involves a target's location and the direction needed to reach it. Forms that support smooth, continuous variation can convey continuous-valued information such as distance or other values that vary over a range. Forms that clearly separate options can convey categorical information, such as distinguishing between multiple targets. When categories are ordered, forms that present successive states distinctly can make their transitions perceptible.

Choices in this dimension can also affect cue noticeability both within a single modality and when guidance is coordinated across multiple modalities. One source of distinctiveness comes from **contrast-based** signals that diverge from the dominant patterns present in the environment (Duncan and Humphreys, 1989; Itti and Koch, 2001). Another source comes from **novelty-based** signals that introduce forms the environment does not naturally contain (Posner et al., 1980; Coull and Nobre, 1998). Both sources help cues remain noticeable even in cluttered or changing scenes. Beyond single-modality design, distributing cues across multiple sensory channels helps maintain access to the relevant information when one channel becomes overloaded or less accessible (Desimone and Duncan, 1995). Such multisensory redundancy can reduce attentional effort and stabilize cue noticeability in complex environments (Nobre, 2001; Cantarella et al., 2024).

Visual Modality. Visual modality offers flexible representational forms and supports intuitive perception of spatial relations between cues and targets. We identified 9 visual forms that fell into two main design approaches. The first approach enhances the salience by adjusting existing visual content in the

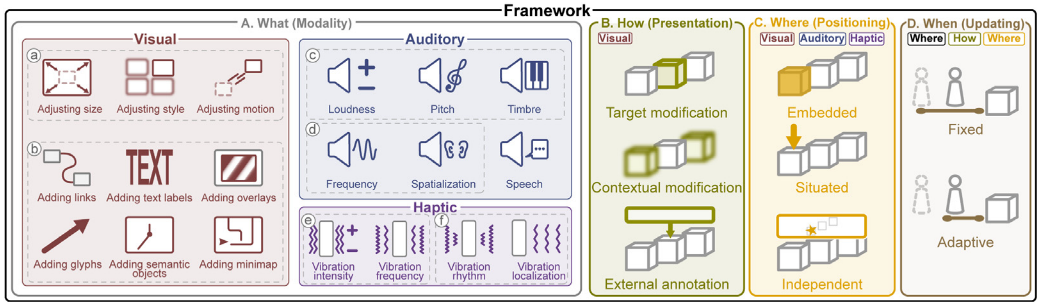

Fig. 2. The framework of cue design is structured around four dimensions for attention guidance in AR localization tasks: What (Modality), How (Presentation), Where (Positioning), and When (Updating). Within the framework, the What dimension organizes attention-guidance cues across three modalities in AR through a design space. Correspondingly, applicable ranges for modalities are marked on the How, Where, and When dimensions: How applies only to visual cues, whereas Where and When apply to all modalities.

environment, as illustrated in Fig. 2A(a). By amplifying differences from surrounding visuals, this approach aligns well with contrast-based distinctiveness. The second approach introduces new visual elements, as shown in Fig. 2A(b). These forms can support novelty-based distinctiveness and help the cues stand out in the scene.

Adjusting size. This technique captures the user's attention by altering the geometric properties of the target object, such as making it visually longer or wider. The resulting deformation makes the target appear distinctly different from its original shape, enhancing its salience within the user's view.

Adjusting style. This technique enhances the prominence of a target object by altering its visual attributes or modifying the surrounding environment to make it stand out. Style adjustments may include changes to texture, color, and transparency.

Adjusting motion. This technique directs user attention by introducing dynamic visual effects to target objects, creating a contrast with the static context. For instance, implementing flashing contours around an object or using a pulsating line to guide the user from their current position to a target.

Adding overlays. This technique enhances the prominence of target objects by superimposing additional visual elements over them. Adjustments to the overlay's color and size are made to ensure clear visibility and effective differentiation from the surrounding environment.

Adding links. This technique directs user attention by visually connecting the observer's position to the target's location using lines. The visual properties of these lines, such as color, thickness, and style (e. g., dashed or solid), can be adjusted to convey additional information.

Adding text labels. This technique enhances user attention by attaching meaningful labels to target objects, providing extra information like distances, time estimates, or customized descriptions. These labels offer additional context, improving clarity and user understanding.

Adding glyphs. This technique highlights target objects by placing icons or symbols near them. These glyphs can range from simple shapes like arrows or circles to more detailed and semantically meaningful symbols, depending on the context and design requirements.

Adding semantic objects. This technique enhances user experience by integrating context-specific, recognizable objects into the scene to convey information intuitively. Unlike abstract symbols or glyphs, these semantic objects are diegetic elements that blend seamlessly into environment, reducing cognitive load and fostering engaging interaction (Norouzi et al., 2021).

Adding minimap. This technique enhances user navigation by using a minimap, a compact, real-time representation of the environment, into the interface. The minimap shows both the user's and the target's locations, along with their orientations, on either 2D or 3D maps. However, interpreting minimaps, particularly in 3D, may require spatial reasoning and map-reading proficiency, which could increase cognitive load for some users.

Auditory Modality. Auditory modality convey information through temporal, directional, and semantic variations. We identified 6 primary auditory forms, grouped into two patterns: continuous sounds and temporal sequences. Continuous sounds are sustained tones that vary in loudness, pitch, and timbre (Fig. 2A(c)). Temporal sequences consist of discrete tones arranged over time and can vary in repetition pattern and channel (Fig. 2A(d)).

Loudness. Variations in loudness can draw user attention and emphasize urgency. A gradual increase in volume may suggest rising importance, while sudden changes can highlight events that require immediate response.

Pitch. Pitch variations can effectively attract attention and convey information. Gradual changes (e. g., ascending or descending tones) can guide visual-spatial attention, while sudden shifts in pitch signal urgency and prompt immediate action.

Timbre. Timbre, often referred to "tone color", provides guidance through unique sounds produced by various instruments, allowing these sounds to convey categorical information. For instance, distinct timbres can represent different categories or groups, enabling users to quickly associate a sound with a specific context or meaning.

Frequency. Frequency captures the user's attention by adjusting the repetition of sound per unit of time. The number of repetitions depends on the duration of each sound unit and the interval between them. For instance, shorter sound unit durations and intervals result in higher frequencies, indicating potential changes in the context.

Spatialization. Spatialization guides attention by positioning sound in 3D space. This is achieved through techniques such as stereo panning and binaural rendering, which create directional effects based on interaural differences between two ears. These effects help users localize sounds in space and support orientation and decision-making.

Speech. Incorporating speech offers a direct and efficient means of capturing users' attention by delivering detailed semantic information and clear instructions.

Haptic Modality. Haptic modality provides a bodily anchored channel that can express information through tactile variations. We identified 4 vibration forms in our design space, grouped into two patterns: continuous vibrations and discrete vibration pulses. Continuous vibrations are sustained signals that can vary in intensity and frequency (Fig. 2A(e)). Discrete vibrations consist of individual pulses arranged over time and can vary in rhythm and starting location (Fig. 2A(f)).

Vibration intensity. Attention is directed by delivering vibrations to specific body locations. When multiple actuators are used on different body parts, users can perceive the vibration source, allowing for effective direction judgment and enhanced spatial awareness.

Vibration frequency. User attention is guided through changes in vibration speed, determined by the number of cycles per unit time. Higher frequencies produce faster-perceived vibrations, which are effective for signaling critical information, such as approaching a target location.

Vibration rhythm. Users are guided through distinct vibration patterns, created by varying the duration and intervals between pulses. These rhythms are easily distinguishable, allowing for the encoding of discrete information.

Vibration localization. Attention is guided by delivering vibrations to specific body locations. With multiple actuators on different body parts, users can distinguish the vibration source, supporting directional judgment and spatial awareness. Note that this approach may require hardware beyond standard commercial AR devices.

4.2. How (presentation)

The How (Presentation) dimension (as illustrated in Fig. 2B) describes how the visual forms can be effectively integrated into the given context so that users' attention is guided toward targets and related information. Different presentation strategies can change visual salience (Treisman and Gelade, 1980; Duncan and Humphreys, 1989) and produce distinct attentional and cognitive effects (Posner, 1980; Müller and Rabbitt, 1989), even when they are applied to the same visual form defined in the What dimension. In our design space, we define this dimension specifically for the visual modality.

Target modification. This technique enhances target visibility by modifying their appearance through adjustments to surface (Adjusting style), geometry (Adjusting size), or motion (Adjusting motion), while minimizing interference with surrounding objects. Motion-based highlighting is particularly effective in capturing attention in cluttered environments due to its dynamic nature.

Context adjustment. This technique enhances target visibility by modifying the surrounding environment while keeping the target unchanged. Adjusting contextual elements, such as surface styles (Adjusting style), increases visual contrast, making targets more distinguishable. This approach is particularly effective when targets are within the user's field of view (FOV) and experience minimal occlusion. In cluttered scenes, it helps reduce distractions, ensuring targets stand out more clearly.

