

Article

A Comparative Study of Augmented Reality Cues for Robotic Three-Dimensional Pointing

Chih-Hsing Chu *, I-Fang Hsiao and Kuang-Jou Chen

Department of Industrial Engineering and Engineering Management, National Tsing Hua University,
Hsinchu 300044, Taiwan

* Correspondence: chchu@ie.nthu.edu.tw

How to Cite: Chu, C.-H.; Hsiao, I.-F.; Chen, K.-J. A Comparative Study of Augmented Reality Cues for Robotic Three-Dimensional Pointing. *Human-Machine Interaction* 2026, 1(1), 3.

Received: 4 May 2026

Revised: 15 August 2026

Accepted: 15 August 2026

Published: 27 August 2026

Abstract: In manufacturing environments, operators frequently need to input three-dimensional (3D) spatial positions to robotic systems to perform tasks such as grasping, assembly, and placement. Consequently, 3D pointing has become an important input modality in human-robot interaction for industrial applications. Previous studies have shown that different sensory cues delivered through augmented reality can improve operator performance in pointing tasks. This paper systematically evaluates the effects of visual, auditory, and haptic cues on operational performance during the manipulation of a Phantom Touch pointer to reach a target point across four real-virtual interaction configurations. The effects of different sensory modalities are compared using metrics including operational accuracy, task completion time, and subjective measures. The results indicate that sensory cues do not significantly improve pointing accuracy under aligned configurations, where both the pointer and target are either real or virtual. In contrast, under misaligned configurations, visual cues outperform other modalities in pointing accuracy, while haptic cues yield the shortest task completion time among the three modalities. These findings support a cue design principle for selecting appropriate sensory cues based on the interaction configuration and task objectives, providing practical guidance for AR-assisted human-robot interaction.

Keywords: human-robot interaction; human-machine interaction; interface design; augmented reality; pointing; sensory cues

1. Introduction

Industrial robots have become an increasingly indispensable element in modern smart manufacturing, collaborating with human operators on the shop floor. Robots are typically responsible for repetitive, precise, or highly physical tasks, while humans handle higher-level decision-making and command execution [1]. Human-robot collaboration is now commonly applied in manufacturing tasks such as welding, assembly, and transport operations, where human operators issue commands to control robotic movement [2]. Collaborative robots (cobots) can be moved through direct physical guidance or interface-based control, depending on the task and system design. In direct teaching mode, the operator physically holds the robot arm and manually guides it along designated waypoints. In interface-based control, the user specifies movement commands through devices such as a teach pendant or joystick. As a result, three-dimensional (3D) pointing has become an important input modality in human-robot interaction in manufacturing environments [3].

Augmented reality (AR) integrates virtual elements into the physical world, thereby enhancing users' perception of the environment and improving situational awareness. Previous studies have demonstrated the advantages of robot motion planning through multiple interaction modalities within AR interfaces, such as gestures, eye gaze, voice commands, and head movements [1,3–5]. These human inputs are translated into



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

executable commands that enable the robot to perform tasks in 3D space. During task execution, appropriate sensory cues, such as visual images, auditory signals, or tactile vibration feedback, can be delivered through AR in a timely and intuitive manner, allowing users to perform robot planning more effectively. Consequently, AR interfaces for 3D input tasks in robot motion planning have received increasing attention from the perspective of human-robot interaction in smart manufacturing [3,6–8].

In pointing tasks, both the moving object and the target may exist in either real or virtual forms, resulting in four different real-virtual interaction configurations. Previous studies have separately investigated pointing accuracy under these conditions in various application scenarios; however, a systematic comparison and analysis of all four conditions remains limited. Therefore, this study focuses on the use of AR-based sensory cues in robotic pointing tasks across different real-virtual configurations and evaluates how cue modalities affect operational performance. During the pointing tasks, users directly manipulate a robot arm in 3D space to move its pointer to a designated target position. User performance is compared across different combinations of sensory cues and interaction configurations using both objective and subjective measures. By examining performance across different real-virtual interaction configurations, this study proposes a cue design principle for selecting appropriate sensory modalities based on the interaction context. The findings provide design guidelines for AR-assisted pointing interfaces and effective human-robot interaction systems.

2. Literature Review

2.1. Sensory Cues in AR-Assisted 3D Input

Previous studies have demonstrated the advantages of different sensory cues across various aspects of performance in human-machine interaction [9,10]. In visual search tasks, tactile cues have been shown to facilitate faster reaction times, whereas auditory cues tend to reduce error rates [9]. Haptic cues are more effective than visual cues in preventing robot collisions with surrounding surfaces, while visual cues outperform haptic cues in avoiding unintended contact with target objects [10]. In general, haptic cues are well suited for conveying time-critical information that requires immediate attention, whereas visual cues are more effective for communicating continuous system status [10].