External annotation. This technique guides users' attention by introducing new visual elements, such as glyphs or links (e. g., Adding glyphs or Adding links), into the scene to connect external information with targets. While effective, excessive use of these elements can increase cognitive load and reduce readability due to potential occlusion or clutter. Careful design is essential to balance informativeness with clarity.

4.3. Where (positioning)

The Where (Positioning) dimension (as illustrated in Fig. 2C) describes how attention-guidance cues are spatially placed relative to their targets. The spatial proximity between cues and targets can affect the attentional shift required to locate the relevant region (Posner, 1980; Eriksen and St. James, 1986). This dimension applies to all modalities defined in the What dimension, as spatial placement can be designed consistently across modalities under AR's world-aligned spatial reference.

Embedded. Embedded positioning involves aligning cues directly with their targets, making them appear as integral parts of the targets. This alignment ensures a strong visual connection, enhancing the effectiveness of cues. For visual cues, embedding involves modifying the target's appearance, such as Adjusting size, Adding overlays, or text labels, to enhance visibility, which requires minimal occlusion to maintain clarity. In auditory cues, spatial audio positions the source precisely at the target's location, improving alignment and attention guidance.

Situated. Situated positioning involves placing attention cues near, but not directly on, their targets, maintaining a close association. For visual cues, elements are strategically positioned within the user's FOV to enhance accessibility while preserving the relative spatial relationship to the targets. For auditory and haptic cues, spatial parameters such as distance and direction are taken into consideration to reflect the spatial relationship between the user and the targets.

Independent. Independent positioning places attention cues based on user location or FOV, rather than the target's position, to prevent interference with the original scene. For visual cues, independent elements are consistently displayed above real-world content, avoiding overlap with scene. For auditory and haptic cues, independent cues are designed without encoding spatial information, ensuring perception is unaffected by user-target positioning. However, this technique offers the weakest connection to targets compared to embedded or situated positioning. When targets are fully occluded, users may need more time to locate them, potentially impacting response efficiency.

4.4. When (updating)

The When (Updating) dimension (as illustrated in Fig. 2D) describes how attention-guidance cues relate to changes in the user's relationship to the target. Choices relate to how different user-target conditions draw on selective (Desimone and Duncan, 1995) and sustained attention (Warm et al., 2008) during localization tasks. This dimension operates across the other three dimensions in the design framework, as AR's world-aligned spatial reference provides real-time contextual information needed to organize cues consistently across other dimensions.

Fixed. Fixed cues are attention cues that remain constant once presented, unaffected by the user's actions or state.

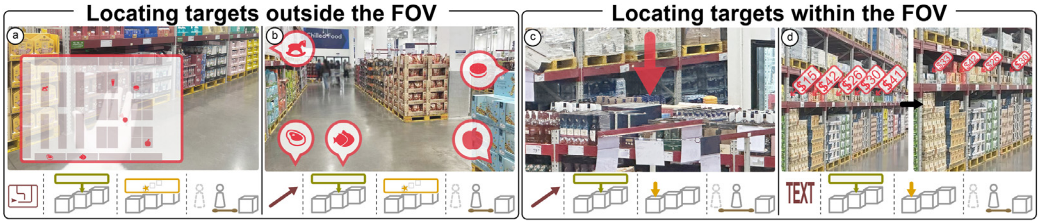

Fig. 3. In the supermarket scenario, we illustrate how different design elements support users in identifying targets. When a target lies outside the user's FOV, (a) a minimap provides global context and (b) icons indicate the target's orientation. When the target is within the FOV, arrows (c) and text labels (d) offer precise and context-sensitive cues that guide attention.

They provide a stable reference point, reducing the risk of misinterpretation due to changing conditions. This stability makes fixed cues particularly suitable for scenarios where targets do not require real-time adjustments or dynamic guidance.

Adaptive. Adaptive cues dynamically adjust to reflect the user's current state relative to targets, utilizing parameters such as spatial position and input state to provide real-time feedback. This ensures continuous and effective guidance as the user moves or interacts within the environment. By updating at moments when attention needs to be redirected, adaptive cues can temporarily restore temporal salience and help users refocus on the relevant spatial region. To minimize disruptions to contextual awareness, the intensity of embedded visual cues, as well as the size or prominence of situated and independent cues, can gradually diminish as the user approaches the target.

In summary, these four dimensions provide a structured way to reason about attention cue design. They clarify how information should be encoded, expressed, situated, and adapted as users perform target localization tasks with AR.

5. Working scenarios

We now illustrate how the framework functions in practice. We discuss design choices across the four dimensions for three common AR localization tasks: **Identifying Targets** (Schwerdtfeger and Klinker, 2008), **Following Routes** (Neges et al., 2017), and **Tracking Process** (Shi et al., 2013).

5.1. Task one: identifying targets

We use a supermarket scenario (Fig. 3) to demonstrate how design elements assist users in identifying targets. Two situations are considered based on the target's spatial setting: within the FOV and outside it.

Context and Design Considerations. This scenario focuses on directing users' attention to specific targets (e.g., objects, locations, or areas in the physical environment). Design considerations depend on factors such as the number, size, and visibility of the targets within the given context:

- *Target number.* The number of targets influences the choice of modalities and forms for attention guidance. For a few easily distinguishable targets, simple visual forms like adjusting size and style, adding overlays or text labels, are effective. These methods enhance target visibility without introducing unnecessary complexity when multiple targets are distributed across a wide area, especially beyond users' FOV. In such cases, providing an overview, such as a minimap (Fig. 3(a)), becomes essential for conveying the broader spatial layout of the targets. In addition, when many targets are involved, auditory and haptic forms may become less intuitive due to additional effort required to mentally map signals to specific spatial positions.
- *Target size.* The size of a target influences which visual forms are appropriate, as size determines the target's baseline visibility. This influence is most noticeable when the target is within the FOV. Large targets are naturally noticeable and often require minimal enhancement. Most visual forms, such as overlays or text labels (Fig. 3(d)), are effective in drawing attention to these targets. Small targets may not stand out as readily. Visual forms like adjusting size and motion can help emphasize the target and distinguish it from the surrounding context. External indications, such as an explicit arrow (Fig. 3(c)), are useful. In contrast, target size plays little role for auditory or haptic forms, as these modalities convey information independently of the target's visual visibility.
- *Target visibility.* The visibility of targets affects which cue placements are feasible and which modalities and forms are suitable. When targets are clearly visible within the FOV, embedded and situated visual cues like adjusting size, style and adding overlays on the targets work well. For occluded targets, situated cues, such as adding glyphs and links, help direct attention. Auditory and haptic cues can also be situated near occluded target, but the spatial resolution users perceive from many of these forms is coarse. This limits how precisely they can indicate the occluded targets. For targets outside the FOV, independent cues are more effective. A minimap can offer an overview of target distribution and users' location and direction, while directional cues like glyphs (Fig. 3(b)), loudness, stereo channel and localization support direction finding.

5.2. Task two: following routes

We present a navigation task in a park (Fig. 4) to demonstrate how guidance cues are designed to support wayfinding.

Context and Design Considerations. This scenario aims to guide users to the destinations by following routes indicated by cues. Design considerations can depend on both the route characteristics (e.g., route landmarks and optional routes) and task-related situational demands (e.g., unforeseen issues):

- *Route landmarks.* Route landmarks in physical navigation can serve as visual anchors that break down long routes into manageable segments and influence how attention-guiding cues are updated. When clear landmarks are available, attention-guiding cues can dynamically adapt to user movement. For example, through dynamic text or speech cues (Fig. 4(a)) to guide users step by step toward the next landmark. In contrast, when landmarks are lacking, cues can offer continuous guidance along the entire route as users lack the intermediate reference points needed for orientation. This can be achieved using a fixed mode, providing consistent directional cues from origin to destination to maintain a stable navigational reference.
- *Optional routes.* The presence of optional routes affects the type and amount of information that cues are expected to

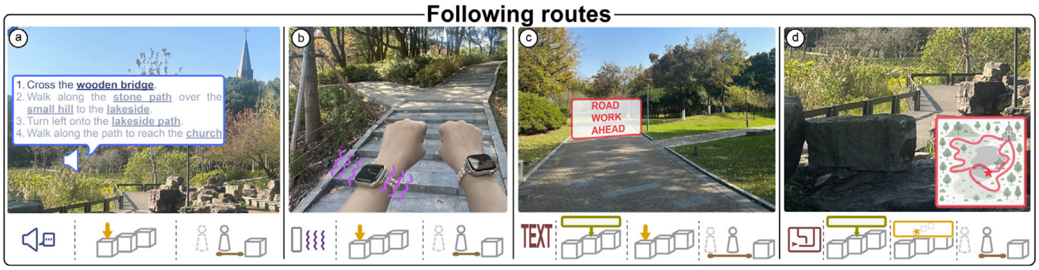

Fig. 4. Following routes in a park scenario. We illustrate four types of guidance cues designed to support wayfinding along a route: (a) speech cues that provide step-by-step instructions, (b) vibration localization that conveys directional information through haptic feedback, (c) situated indicators that embed visual cues directly in the environment to highlight upcoming turns or notable events, and (d) a minimap that offers a global overview of the route and assists users in comparing optional paths.

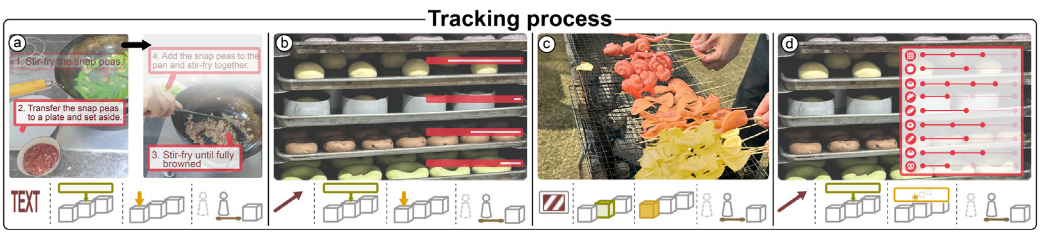

Fig. 5. We illustrate four types of design elements that support users in monitoring ongoing tasks over time: (a) text labels that convey detailed instructions of each step, (b) progress bars that visually represent remaining time or completion levels, (c) color overlays that highlight urgent or status-changing events to draw immediate attention, and (d) progress lines that summarize multiple concurrent timelines and allow users to compare the states of parallel tasks.

convey and therefore influences the suitability of modalities and forms. For single-option routes, especially when the path is clear and visually distinguishable, simple directional cues (e. g., arrows, stereo audio, or vibration localization; see Fig. 4(b)) are sufficient to guide users. In contrast, when multiple optional routes exist, additional information such as route length, travel time, and route geometry is introduced for comparison. In such cases, cues benefit from presenting this comparative information more explicitly, such as distances, directions, and regions associated with each route. Visual forms such as added links or a minimap (Fig. 4(d)) can support this comparison.

- *Unforeseen Issues.* In navigation tasks, unforeseen events or specific arrangements can affect how noticeable cues are expected to be, and may therefore influence choices of modality, forms, and cue positioning. Examples include traffic congestion ahead, construction work, or important features such as highway entrances or exits that require special attention. Such situations are often overlooked during normal guidance, increasing the need for cues with greater noticeability. More explicit indications, such as blinking animations, distinct voice messages, or changes in audio timbre, can help capture attention. Additionally, embedded and situated indicators (Fig. 4(c)) are useful for clearly showing the location of these events within the environment, while independent notifications (e. g., floating messages or audio alerts) are helpful when these locations are occluded or outside the user's FOV.

5.3. Task three: tracking process

We use a cooking task (Fig. 5) to demonstrate how the proposed design elements support users in the process-tracking task.

Context and Design Considerations. Different from the previous two scenarios, tracking requires users to not only find specific targets but also track new and evolving information over time. Users need to continuously monitor the status and progression of ongoing activities, which demands sustained attention and temporal awareness, thereby introducing a higher cognitive load. Beyond environmental context, design considerations can also depend on task-related factors, such as the type of progress, the presence of concurrent tasks, and their priority levels:

- *Process content.* Process content refers to types of task-related information, which can influence the suitable modalities and forms for attention guidance. Progress-related quantities, such as remaining time, can be conveyed through visual forms that show the proportion of what is completed and what remains. For example, a diminishing progress bar can intuitively convey remaining time (Fig. 5(b)). For categorical values, such as indicating current stage, simple glyphs like standalone icons are straightforward and semantically meaningful. For detailed instructions, textual displays (Fig. 5(a)) or speech indication can help ensure clarity and precision.
- *Multiple progresses.* When multiple tasks are tracked in parallel, different spatial arrangements of cues can direct attention toward different aspects of progress, such as overall status or object-specific details. Managing several tasks, such as cooking multiple dishes, requires tracking multiple timelines with different priorities and actions, which increases cognitive load. Simple visual forms for each task can be presented in a unified layout to help users quickly assess overall status and compare progress across tasks, such as multi-task timelines where lines represent tasks and nodes mark stages (Fig. 5(d)). Alternatively, presenting cues in a situated or embedded manner near corresponding physical objects (e. g., pots or proofing bread; Fig. 5(b)) can make it easier to associate each progress with its object, thereby reducing mental effort.
- *Priority shifts.* Priority changes can influence how noticeable cues are expected to be, and this noticeability can be supported by design choices about when cues update and which modalities or forms are used. In multitasking scenarios, task priorities

often change dynamically due to evolving conditions or user actions. Some shifts occur naturally, such as water nearing a boil or a timer running out, while others result from user adjustments, like adding ingredients early or skipping steps. The system can reassess priorities in real time and update attention cues accordingly to redirect attention to urgent tasks. Improving contrast between cues and the surrounding environment, such as by adjusting the intensity of visual forms (Fig. 5(c)) or auditory form can further enhance noticeability and help convey urgency, supporting timely user responses.

6. User study

To assess how well our framework supports real-world design and user needs, we conducted a two-stage evaluation. The first stage involved domain stakeholders, individuals whose work requires rapid and accurate target localization, and experienced designers. Here, our goal was to determine whether the framework can operate as a clear and usable communication framework for describing tasks, constraints, and cue designs. The second stage focused on end users: we tested whether cues designed using the four dimensions actually help users accomplish target localization effectively in practice. This combined evaluation allows us to examine both how well the framework guides design and how well the resulting designs perform in realistic localization tasks.

6.1. Stakeholder validation of the design framework

This evaluation includes two scenarios from different domains, each with distinct task demands and working contexts. The first is a library setting, where the system needs to highlight specific targets (books) within a structured indoor environment. The second is a cultural-tourism scenario, where attention-guidance cues are used to direct visitors toward localization tasks in an open outdoor space with richer and more variable contextual conditions. We invited two stakeholders, P1 from the library scenario and P3 from the cultural tourism scenario, and two experienced designers (P2 and P4). We examined how our framework, with its modality design space, facilitated their discussions and guided their proposed cue designs.

Task and Procedure. Each stakeholder-designer pair completed four sequential steps, followed by individual interviews.

- **Introduction.** We began by introducing the purpose of attention-cue design, the design space and the four dimensions of the design framework. Stakeholders and designers asked clarifying questions, and we explained how the framework can support early-stage communication where task requirements and design ideas need to be aligned.
- **Contextual Grounding.** Stakeholders described their working scenario, localization tasks, requirements, and the challenges they encounter. Together, each pair established a shared understanding of the task demands and context. At this stage, we observed whether participants naturally referenced elements of the design framework and to what extent.
- **Communication.** Participants then revisited the scenario using the four-dimension design framework to articulate concrete task needs: what information needs to be conveyed, how cues should be encoded, where cues should appear, and when they should update. They identified key choices in each dimension and explained their rationale. At this stage, we observed how the design framework shaped communication, how frequently each dimension was invoked, and whether the framework enabled more complete reasoning.
- **Idea Proposing and Discussion.** Designers generated candidate cue designs based on the refined requirements. Stakeholders responded to each design, discussing necessary adjustments or constraints. At this stage, we examined whether the framework supported designers in justifying their concepts and whether stakeholders used the dimensions to evaluate completeness or identify missing considerations.

After all steps, we conducted individual interviews. Stakeholders reflected on describing their needs and communicating with designers in both the unstructured and structured phases. Designers reflected on interpreting stakeholder requirements and how the framework influenced their reasoning and collaboration. Across the entire process, we evaluated how effectively the framework, including its modality design space, functioned as a shared language between designers and stakeholders.

Results. Across sessions, stakeholders and designers reported that the framework improved clarity, depth, and mutual understanding during early design communication.

- **Requirements Articulation.** Stakeholders became **more specific and structured when describing their needs**. For example, P1 explained that library users often know which books they want and therefore require direct, unambiguous cues. With the design framework, she could clearly specify where guidance should appear and when it should be shown during the workflow. Similarly, P3 clarified that attention cues in cultural tourism must both attract visitors' attention and guide them toward events or shops. Using the design framework, he was able to provide concrete examples of where guidance should occur along the route and how to draw visitors' attention effectively.
- **Requirements Verification.** Designers likewise reported that the design framework **guided them toward asking more targeted and relevant questions**. For instance, P2 clarified whether targets outside the FOV required guidance, and P4 asked about riverside safety zones, a constraint that would restrict cue placement. The design framework also helped stakeholders articulate design preferences with rationale: P1 preferred target-modification cues over contextual augmentation because books placed close together made contextual cues cluttered, and P3 wanted ground-based route guidance that matched the cultural style of the folk village.
- **Design Alternatives.** Designers found that the design framework helped them **narrow and refine design choices in a scenario-appropriate way**. For example, P2 proposed a two-stage highlighting approach, which was shaped by the low-occlusion and fixed-route nature of the library setting. The approach uses color overlays at a distance and transitions to book-level labels as the user approaches the correct shelf. In contrast, the cultural-tourism scenario prompted more diverse cue combinations to accommodate different visitor abilities and the stylistic constraints of the environment.
- **Shared Understanding.** Finally, the framework **improved shared understanding and communication during co-design**. P1 and P2 quickly agreed that the "Where" dimension was most critical for the library, while P3 said he could easily see how new design proposals corresponded to earlier stated needs. Designers noted that clear stakeholder preferences sometimes narrowed exploration (as in the library case) or encouraged broader ideation (as in the tourism case). Several participants mentioned that, because many visual options existed, simple visual examples could help ensure both sides were referring to the same cue type.