Vision is the most extensively studied sensory modality and is often regarded as particularly important for perceiving and interacting with the surrounding environment [11]. Accordingly, visual feedback is commonly used to support 3D pointing tasks. When performing tasks in 3D space, AR can be used to present spatial references and visualizations, including points and positions, paths and trajectories, regions and boundaries, as well as other visual representations that convey information about the robot or environment, such as direction, velocity, targets, and obstacles [12]. Visual feedback includes both inherent sensory information, such as direct visual observation and views obtained through head-mounted displays (HMD), and designed augmented visual cues overlaid on the interface. An AR-assisted teleoperation system for pick-and-place tasks proposed five types of visual assistance cues to support human-robot interaction, including motion guidance, obstacle indication, target hints, autonomy activation, and autonomy intention [13]. A study on pointing tasks compared different visual cues, including texture, support cylinders, and color highlighting. The findings indicated that highlighting was the most effective cue, functioning as selection feedback that significantly improved pointing performance [14]. In 3D input tasks, commonly used AR visual cues include directional guidance, distance-to-target representation, and target confirmation. Directional cues are typically visualized using arrow vectors or connecting lines to indicate the intended direction of movement [10,15]. Distance cues are often represented by lines linking the current position to the target, where distance is encoded through line length, numerical displays, or height indicators [10,15,16]. Target confirmation cues commonly include highlighting the target object, color changes, and text-based UI elements that indicate interaction states, such as affordances or task completion [10,17].

2.2. Depth Perception in Real-Virtual Interaction Configurations

3D pointing is more complex than planar pointing due to the increased degrees of freedom and the need to control movement in depth [18]. Depth perception arises from the integration of multiple depth cues, including occlusion, relative size, relative density, height in the visual field, aerial perspective, motion parallax, vergence, accommodation, and binocular disparity [19]. These cues do not contribute equally across all distances; instead, their relative importance varies with spatial scale. In near or personal space, depth perception is primarily supported by binocular disparity, convergence, and accommodation, whereas occlusion and relative size remain an effective and reliable source of depth information across all spatial ranges [19,20]. Perceptual matching tasks are used to evaluate how rendering design influences depth perception by requiring users to align virtual and

physical objects. Unlike real objects, virtual depth is not physically grounded but is instead conveyed through rendering cues [21].

Depth cues may be influenced by the interaction interface and operational context. Virtual-virtual pointing performance on stereoscopic display is reduced compared to real-real pointing in the physical world. This reduction may be attributed to lower ballistic movement accuracy and conflicting visual feedback due to insufficient depth perception [22]. Such visual conflicts may lead to more difficult and time-consuming corrective movements in virtual pointing [22]. When comparing virtual-virtual and real-virtual configurations using a virtual reality (VR) HMD and an optical see-through AR HMD, respectively, both devices exhibited similar overall performance. However, depth perception issues had a more pronounced impact in VR, likely due to the availability of real-world visual references in AR that provide richer real-world spatial information and motion parallax [23]. Different interaction configurations are also associated with various display modalities, including AR/VR HMDs, stereoscopic displays, and 2D displays. Performance differences may arise from the availability of depth cues, with 3D selection being affected by vergence-accommodation conflict and degraded under depth cue conflicts [24,25]. Teleoperation systems have attempted to enhance spatial perception through the integration of 3D sensing technologies, VR/AR interfaces, and visual annotations [26]. An AR-based VR interface provides preliminary evidence that visualizing depth information through AR overlays may enhance teleoperation performance by reducing perceptual ambiguity and improving spatial awareness [27].

The above literature review identifies two research gaps in existing studies. First, although different sensory cues have been applied to support 3D pointing tasks across various application contexts, systematic cross-comparisons under unified experimental conditions remain limited. Second, the influence of real-virtual interaction configurations on the effectiveness of different sensory modalities has not been thoroughly examined. This study aims to address these gaps through a comprehensive experimental investigation. The remainder of this paper is organized as follows. Section 3 presents the experimental design for evaluating pointing performance under different sensory cues in AR and real-virtual interaction configurations. Section 4 reports the key findings derived from statistical analysis. Section 5 concludes the paper and outlines directions for future research.

3. 3D Pointing Experiment

This section describes the experimental design and the implementation of the required interactive functions. It includes the design of sensory modalities, the hardware and software used, as well as the experimental procedures and evaluation metrics. Prior to conducting the experiments, the study protocol was approved by the Research Ethics Committee of National Tsing Hua University (REC No.: 11404HE064).

3.1. Experimental Setup

Figure 1 illustrates the experimental setup, which consists of a HoloLens 2 HMD, a Phantom Touch haptic device, and a desktop Unity server to support multimodal interaction in 3D pointing tasks. The HoloLens 2 renders the experimental AR scene, including visual and auditory cues, and displays the virtual pointer and target. The Unity server performs coordinate transformations between real and virtual spaces and records experimental data, communicating with the HoloLens 2 via TCP/IP over Wi-Fi. The Phantom Touch device is connected to the Unity server through a USB cable, providing haptic feedback while simultaneously reporting pointer position data. The setup allows the user, wearing the HoloLens 2, to manipulate the haptic device to control the pointer and reach a designated target, while receiving real-time multimodal feedback for pointing assistance in the AR environment.