Overall, the evidence indicates that the design framework functions effectively as a shared design language, helping stakeholders articulate needs more fully, prompting designers to ask sharper questions and generate more relevant options, and enabling both sides to reason about localization-cue design in a structured, comprehensive way.

6.2. User study on cue effectiveness

We designed a prototype system integrating selected guidance cues and evaluated their performance across the three localization tasks described in Section 5. In this study, we focused on visual and auditory cues because these modalities are natïvely supported by current AR headsets and represent the most commonly used forms of attentional guidance in practical AR applications. We did not include haptic cues, as haptics typically require external hardware such as gloves or handheld controllers, which would introduce device-specific constraints and variability into the experiment. Since our goal was to assess how well framework-designed cues support localization tasks, rather than to compare different hardware modalities, restricting the study to built-in AR modalities ensured a fair and interpretable evaluation.

Participants. We recruited 15 participants (9 males, 6 females) from the local university, 21–24 years old (M=22.3, SD=1.8). None of them had participated in the ideation workshop. All participants had a background in computer science and were right-handed. They reported regular use of VR/AR devices (at least once a week). They all had normal or corrected-to-normal vision, with no difficulty distinguishing the colors used in our prototype.

Apparatus. The experiment was conducted in a 5 m × 5 m (Fig. D.11 in Appendix D), allowing participants to move freely without obstacles. We used a Microsoft's HoloLens 2 (2,048 × 1,080 pixels per eye, 52° FoV) as the AR device, connected to a PC (Intel Core™ i7, 3.5 GHz, 32 GB RAM, GeForce RTX1080) via holographic remoting. Participants wore AirPods Pro earphones, wirelessly connected to the PC. To minimize distractions, both Active Noise Cancellation and Transparency Mode were disabled. The prototype system was developed using Unity3D (version 2020.3.28f1) and deployed on the HoloLens.

Task and Procedure. We asked participants to complete three tasks with the guidance of cues as fast and accurately as possible. For each task, we leveraged three visual cues and two auditory cues that were the most popular in the ideation workshop.

Task One: Locating Target. Participants were instructed to locate a target positioned outside their initial FOV.

Angle. Two rays visualize direction: one shows the participant's forward direction, and the other points to the target. It updates dynamically as the participant moves.

Arrow. A curved arrow appears 30 cm ahead of the participant, extending 40 cm toward the target. It originates from the participant's forward direction and updates dynamically with their movement.

Compass. The compass displays a yellow star indicating the target direction and a dynamic pointer showing the participant's current facing direction.

Voice message. This cue delivers verbal instructions, such as "turn left/right", to guide the participant until they face the target direction.

Sound. This cue plays a beep sound based on the target's relative position. For example, the left earphone beeps when the target is to the participant's left.

Task Two: Following Path. Participants were instructed to follow a designated route and reach a destination located 5 m away, guided by the warning cue. To prevent reliance on environmental features, the destination was randomly assigned before each trial within a 60 ° angle relative to the participant's current facing direction, while maintaining a fixed distance of 5 m.

Arrow. An arrow links the start and destination.

Color block. A red block appears at a fixed distance ahead. As the participant moves forward, the color gradually shifts to green, turning fully green upon arrival.

Visual metaphor. A metaphorical guide (e. g., a fox) leads the way to the destination. If the participant falls over 2 m behind, it pauses and waits until they catch up.

Voice message. Speech announces remaining distance at one-meter intervals as the participant approaches destination.

Sound. A beep plays through the earphones, increasing in volume as the participant approaches the destination, peaking upon arrival.

Task Three: Observing Progress. Inspired by traffic systems that display remaining time through numeric countdowns or progress bars, this task evaluates whether participants can effectively perceive time progression with the aid of different cues.

Progress Bar. A visual bar represents task progress, with its length decreasing over time to reflect the countdown.

Clock. A clock provides a continuous and intuitive indication of the time remaining.

TEXT *Text.* A text box displays a countdown, showing the remaining time in seconds from the start time down to zero.

Voice message. Voice reminders are given every 5 s, increasing to every second during the final 5 s.

Sound. The sound gradually decreases in volume throughout the countdown, becoming silent at the end.

Design. In the study, participants were first given an overview of the research background and a brief explanation of the procedures. They then wore the HoloLens HMD and earphones, adjusting them for comfort and optimal clarity of the AR content. Next, participants were trained to start and complete a task using voice commands, saying "begin" to start and "finish" when they believed the task was complete.

The participants complete three tasks in a fixed order: *Locating Target*, *Following Path*, and *Observing Progress*. For a given participant with a specific P_{ID}, where ID is unique and $\in [0, 4]$ (the first batch of 5 participants), we counter-balanced the order of cues using a Latin square as "$P_{ID}/5$". We repeated the Latin square order for the second and third batches of 5 participants.

After each task, participants completed a cue-comparison questionnaire and rated workload and fatigue using NASA's Task Load Index (TLX) (Hart, 2006). The entire session lasted approximately 30 min. The study protocol was approved by the University Research Ethics Review Panel (IRB). In summary, we had 15 participants × 3 tasks × 5 cues × 4 repetitions = 900 trails.

Measures and Analysis. To reduce the impact of the learning effect, we excluded the first repetition of each cue. This left 675 trails. Due to limitations of null hypothesis significance testing (NHST) (Baguley, 2009; Cumming, 2014; Dragicevic, 2016; Dragicevic et al., 2014), we state results by estimation techniques with effect sizes and confidence intervals (CIs) recommended by APA (Yao, 2011).

Accuracy and Completion Time. We recorded user performance data, including completion time, accuracy, and movement, for further analysis. Completion time was measured from the moment participants verbally said "start" to when they said "finish".

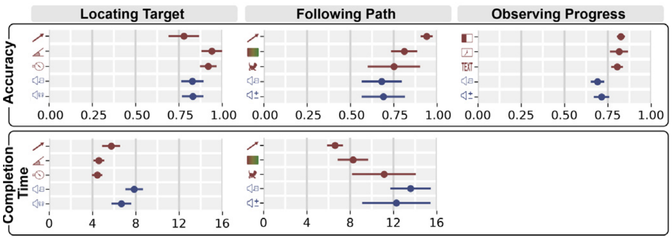

Fig. 6. The accuracy for three tasks (locating the target, following the path and observing progress) and completion time for two tasks (locating the target and following the path). Error bars show 95% CIs.

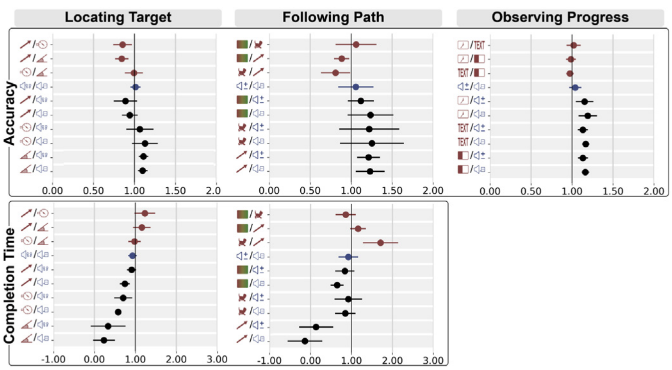

Fig. 7. The overall pairwise ratio of accuracy for three tasks (locating the target, following the path and observing progress) and completion time for two tasks (locating the target and following the path). Error bars show 95% CIs.

Accuracy was determined by comparing participants' final state with the corresponding ground truth values, using task-specific measures including orientation, position, or button press timing. Details of the accuracy calculation are provided in Appendix B. For both accuracy and completion time, we computed the mean and 95% bootstrap CIs ($n = 15$).

Results: Accuracy and Time. The mean accuracy, completion time, and corresponding 95% CIs are shown in Fig. 6, with pairwise comparisons illustrated in Fig. 7. Detailed results are provided in Table C.1 in Appendix C, along with additional statistical analyses in Table C.2 to C.8. In the task of Locating Target, participants took less time when guided by visual cues compared to auditory cues. Among the three visual cues, the Arrow resulted in lower accuracy and longer completion time than both Compass and Angle cues. In the task of *Following Path*, visual cues again led to higher accuracy and faster completion time than auditory cues. The Arrow cue outperformed both Color Block and Visual Metaphor in terms of accuracy and completion time. In the task of *Observing Progress*, all visual cues performed similarly in both accuracy and completion time, and were more accurate than auditory cues. Across all three tasks, the two auditory cues, Sound and Voice message, showed comparable performance in both accuracy and completion time.