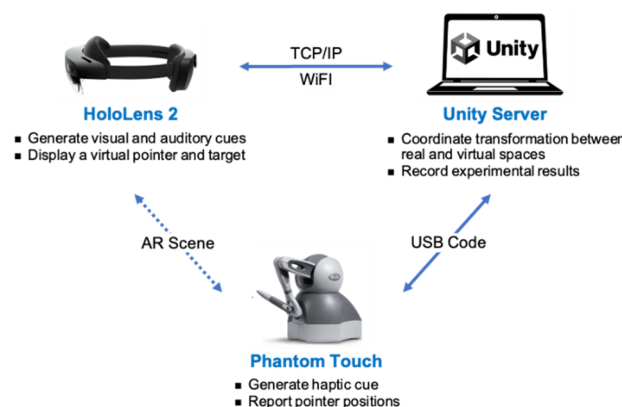


Figure 1. Experimental setup for AR-assisted 3D pointing with sensory cues.

The apex of a square pyramid serves as the real target, while its corresponding 3D model is presented as the virtual target in AR. As each trial involves multiple pointing tasks, pyramids of four different heights (60, 80, 100, and 120 mm) are randomly assigned to reduce potential learning effects. As shown in Figure 2, a custom-designed base is used to ensure stability and prevent tipping during the experiment. The orange marker at the tip of the stylus of the haptic device is defined as the real pointer controlled by the participant, which must be aligned with the target position (see Figure 3a). In contrast, the virtual pointer is defined as a point located 3 cm beyond the stylus tip along the direction from the stylus base to the stylus tip. As shown in Figure 3b, the blue ray represents this extension, and the red cursor at its end denotes the virtual pointer controlled by the participant.

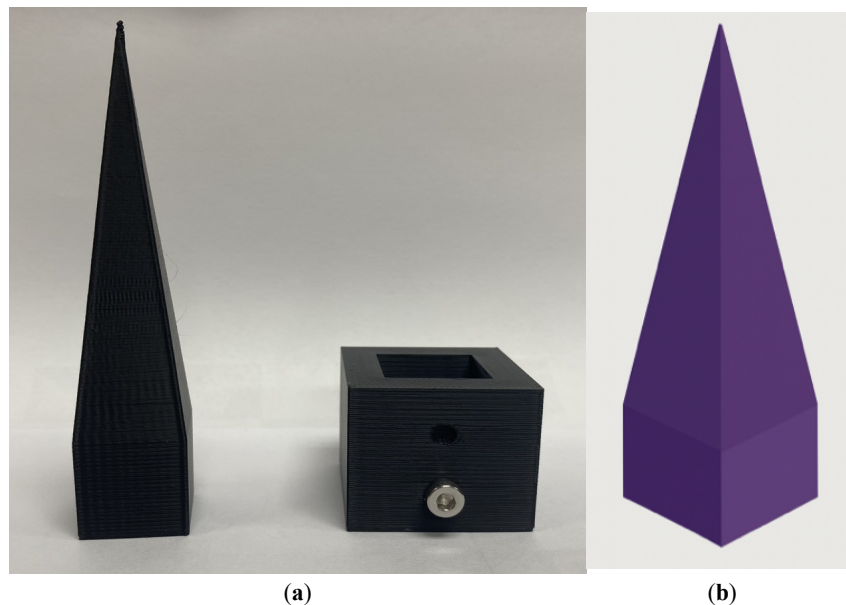


Figure 2. (a) Real square pyramid and its supporting base and (b) virtual target.

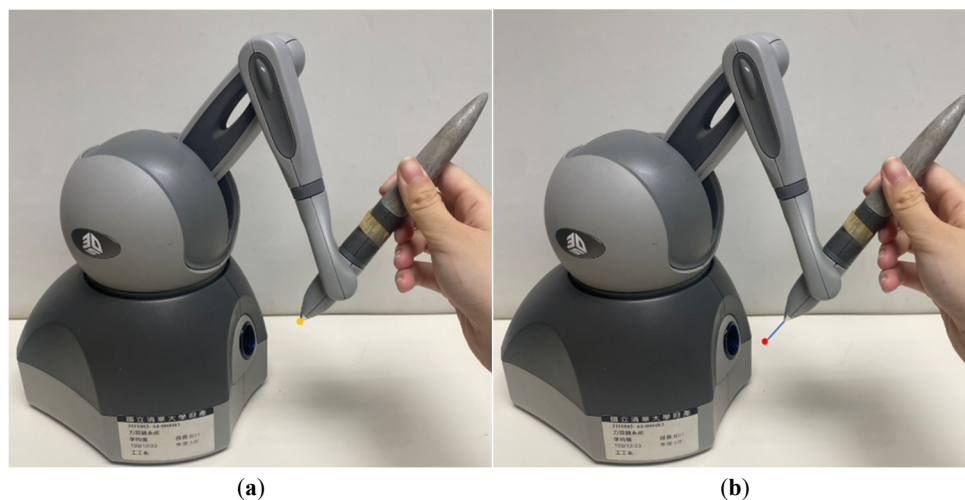


Figure 3. (a) Real pointer indicated by the orange marker and (b) representation of the virtual pointer.

To avoid interference from lighting variations on spatial perception, the experiment was conducted in an enclosed indoor environment with stable and uniform illumination. The experimental layout is shown in Figure 4. The target square pyramid was positioned 37 cm from the front edge of the table to accommodate a seated posture, with the upper arm hanging naturally and the forearm resting horizontally on the table. To eliminate visual interference, a white partition was positioned between the haptic device and the target. To reduce the influence of head movement on alignment accuracy between virtual content and corresponding real-world location, participants were instructed to place their chin on a chin rest to maintain a stable head posture. A rectangular platform was used to define the initial position of each trial; the stylus tip was placed on its top surface before initiating each pointing task. Throughout the experiment, participants wore the HoloLens 2 and performed the task using a natural pen-holding posture, with the stylus tip kept visible and not occluded by the hand.

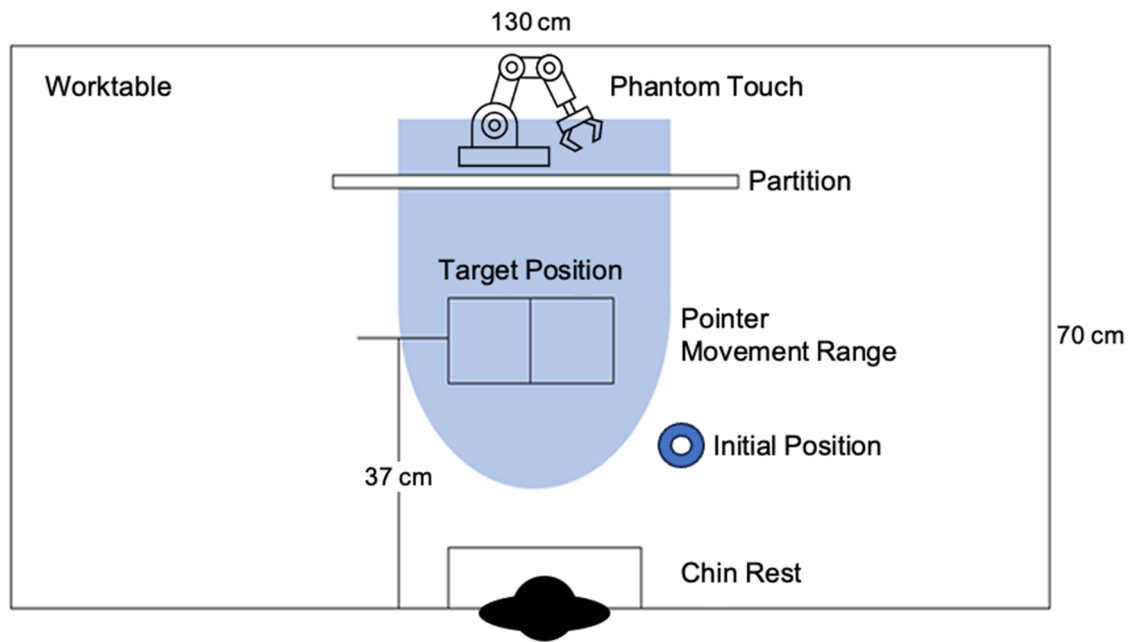


Figure 4. Experimental layout.

3.2. Sensory Cue Designs

3.2.1. Visual Cue Design

The design aims to assist users in perceiving the positional discrepancy between the pointer and the target along three spatial directions. As shown in Figure 5, three semi-transparent white planes represent the front-back, left-right, and up-down directions, respectively. A dynamically varying green square is overlaid on each plane to indicate the positional deviation along the corresponding axis. When the pointer is at the initial position, the square is at its minimum size $d_{\min}$. As it approaches the target from a distance, the square gradually increases in size; once the target is reached, the semi-transparent plane is fully covered by the square at its maximum size $d_{\max}$. This design provides complete three-axis spatial cues, allowing users to adjust one direction at a time. An advantage of this approach is that it integrates directional guidance and distance-related visual variation into a single cue, thereby reducing the process of integrating multiple sources of information.



Figure 5. Dynamically varying visual cue: (1) initial position, (2) approaching the target, and (3) target reached.

3.2.2. Auditory Cue Design

Based on the experimental findings reported by Bazilinskyy et al. [28], varying inter-stimulus intervals has been shown to improve performance in distance estimation tasks. Accordingly, the auditory cue in this study adopts this approach to assist users in perceiving the distance between the pointer and the target. As the cursor moves closer to the target, the interval between successive auditory beeps decreases (i.e., the repetition rate increases from $H_{\min}$ to $H_{\max}$). Throughout the pointing task, the sound pressure level was maintained at 66 dB SPL, and the pitch was fixed at approximately 2093 Hz.