Results: User Workload and Preference. We found that participants experienced increased efficiency, reduced frustration, decreased time pressure, and required less mental effort when they completed tasks using visual cues compared to auditory cues (Fig. C.10 in Appendix C), especially in *Following Path* task. Moreover, our findings revealed that in the *Following Path* task, participants reported more favorable subjective experiences when guided by Visual Metaphor than two auditory cues. Interestingly, despite all of them having similar accuracy, participants showed a preference for visual cues over auditory ones. Another interesting finding is that in the *Locating Target* task, although the Compass resulted in higher accuracy compared to the Arrow, participants reported better subjective experiences with the Arrow cue. After analyzing the preference ranking of attention guidance cues for three tasks (Fig. C.9 in Appendix C), we found that participants generally preferred visual cues over auditory ones, which is consistent with subjective experiences. In *Following Path* and *Locating Target* tasks, users both preferred the Arrow cues the most. The progress bar was most favored in the *Observing progress* task, aligning with task performance and subjective experiences.

7. Discussion, limitation, and future work

We further discuss the key insights gained from this work.

7.1. How the framework structures early-stage design reasoning

In the first evaluation, the design framework guided each stakeholder-designer pair to change their communication approach. They started to break down the localization process into concrete tasks and to clarify the contextual constraints for each task. This shift made missing task details and contextual constraints more explicit and helped them supply the needed information with greater clarity. For each task, designers were also motivated to ask contrastive questions to clarify the flexibility of each design dimension. As a result, they reached a clearer shared understanding of the design requirements.

With this clearer shared understanding, our design framework also motivated designers to make the role of each dimension more explicit in their reasoning. Instead of focusing mainly on how a cue expressed target information, they began to consider how each dimension influenced its noticeability and the cognitive effort required to use it. With the task decomposition during communication, they could more easily identify the main contextual factors for cue noticeability and potential cognitive load in each task, such as spatial layout or ongoing activities. This helped them identify the most important dimension for each task and narrow the design choices in the other dimensions. This shift encouraged more flexible combinations and sometimes led to design concepts that had not been considered before.

7.2. Understanding cue effectiveness and design guidelines

Across the three tasks, the more effective cues were those that conveyed both the current state and the progression toward the goal. For example, indicating the orientations of the user and target, their relative locations, or the distance between them. Making these relationships explicit reduced cognitive effort and enabled more direct and confident task responses.

In the *Following Path* task, participants achieved the highest accuracy and the shortest completion time when using the Arrow cue. This cue was attached to the ground and along the entire target route and clearly indicated the start and target positions. This stable spatial reference helped participants determine the correct heading direction and estimate how far they still needed to travel. These results indicate that pairing clear target information with a spatial reference effectively supports path-following tasks, even when the cue remains static during movement.

In the *Locating Target* task, participants achieved high accuracy and short completion times when using either Compass or

Angle. Compass is presented as an independent indicator, whereas Angle adds a spatial anchor along the target direction. Both cues encode the start and end directions and update in real time with the user's rotation. These shared properties helped participants adjust their orientation precisely. The similar performance of the two cues suggests that, in this task, clear directional encoding combined with timely updates can effectively support orientation, while the additional spatial anchor in Angle did not show a clear performance advantage in our study.

In the *Observing Progress* task, participants achieved high accuracy when using either the Progress Bar or the Clock. Both cues encode the remaining time and offer a clear temporal anchor through continuous updates. These properties helped participants judge how much time was left and identify the moment of completion. These results suggest that pairing clear temporal information with timely updates can effectively support progress monitoring tasks.

These findings suggest two guidelines for individual cues.

- **G1.** The What dimension should make the gap between the user's current state and the target state in a form that remains readily perceptible during the task.
- **G2.** At least one of the Where or When dimensions should align with the perceptual demands of the target.

Together, the two guidelines capture the properties of the cues that better supported users in our study, linking clear gap information with task-appropriate spatial or temporal support.

7.3. Limitations

Our design space currently covers modalities and forms that are natively supported on mainstream commercial AR devices. Native support for haptic feedback, however, remains uneven across current headsets, which limits the range of haptic forms that can be directly implemented on commercial AR devices. Despite this limitation, haptic feedback remains an important and expanding part of the What dimension. A growing range of third-party devices, such as haptic gloves (Uchiyama et al., 2008) and vests (Tan et al., 2003), offers more diverse and expressive tactile patterns. Prior work has also explored motion-based tactile feedback (Israr and Poupyrev, 2011), olfactory cues (Dozio et al., 2021), and gustatory stimulation, showing that the haptic modality has broader potential for attention guidance. Future work could extend this dimension by incorporating a wider range of haptic forms as supporting hardware becomes more accessible.

In this study, we concentrated on comprehensive design strategies for localization tasks, including identifying target locations, navigating toward them, and tracking progress. We discussed attention cue designs for scenarios requiring users to notice specific types of information. However, this goal-directed focus should be extended to open-ended exploratory tasks, where users may not know in advance what is relevant. In such contexts, attention cues could help highlight unexpected or significant features. This highlights the need for future investigation into how such cues can be designed to support discovery-driven attention.

Additionally, the scenarios and user study primarily explored visual and auditory cues, as these were consistently rated most effective during the ideation workshop. The user study was conducted in a controlled laboratory environment with minimal distractions, which does not reflect the complexity and unpredictability of real-world settings, such as ambient noise from pedestrians or traffic. Moreover, most participants were young adults with similar educational and technical backgrounds, which limits the diversity of perspectives represented in our sample. The evaluation period was also brief, capturing only immediate task performance. These study settings limit the generalizability of our findings. In future work, we plan to address this by conducting studies in more ecologically valid environments and involving a larger and more diverse participant population, enabling a broader evaluation of the proposed cues.

Finally, while our study focused on short-term task performance, the long-term cognitive effects of repeated cue exposure, such as fatigue, habituation, or learning, remain unclear. In addition, the long-term effects of different guidance modalities on user experience and adaptation were not examined. Future work could investigate these long-term processes through longitudinal or multi-session studies.

8. Conclusion

This work establishes a framework for designing attention-guidance cues in AR, offering a unified way to reason about how cues express target-related information and remain perceptible under diverse task and context conditions. Building on insights from an ideation workshop and attentional theory, we articulated the framework through four dimensions: What, How, Where, and When. We demonstrated how these dimensions apply in representative localization scenarios that illustrate challenges in identifying targets, following routes, and monitoring ongoing processes. We further examined how users interpreted and applied the framework through a complementary interview study, providing insight into how users reason about attention-guidance cues. We also conducted a two-stage evaluation: one stage examined how well the framework supports design reasoning in real-world settings, and the other tested whether cues created with it help users perform localization tasks effectively. In addition, we derived design guidelines for individual cues. Rather than treating cues as modality-specific artifacts, our design framework helps relate cue behavior to the underlying demands of localization tasks and the constraints of context, providing a transferable structure that can be used to design, analyze, and compare cues across AR scenarios. Future work may explore broader sensory channels, more dynamic environments, and long-term use cases to further extend the applicability of this approach.

CRediT authorship contribution statement

Fuqi Xie: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Luyan Jiang:** Validation, Software, Investigation, Formal analysis, Data curation. **Siqi Xie:** Investigation, Formal analysis, Data curation. **Qianru Liu:** Investigation. **Hai-Ning Liang:** Resources. **Lingyun Yu:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Ethical approval

This study was approved by the School Ethics Review Panel of Xi'an Jiaotong-Liverpool University (No. ER-SAT-00100001205202-30407163156).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62272396).

Appendix A. Example materials from the ideation workshop

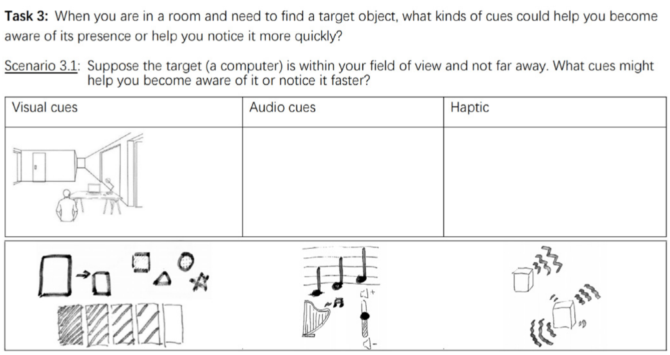

Fig. A.8. An example task sheet (top) outlines three cue modalities (visual, audio, and haptic) providing participants with a structured layout for proposing attention-guiding solutions. The example sketches (bottom) illustrate early conceptual ideas generated by participants during the ideation workshop.