3.2.3. Haptic Cue Design

The haptic cue design follows the approach proposed by Wang et al. [8]. In their study, the same haptic feedback device generated a pulling force to guide the pointer toward the target during 3D input tasks. When the user moves the pointer away from the target, the applied force opposes the movement direction, creating noticeable resistance and indicating that the movement direction should be adjusted. The magnitude of the force feedback is dynamically adjusted according to the positional discrepancy between the pointer and the target: a larger force is applied when the pointer is farther from the target, and a smaller force is applied as it approaches the target. The force magnitude ranges from $F_{\min}$ to $F_{\max}$. When the pointer coincided with the target, no force was applied. Once a positional discrepancy occurred, the force resumed from $F_{\min}$ in the direction of the target..

3.3. Experimental Procedure

A pilot experiment was first conducted to determine an optimal modulation pattern for each sensory cue design. Five participants (three males, two females) were recruited to directly manipulate the Phantom Touch pointer to a real target (see Figure 6). The cue parameters selected based on the pilot experiment are summarized in Table 1. It has been shown that during the fine-adjustment stage, when the pointer is very close to the target, non-linear modulation of assistance cues can improve user accuracy [7]. Therefore, this study evaluates the effects of different modulation patterns on pointing accuracy by conducting a pilot experiment that compares three modulation functions, square, linear, and square-root, to identify the most effective approach for each sensory modality.

Table 1. Design features of the implemented sensory cues.

Sensory Cue	Modulation Pattern	Design Features
Visual cue	Linear transformation	$d_{\min} = 15 \text{ mm}$, $d_{\max} = 50 \text{ mm}$
Auditory cue	Square transformation	$H_{\min} = 1 \text{ Hz}$, $H_{\max} = 8 \text{ Hz}$ 2093 Hz tone at 66 dB
Haptic cue	Square-root (sqrt) transformation	$F_{\min} = 0.1 \text{ N}$; $F_{\max} = 0.4 \text{ N}$

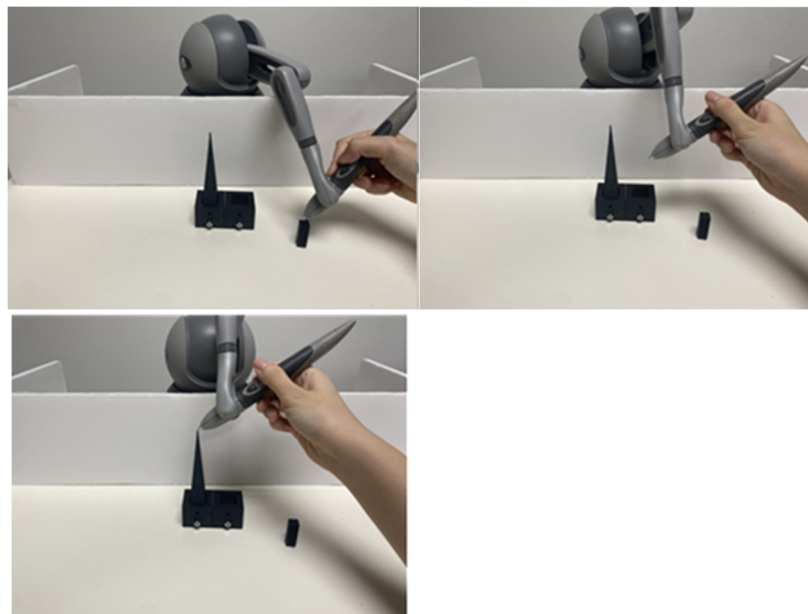


Figure 6. Manipulation of a Phantom Touch pointer to reach a vertex on a real target.

The main experiment employed a within-subject design. Participants were recruited based on the following criteria: 18–30 years of age, normal or corrected-to-normal vision, no color vision deficiency, and right-handedness. All participants provided informed consent before the experiment and completed a practice session of the pointing tasks with an emphasis on accuracy. In total, 20 participants with an equal gender ratio were recruited to examine the effects of sensory cues and pointer-target configurations on pointing performance. Each participant experienced all combinations of the three sensory cue conditions and the no-feedback baseline across the four real-virtual interaction configurations. The real-virtual interaction configuration factor comprised four

pointer-target configurations: real-real (R-R), real-virtual (R-V), virtual-real (V-R), and virtual-virtual (V-V) conditions. Within each condition, the pointing task was performed four times at different target heights. Participants were also instructed to prioritize accuracy when completing the pointing task. They were not required to deliberately increase their movement speed, but intentional slowing was also discouraged.

Both objective and subjective measures were collected in the experiment. As summarized in Table 2, task performance was evaluated using pointing accuracy and completion time, while subjective workload was assessed by a custom questionnaire. Accuracy is defined as the Euclidean distance between the final pointer position and the target, and completion time is measured as the duration from the start signal to voluntary trial termination. For statistical analysis, the real-virtual interaction configuration factor was represented by two binary factors, pointer and target, corresponding to whether the pointer and target were real or virtual. The data were analyzed using linear mixed models (LMMs), with sensory cue, pointer, target, and their interactions specified as fixed effects, and the subject treated as a random effect. For skewed dependent variables, including accuracy and completion time, a logarithmic transformation was applied prior to statistical analysis. Partial η^2 values (η_p^2) were calculated from the F statistics to estimate the effect sizes of the fixed effects. Tukey's post hoc test was conducted for pairwise comparisons.