Appendix B. Calculation of accuracy

In *Locating Target* task, the accuracy (FTacc) is calculated based on the participant's start direction $\vec{\mathbf{sd}}$, the participant's end direction $\vec{\mathbf{ed}}$ and ground truth direction $\vec{\mathbf{gd}}$:

$$FTacc = 1 - \frac{acos\left(\frac{\vec{\mathbf{ed}}\cdot\vec{\mathbf{gd}}}{|\vec{\mathbf{ed}}||\vec{\mathbf{gd}}|}\right)}{acos\left(\frac{\vec{\mathbf{sd}}\cdot\vec{\mathbf{gd}}}{|\vec{\mathbf{sd}}||\vec{\mathbf{gd}}|}\right)} \tag{B.1}$$

In *Following Path* task, the accuracy (FPacc) is calculated based on the participant's start location $\mathbf{r}^{(\mathbf{s})}$, the participant's end location $\mathbf{r}^{(\mathbf{e})}$ and ground truth location $\mathbf{r}^{(\mathbf{g})}$:

$$FPacc = 1 - \frac{\left|\mathbf{r}^{(\mathbf{e})} - \mathbf{r}^{(\mathbf{g})}\right|}{\left|\mathbf{r}^{(\mathbf{s})} - \mathbf{r}^{(\mathbf{g})}\right|} \tag{B.2}$$

In *Observing Progress* task, the accuracy (ETacc) is based on the start time moment $\mathbf{t}^{(\mathbf{s})}$, end time moment $\mathbf{t}^{(\mathbf{e})}$ ground truth countdown time $\mathbf{t}^{(\mathbf{g})}$:

$$ETacc = 1 - \frac{\left|\mathbf{t}^{(\mathbf{g})} - \mathbf{t}^{(\mathbf{e})}\right|}{\left|\mathbf{t}^{(\mathbf{g})} - \mathbf{t}^{(\mathbf{s})}\right|} \tag{B.3}$$

Appendix C. Additional result data from the study

The Tables C.1–C.8, Figs. C.9 and C.10 show additional results from the experiment we described in Section 6. Table C.1 shows all mean task completion times, accuracy scores, and their corresponding 95% confidence intervals. We assessed the overall effects of cue conditions on user accuracy (three tasks) and completion time (two tasks). Normality was tested using the Shapiro–Wilk test. Repeated measures ANOVA was applied to normally distributed data; otherwise, the Friedman test was used. Effect sizes were partial eta squared (η_p^2) for ANOVA and Kendall's W for Friedman. Tables C.2 and C.3 summarize the main effects. Tables C.4–C.6 report pairwise p-values for accuracy; Tables C.7 and C.8 for completion time. Effect size thresholds: Kendall's W = 0.1 (small), 0.3 (medium), 0.5 (large); η_p^2 = 0.01 (small), 0.06 (medium), 0.14 (large). Fig. C.9 shows the overall user preference rank. Fig. C.10 shows the overall user workload and performance.

Table C.1
The mean accuracy for three tasks and completion times two tasks, and their corresponding 95% CIs.

	Cue	Accu.	95% CI	Time (s)	95% CI
Locating Target	Arrow	0.78	[0.69, 0.87]	5.73	[4.89, 6.56]
	Angle	0.94	[0.88, 1.00]	4.59	[4.09, 5.08]
	Compass	0.92	[0.87, 0.97]	4.44	[3.96, 4.93]
	Voice message	0.83	[0.76, 0.89]	7.85	[7.04, 8.66]
	Sound	0.83	[0.77, 0.89]	6.67	[5.77, 7.58]
Following Path	Arrow	0.94	[0.91, 0.98]	6.60	[5.87, 7.33]
	Color block	0.81	[0.73, 0.89]	8.27	[6.86, 9.68]
	Visual metaphor	0.75	[0.60, 0.90]	11.13	[8.19, 14.08]
	Voice message	0.68	[0.57, 0.80]	13.58	[11.71, 15.44]
	Sound	0.69	[0.57, 0.82]	12.28	[9.12, 15.43]
Observing Progress	Progress Bar	0.83	[0.80, 0.85]		
	Clock	0.81	[0.76, 0.87]		
	Text	0.80	[0.77, 0.84]		
	Voice message	0.69	[0.65, 0.73]		
	Sound	0.71	[0.67, 0.76]		

Table C.2
Summary of main effects using the Friedman test. Effect sizes are reported as Kendall's W.

Metrics	Task	χ^2	p-value	Kendall's W	Effect size
Accuracy	Locating Target	14.89	0.005	0.286	Small
	Following Path	14.09	0.007	0.271	Small
Time	Locating Target	35.69	<0.001	0.686	Large
	Following Path	25.42	<0.001	0.489	Medium

Table C.3
Summary of main effects using repeated measures ANOVA. Effect sizes are reported as partial eta squared (η_p^2).

Metric	Task	F Value	p-value	η_p^2	Effect size
Accuracy	Observing Progress	12.56	<0.001	0.511	Large

Table C.4
Pairwise comparisons of accuracy in the Locating Target task using the Friedman–Wilcoxon procedure.

Cue1	Cue2	Adj. p-value	Reject	r	Effect size
		0.002	True	−1.000	Large
		1.000	False	−0.253	Small
		0.033	True	−0.670	Large
		0.017	True	−0.736	Large
		0.046	True	−0.846	Large
		0.942	False	−0.539	Large
		1.000	False	−0.385	Medium
		0.803	False	0.560	Large
		0.574	False	0.604	Large
		1.000	False	−0.208	Small

Table C.5
Pairwise comparisons of accuracy in the Following Path task using the Friedman test followed by Wilcoxon signed-rank tests.

Cue1	Cue2	Adj. p-value	Reject	r	Effect size
		0.007	True	−0.956	Large
		0.839	False	0.076	Negligible
		1.000	False	0.296	Small
		1.000	False	0.341	Medium
		0.024	True	−0.890	Large
		0.012	True	−0.934	Large
		0.024	True	−0.890	Large
		0.215	False	0.714	Large
		0.479	False	0.626	Large
		1.000	False	0.143	Small

Table C.6
Pairwise comparisons of accuracy in the Observing Progress task using the ANOVA–t-test procedure with Holm correction.

Cue1	Cue2	Adj. p-value	Reject	Cohen's d	Effect size
Clock	TEXT	1.000	False	0.146	Negligible
Clock	Voice message	0.002	True	1.630	Large
Clock	Sound	0.019	True	1.255	Large
Clock	ProgressBar	1.000	False	−0.166	Negligible
TEXT	Voice message	<0.001	True	1.876	Large
TEXT	Sound	0.011	True	1.374	Large
TEXT	ProgressBar	1.000	False	−0.450	Small
Voice message	Sound	1.000	False	0.341	Small
Voice message	ProgressBar	<0.001	True	−2.544	Large
Sound	ProgressBar	<0.001	True	−1.901	Large

Table C.7
Pairwise comparisons of completion time in the Locating Target task using the Friedman–Wilcoxon procedure.

Cue1	Cue2	Adj. p-value	Reject	r	Effect size
Angle	Arrow	0.134	False	0.758	Large
Angle	Compass	1.000	False	−0.099	Negligible
Angle	Voice message	0.002	True	1.000	Large
Angle	Sound	0.002	True	1.000	Large
Arrow	Compass	0.479	False	0.626	Large
Arrow	Voice message	0.017	True	−0.912	Large
Arrow	Sound	0.046	True	−0.846	Large
Compass	Voice message	0.017	True	−0.736	Large
Compass	Sound	0.013	True	−0.758	Large
Voice message	Sound	0.398	False	−0.648	Large

Table C.8
Pairwise comparisons of completion time in the Following Path task using the Friedman–Wilcoxon procedure.

Cue1	Cue2	Adj. p-value	Reject	r	Effect size
Visual metaphor	Arrow	0.024	True	0.890	Large
Visual metaphor	Color block	0.681	False	−0.582	Large
Visual metaphor	Voice message	1.000	False	−0.451	Medium
Visual metaphor	Sound	1.000	False	−0.121	Small
Arrow	Color block	0.215	False	0.714	Large
Arrow	Voice message	0.002	True	1.000	Large
Arrow	Sound	0.004	True	0.978	Large
Color block	Voice message	0.012	True	−0.934	Large
Color block	Sound	0.327	False	−0.670	Large
Voice message	Sound	1.000	False	−0.341	Medium

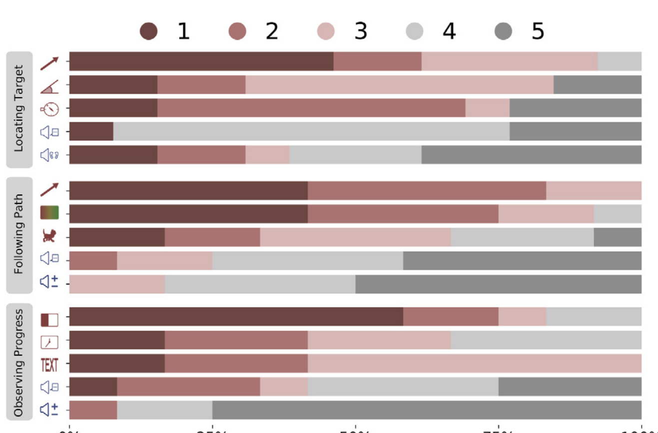

Fig. C.9. The rank of user preference in three tasks.