Table 2. Evaluation criteria during the experiment.

Evaluation Criteria	Definition
Accuracy	Euclidean distance between the indicated position and the target.
Completion time	The duration from start signal to voluntary trial termination.
Subjective questionnaire	A custom five-point Likert-scale questionnaire adapted from the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX) to evaluate mental workload, perceived complexity, intuitiveness, helpfulness, ease of use, confidence, and predictability.

In the reported F statistics, the numerator degrees of freedom were determined by the number of levels of fixed-effect factors. For example, the sensory cue factor contained four levels, resulting in three numerator degrees of freedom ($4 - 1 = 3$) in the accuracy and completion time analyses. The denominator degrees of freedom were estimated using the Satterthwaite approximation based on the fitted linear mixed model. The analyses were based on 1280 observations (20 participants $\times$ 4 sensory cue conditions $\times$ 4 real-virtual interaction configurations $\times$ 4 repetitions at different target heights). The subjective questionnaire analysis included 240 observations (20 participants $\times$ 3 sensory cue conditions $\times$ 4 real-virtual interaction configurations) as questionnaire data were collected only for the three sensory cue conditions, excluding the no-feedback baseline. Hence, the sensory cue factor had two numerator degrees of freedom ($3 - 1 = 2$) in the subjective questionnaire analysis.

4. Statistical Analysis of Experimental Results

4.1. Statistical Analysis Results

4.1.1. Pointing Accuracy

Figure 7 presents the statistical results of pointing accuracy across the three sensory cues and the no-feedback baseline. There was a significant main effect of sensory cue ($F(3, 1245) = 28.349, p < 0.001, \eta_p^2 = 0.064$). Significant interactions were also observed between pointer and target ($F(1, 1245) = 484.222, p < 0.001, \eta_p^2 = 0.280$), and among sensory cue, pointer, and target ($F(3, 1245) = 13.487, p < 0.001, \eta_p^2 = 0.031$). In the R-R and V-V conditions, no significant differences in pointing accuracy across sensory cues were observed. In contrast, the effect of sensory cue was significant in the R-V condition ($p < 0.001$). Post hoc comparisons revealed that visual cues significantly outperformed no feedback, haptic cues, and auditory cues in pointing accuracy. A similar pattern was observed in the V-R condition ($p < 0.001$), in which the visual cues yielded significantly greater pointing accuracy than all other conditions. Regardless of sensory cue, the R-R and V-V conditions yielded significantly greater pointing accuracy than the R-V and V-R conditions.

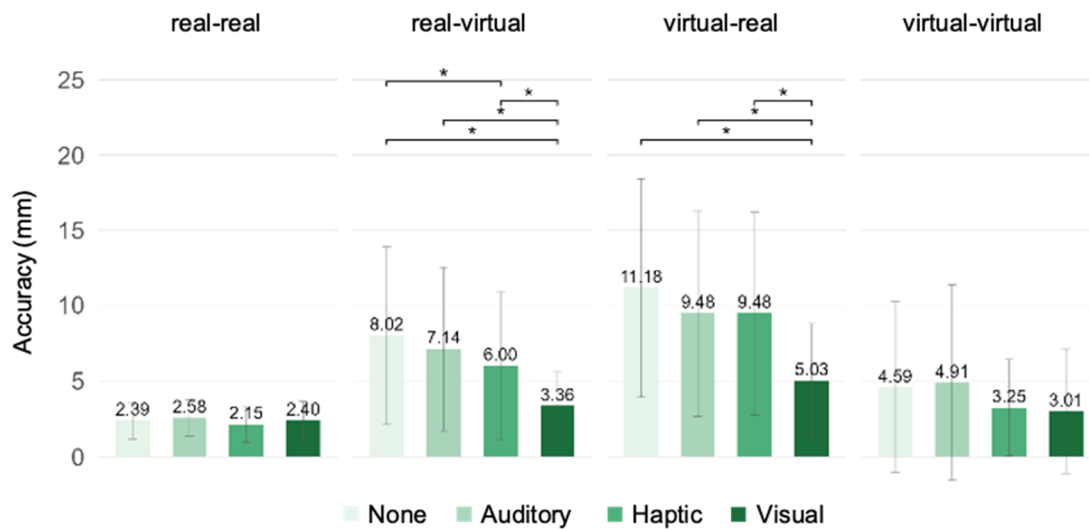


Figure 7. Comparison of accuracy across modality cues under four interaction configurations, with error bars representing standard deviations and horizontal brackets indicating pairwise comparisons based on Tukey's post hoc test; asterisks denote statistically significant differences ($p < 0.05$).