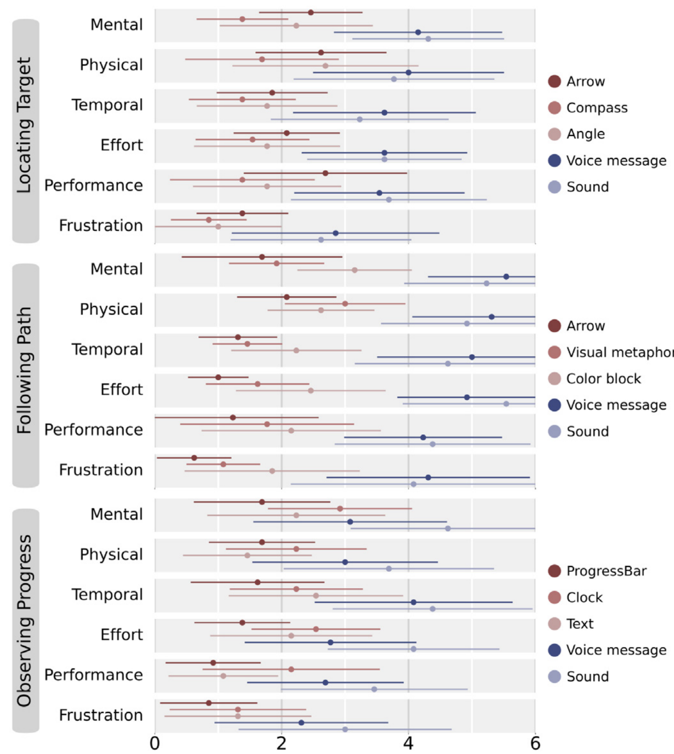

Fig. C.10. User workload and performance. Error bars show 95% CIs.

Appendix D. Environment setting of the study

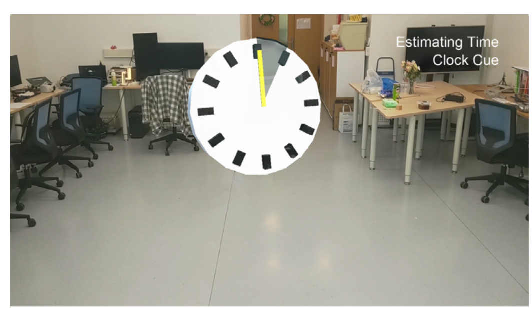

Fig. D.11. User study setup in the lab, the clock cue from the *Observing Progress* task as an example.

Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.visinf.2026.100309.

References

Arbeláez, J., Viganò, R., Osorio-Gómez, G., 2019. Haptic augmented reality (HapticAR) for assembly guidance. Int. J. Interact. Des. Manuf. 13, 673–687. http://dx.doi.org/10.1007/S12008-019-00532-3.

Baguley, T., 2009. Standardized or simple effect size: What should be reported? Br. J. Psychol. 100 (3), 603–617. http://dx.doi.org/10.1348/000712608X377117.

Billinghurst, M., Kato, H., 2002. Collaborative augmented reality. Commun. ACM 45 (7), 64–70. http://dx.doi.org/10.1145/514236.514265.

Bowman, D.A., Hodges, L.F., 1999. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. J. Vis. Lang. Comput. 10 (1), 37–53. http://dx.doi.org/10.1006/jvlc.1998.0111.

Cantarella, G., Mioni, G., Bisiacchi, P.S., 2024. Young adults and multisensory time perception: Visual and auditory pathways in comparison. Atten. Percept. Psychophys. 86 (4), 1386–1399. http://dx.doi.org/10.3758/s13414-023-02773-7.

Coull, J.T., Nobre, A.C., 1998. Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J. Neurosci. 18 (18), 7426–7435. http://dx.doi.org/10.1523/JNEUROSCI.18-18-07426.1998.

Cumming, G., 2014. The new statistics: Why and how. Psychol. Sci. 25 (1), 7–29. http://dx.doi.org/10.1177/0956797613504966.

Cutting, J., Vishton, P., 1995. Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. In: Percept. Space Mot., vol. 5, pp. 69–177.

Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18, 193–222. http://dx.doi.org/10.1146/annurev.ne.18.030195.001205.

Dominic, J., Robb, A., 2020. Exploring effects of screen-fixed and world-fixed annotation on navigation in virtual reality. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces. IEEE, pp. 607–615. http://dx.doi.org/10.1109/VR46266.2020.00083.

Dozio, N., Maggioni, E., Pittera, D., Gallace, A., Obrist, M., 2021. May I smell your attention: Exploration of smell and sound for visuospatial attention in virtual reality. Front. Psychol. 12, 671470. http://dx.doi.org/10.3389/fpsyg.2021.671470.

Dragicevic, P., 2016. Fair statistical communication in HCI. In: Modern Statistical Methods for HCI. Springer International Publishing, Cham, pp. 291–330. http://dx.doi.org/10.1007/978-3-319-26633-6_13.

Dragicevic, P., Chevalier, F., Huot, S., 2014. Running an HCI experiment in multiple parallel universes. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 607–618. http://dx.doi.org/10.1145/2559206.2578881.

Duncan, J., Humphreys, G.W., 1989. Visual search and stimulus similarity. Psychol Rev 96 (3), 433–458. http://dx.doi.org/10.1037/0033-295X.96.3.433.

Ellis, S.R., Menges, B.M., 1998. Localization of virtual objects in the near visual field. Hum. Factors 40 (3), 415–431. http://dx.doi.org/10.1518/001872098779591278.

Elmqvist, N., Tsigas, P., 2008. A taxonomy of 3d occlusion management for visualization. IEEE Trans. Vis. Comput. Graphics 14 (5), 1095–1109.

Eriksen, C.W., St. James, J.D., 1986. Visual attention within and around the field of focal attention: A zoom lens model. Percept. Psychophys. 40 (4), 225–240. http://dx.doi.org/10.3758/BF03211502.

Gaspelin, N., Luck, S.J., 2018. The role of inhibition in avoiding distraction by salient stimuli. Trends Cogn. Sci. 22 (1), 79–92. http://dx.doi.org/10.1016/j.tics.2017.11.001.

Goldstein, E.B., 2001. Blackwell handbook of perception. In: Handbooks of Experimental Psychology. Blackwell, Oxford, UK.

Gröhn, M., Lokki, T., Takala, T., 2005. Comparison of auditory, visual, and audiovisual navigation in a 3D space. ACM Trans. Appl. Percept. 2 (4), 564–570. http://dx.doi.org/10.1145/1101530.1101558.

Hart, S., 2006. NASA-Task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, (9), SAGE Publications, pp. 904–908. http://dx.doi.org/10.1177/154193120605000909.

He, S., Yao, H., Jiang, L., Li, K., Xiang, N., Li, Y., Liang, H.N., Yu, L., 2024. Data cubes in hand: A design space of tangible cubes for visualizing 3D spatio-temporal data in mixed reality. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. ACM, p. 209. http://dx.doi.org/10.1145/3613904.3642740.

Henderson, S., Feiner, S., 2011a. Augmented reality in the psychomotor phase of a procedural task. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 191–200. http://dx.doi.org/10.1109/ISMAR.2011.6092386.

Henderson, S., Feiner, S., 2011b. Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans. Vis. Comput. Graphics 17 (10), 1355–1368. http://dx.doi.org/10.1109/TVCG.2010.245.

Hwang, S., Kim, Y., Seo, Y., Kim, S., 2023. Enhancing seamless walking in virtual reality: Application of bone-conduction vibration in redirected walking. In: 2023 IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 1181–1190. http://dx.doi.org/10.1109/ISMAR59233.2023.00135.

Israr, A., Poupyrev, I., 2011. Tactile brush: drawing on skin with a tactile grid display. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 2019–2028. http://dx.doi.org/10.1145/1978942.1979235.

Itti, L., Koch, C., 2001. Computational modelling of visual attention. Nat. Rev. Neurosci. 2 (3), 194–203. http://dx.doi.org/10.1038/35058500.

Klein, R.M., 2000. Inhibition of return. Trends Cogn. Sci. 4 (4), 138–147. http://dx.doi.org/10.1016/S1364-6613(00)01452-2.

Lee, J., Jin, F., Kim, Y., Lindlbauer, D., 2022. User preference for navigation instructions in mixed reality. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces. IEEE, pp. 802–811. http://dx.doi.org/10.1109/VR51125.2022.00102.

Lehtinen, V., Oulasvirta, A., Salovaara, A., Nurmi, P., 2012. Dynamic tactile guidance for visual search tasks. In: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 445–452. http://dx.doi.org/10.1145/2380116.2380173.

Lepelley, M., Lejeune, L., Lestienne, F., 2009. Study of tactile guidance of the hand in a pointing task by means of a vibrotactile device. Comput. Methods Biomech. Biomed. Eng. 12 (S1), 169–170. http://dx.doi.org/10.1080/10255840903091411.

Lin, T., Yang, Y., Beyer, J., Pfister, H., 2021. Labeling out-of-view objects in immersive analytics to support situated visual searching. IEEE Trans. Vis. Comput. Graphics 29 (3), 1831–1844.