4.1.2. Completion Time

Figure 8 presents the statistical results for completion time across the three sensory cue conditions and the no-feedback baseline. There were significant main effects of sensory cue ($F(3, 1245) = 89.205, p < 0.001, \eta_p^2 = 0.177$). Significant interactions were also observed between pointer and target ($F(1, 1245) = 120.349, p < 0.001, \eta_p^2 = 0.088$), and among sensory cue, pointer, and target ($F(3, 1245) = 7.370, p < 0.001, \eta_p^2 = 0.017$). In the R-R condition, post hoc comparisons revealed that completion time was significantly longer with visual cues than with no feedback ($p < 0.001$) and haptic cues ($p = 0.013$). A similar pattern was observed in the V-V condition, in which completion time was significantly longer with visual cues than with no feedback or haptic cues (both $p < 0.001$). In both the R-V and V-R conditions, haptic cues and no feedback resulted in significantly shorter completion times than visual or auditory cues ($p = 0.004$ for the comparison between no feedback and auditory cues in the R-V condition; all other comparisons, $p < 0.001$). Regardless of sensory cue, the R-R condition yielded significantly shorter completion times than the other three conditions.

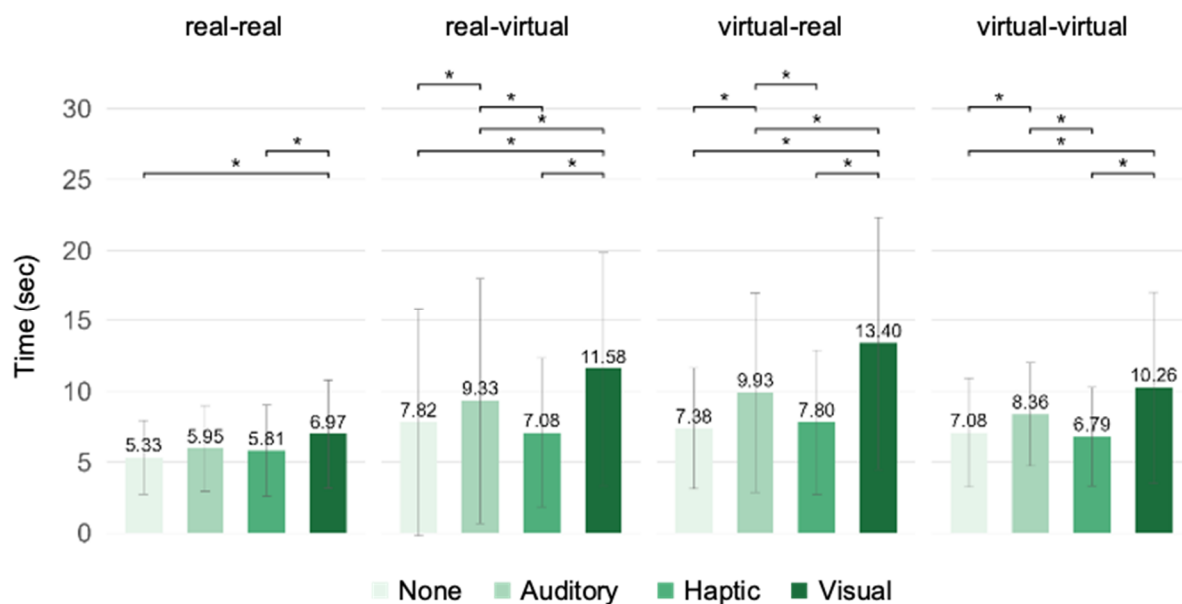


Figure 8. Comparison of completion time across modality cues under four interaction configurations, with error bars representing standard deviations and horizontal brackets indicating pairwise comparisons based on Tukey's post hoc test; asterisks denote statistically significant differences ($p < 0.05$).

4.1.3. Subjective Evaluation

A custom five-point Likert-scale questionnaire, adapted from the SUS and NASA-TLX, was used to examine the trade-off between objective task performance and perceived cognitive workload. Subjective questionnaire data were collected only for the three sensory-cue conditions, excluding the no-feedback baseline (Table 3). A significant main effect of sensory cue was observed for every questionnaire dimension. Auditory cues were rated as more intuitive than haptic cues, whereas both visual and auditory cues received higher ratings than haptic cues for helpfulness, ease of use, predictability, and user confidence.

Table 3. Statistics of subjective evaluation.