Lu, W., Duh, H.B.-L., Feiner, S., Zhao, Q., 2014. Attributes of subtle cues for facilitating visual search in augmented reality. IEEE Trans. Vis. Comput. Graphics 20 (3), 404–412. http://dx.doi.org/10.1109/TVCG.2013.241.

Marquardt, A., Trepkowski, C., Eibich, T.D., Maiero, J., Kruijff, E., Schöning, J., 2020. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. IEEE Trans. Vis. Comput. Graphics 26 (12), 3389–3401. http://dx.doi.org/10.1109/TVCG.2020.3023605.

Matviienko, A., Müller, F., Schön, D., Seesemann, P., Günther, S., Mühlhäuser, M., 2022. BikeAR: Understanding cyclists' crossing decision-making at uncontrolled intersections using augmented reality. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. ACM, p. 366. http://dx.doi.org/10.1145/3491102.3517560.

Müller, H.J., Rabbitt, P., 1989. Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. J. Exp. Psychol. Hum. Percept. Perform. 15 (2), 315–330. http://dx.doi.org/10.1037/0096-1523.15.2.315.

Neges, M., Koch, C., König, M., Abramovici, M., 2017. Combining visual natural markers and IMU for improved AR based indoor navigation. Adv. Eng. Inf. 31, 18–31. http://dx.doi.org/10.1016/j.aei.2015.10.005.

Nobre, A.C., 2001. Orienting attention to instants in time. Neuropsychologia 39 (12), 1317–1328. http://dx.doi.org/10.1016/S0028-3932(01)00120-8.

Nogalski, M., Fohl, W., 2016. Acoustic redirected walking with auditory cues by means of wave field synthesis. In: 2016 IEEE Virtual Reality. IEEE, pp. 245–246. http://dx.doi.org/10.1109/VR.2016.7504745.

Norouzi, N., Bruder, G., Erickson, A., Kim, K., Bailenson, J., Wisniewski, P., Hughes, C., Welch, G., 2021. Virtual animals as diegetic attention guidance mechanisms in 360-degree experiences. IEEE Trans. Vis. Comput. Graphics 27 (11), 4321–4331. http://dx.doi.org/10.1109/TVCG.2021.3106490.

Ogawa, K., Fujita, K., Sakamoto, S., Takashima, K., Kitamura, Y., 2023. Exploring visual-auditory redirected walking using auditory cues in reality. IEEE Trans. Vis. Comput. Graphics 30 (8), 5782–5794.

Parseihian, G., Jouffrais, C., Katz, B.F.G., 2014. Reaching nearby sources: comparison between real and virtual sound and visual targets. Front. Neurosci. 8, http://dx.doi.org/10.3389/fnins.2014.00269.

Posner, M.I., 1980. Orienting of attention. Q. J. Exp. Psychol. 32 (1), 3–25. http://dx.doi.org/10.1080/00335558008248231.

Posner, M.I., Snyder, C.R., Davidson, B.J., 1980. Attention and the detection of signals. J. Exp. Psychol. 109 (2), 160–174.

Pylyshyn, Z.W., Storm, R.W., 1988. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spat. Vis. 3 (3), 179–197. http://dx.doi.org/10.1163/156856888X00122.

Raikwar, A., Plabst, L., Batmaz, A.U., Niebling, F., Ortega, F.R., 2024. Ping! Your food is ready: Comparing different notification techniques in 3D AR cooking environment. In: 2024 IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 1157–1166. http://dx.doi.org/10.1109/ISMAR62088.2024.00132.

Sawaki, R., Luck, S.J., 2010. Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. Atten. Percept. Psychophys. 72 (6), 1455–1470. http://dx.doi.org/10.3758/APP.72.6.1455.

Schwerdtfeger, B., Klinker, G., 2008. Supporting order picking with augmented reality. In: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE/ACM, pp. 91–94. http://dx.doi.org/10.1109/ISMAR.2008.4637331.

Seeliger, A., Merz, G., Holz, C., Feuerriegel, S., 2021. Exploring the effect of visual cues on eye gaze during AR-guided picking and assembly tasks. In: 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct. IEEE, pp. 159–164. http://dx.doi.org/10.1109/ISMAR-Adjunct54149.2021.00041.

Shi, Z., Church, R.M., Meck, W.H., 2013. Bayesian optimization of time perception. Trends Cogn. Sci. 17 (11), 556–564. http://dx.doi.org/10.1016/j.tics.2013.09.009.

Sohlberg, M.M., Fickas, S., Hung, P.F., Fortier, A., 2007. A comparison of four prompt modes for route finding for community travellers with severe cognitive impairments. Brain Inj. 21 (5), 531–538. http://dx.doi.org/10.1080/02699050701311000.

Sutton, J., Langlotz, T., Plopski, A., Hornbæk, K., 2024. Flicker augmentations: Rapid brightness modulation for real-world visual guidance using augmented reality. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. ACM, p. 752. http://dx.doi.org/10.1145/3613904.3642085.

Sutton, J., Langlotz, T., Plopski, A., Zollmann, S., Itoh, Y., Regenbrecht, H., 2022. Look over there! Investigating saliency modulation for visual guidance with augmented reality glasses. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. ACM, p. 81. http://dx.doi.org/10.1145/3526113.3545633.

Tan, H., Gray, R., Young, J.J., Taylor, R., 2003. A haptic back display for attentional and directional cueing. Haptics-E 3 (1), 1–20.

Tong, W., Chen, Z., Xia, M., Lo, L.Y.H., Yuan, L., Bach, B., Qu, H., 2023. Exploring interactions with printed data visualizations in augmented reality. IEEE Trans. Vis. Comput. Graphics 29 (1), 523–533. http://dx.doi.org/10.1109/TVCG.2022.3209386.

Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. Cogn. Psychol. 12 (1), 97–136. http://dx.doi.org/10.1016/0010-0285(80)90005-5.

Trepkowski, C., Marquardt, A., Eibich, T.D., Shikanai, Y., Maiero, J., Kiyokawa, K., Kruijff, E., Schöning, J., König, P., 2022. Multisensory proximity and transition cues for improving target awareness in narrow field of view augmented reality displays. IEEE Trans. Vis. Comput. Graphics 28 (2), 1342–1362. http://dx.doi.org/10.1109/TVCG.2021.3116673.

Uchiyama, H., Covington, M.A., Potter, W.D., 2008. Vibrotactile glove guidance for semi-autonomous wheelchair operations. In: Proceedings of the 46th Annual Southeast Regional Conference on XX. ACM, pp. 336–339. http://dx.doi.org/10.1145/1593105.1593195.

Warm, J.S., Parasuraman, R., Matthews, G., 2008. Vigilance requires hard mental work and is stressful. Hum. Factors 50 (3), 433–441. http://dx.doi.org/10.1518/001872008X312152.

Weber, B., Schätzle, S., Hulin, T., Preusche, C., Deml, B., 2011. Evaluation of a vibrotactile feedback device for spatial guidance. In: 2011 IEEE World Haptics Conference. IEEE, pp. 349–354. http://dx.doi.org/10.1109/WHC.2011.5945511.

Wieland, J., Garcia, R.C.H., Reiterer, H., Feuchtner, T., 2022. Arrow, Bézier Curve, or Halos? – Comparing 3D out-of-view object visualization techniques for handheld augmented reality. In: 2022 IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 797–806. http://dx.doi.org/10.1109/ISMAR55827.2022.00098.

Wolfe, J.M., 1994. Guided search 2.0: A revised model of visual search. Psychon. Bull. Rev. 1 (2), 202–238. http://dx.doi.org/10.3758/BF03200774.

Wolfe, J.M., 2020. Visual search: How do we find what we are looking for? Annu. Rev. Vis. Sci. 6 (1), 539–562. http://dx.doi.org/10.1146/annurev-vision-091718-015048.

Wu, M., Cheng, Y.F., Lindlbauer, D., 2024. New ears: An exploratory study of audio interaction techniques for performing search in a virtual reality environment. In: 2024 IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 386–395. http://dx.doi.org/10.1109/ISMAR62088.2024.00053.

Yao, R., 2011. Publication manual of the American psychological association. Fam. Consum. Sci. Res. J. 39 (4), 442–443. http://dx.doi.org/10.1111/j.1552-3934.2011.02081.x.

Zhang, Y., Nowak, A., Xuan, Y., Romanowski, A., Fjeld, M., 2023. See or hear? Exploring the effect of visual/audio hints and gaze-assisted instant post-task feedback for visual search tasks in AR. In: 2023 IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 1113–1122. http://dx.doi.org/10.1109/ISMAR59233.2023.00128.

Zhao, Y., Stefanucci, J., Creem-Regehr, S., Bodenheimer, B., 2023. Evaluating augmented reality landmark cues and frame of reference displays with virtual reality. IEEE Trans. Vis. Comput. Graphics 29 (5), 2710–2720. http://dx.doi.org/10.1109/TVCG.2023.3247078.