Dimension	Sensory Cue Effect	Comparison
Mental workload	$F(2, 209) = 9.457, p < 0.001, \eta_p^2 = 0.083$	Visual > others ($p < 0.001$)
Complexity	$F(2, 209) = 9.836, p < 0.001, \eta_p^2 = 0.086$	Visual > auditory ($p < 0.001$) Visual > haptic ($p = 0.001$)
Intuitiveness	$F(2, 209) = 4.214, p = 0.016, \eta_p^2 = 0.039$	Auditory > haptic ($p = 0.019$)
Helpfulness	$F(2, 209) = 22.716, p < 0.001, \eta_p^2 = 0.179$	Haptic < others ($p < 0.001$)
Ease of use	$F(2, 209) = 13.170, p < 0.001, \eta_p^2 = 0.112$	Haptic < others ($p < 0.001$)
Confidence	$F(2, 209) = 18.223, p < 0.001, \eta_p^2 = 0.148$	Haptic < others ($p < 0.001$)
Predictability	$F(2, 209) = 17.264, p < 0.001, \eta_p^2 = 0.142$	Haptic < others ($p < 0.001$)

4.2. Discussion

No significant differences in accuracy were observed between the aligned configurations (R-R and V-V), and accuracy did not differ significantly across sensory cue conditions within either configuration. In the previous study [29], participants were required to observe a target and then walk to its location with their eyes closed, preventing continuous updates of spatial cues during movement and resulting in larger distance estimation errors. However, the pointing task in this study allows continuous visual access to the target throughout the trial. Unity rendering maintained correct virtual occlusion relationships between the pointer and the target during movement, providing consistent depth cues. Thus, the V-V condition achieves accuracy comparable to the R-R condition without the need for additional sensory cues. In contrast, aligned configurations demonstrated more accurate pointing than misaligned configurations, which may be attributed to the greater availability of spatial perception cues when the pointer and target maintain consistent configurations. Misaligned configurations not only lack certain spatial cues but may also introduce visual conflicts due to discrepancies between depth perception sources. Consequently, distance estimation may be more difficult under misaligned configurations, resulting in reduced task performance.

Visual cues significantly improved accuracy only in the misaligned configurations, although they increased completion time in both aligned and misaligned configurations. This performance pattern is likely attributable to the more comprehensive spatial information provided by visual cues, which may require additional processing time. Specifically, the design incorporates two elements. First, a 3D planar cue represents positional discrepancies along the left-right, up-down, and front-back directions through changes in the area of colored patches, enabling users to identify misaligned directions and obtain full 3D adjustment information. Second, a bounding-frame cue indicates whether the object has been aligned with the target, serving as a final confirmation of alignment. Visual cues compensate for the depth ambiguity in misaligned configurations; however, integrating the information provided by visual cues with information directly observed from the real-world environment may require additional cognitive processing, resulting in longer completion time and higher mental workload. The haptic cues had the shortest completion time among the sensory cues. This may be attributable to haptic guidance facilitating movement correction by providing continuous force information that can be directly incorporated into motor control. This interpretation is consistent with the concept that tighter perception-action coupling can reduce the need for intermediate stimulus-response translation during movement [30], although perception-action coupling was not directly measured in this study.

The findings reveal trade-offs between performance and cognitive demand across different sensory cues. Although visual cues are effective in improving spatial accuracy, their design complexity should be carefully managed to avoid excessive cognitive burden. In contrast, haptic cues facilitated faster task completion by providing more direct guidance through physical movement, thereby reducing the need for extensive visual processing. This suggests that haptic information supports rapid motor adjustment, which is particularly beneficial in time-sensitive operations. However, compared to visual cues, haptic cues did not achieve the same level of accuracy improvement, indicating that different sensory modalities may serve complementary roles depending on

task objectives. From an interaction design perspective, these findings suggest that sensory cues should be selected based on both the interaction configuration and task objectives to achieve an appropriate balance between operational performance and cognitive workload.

4.3. Limitations

This study experimentally identified how real-virtual interaction configurations influence pointing performance and demonstrated that the effects of sensory cues were particularly pronounced under misaligned configurations, with modality-specific benefits in accuracy and completion time. While these findings provide empirical evidence for sensory cue design in 3D pointing, their generalizability to more complex robotic tasks requires further investigation. First, the objective of the present study was not to replicate an industrial robot programming environment, but to establish a controlled experimental setting for investigating interaction design principles in three-dimensional pointing. The Phantom device, workspace scale, and static pointing task were intentionally selected to isolate the effects of sensory modalities and interaction configurations. Second, the experimental setup was smaller than a typical industrial workspace, and differences in movement range and target distance may have influenced the observed performance. Third, although this study focused on three-dimensional pointing to static targets, this interaction serves as the basis for many robotic programming tasks, including trajectory teaching, orientation specification, assembly path planning, and dynamic target selection. Therefore, the proposed design principles may provide a theoretical foundation for interaction design beyond the tasks examined in this study. Future research should validate these principles in more complex robotic scenarios. Fourth, the visual, auditory, and haptic cues conveyed different types of task-relevant information. Specifically, the visual cue provided 3D spatial deviations and target confirmation, whereas the auditory and haptic cues conveyed distance and directional guidance, respectively. Consequently, the observed differences may reflect both sensory modality and information representation. Although this limits direct comparisons across modalities, the cue designs also reflect the inherent characteristics of each sensory modality. Future research should develop more comparable cue designs to better isolate the effects of sensory modality and evaluate their effectiveness in full-scale robotic operations.

5. Conclusions

In the supervision mode of human-robot collaboration, humans assume a decision-making role by issuing high-level commands, while industrial robots execute motion and manipulation tasks accordingly. In such contexts, operators frequently specify target positions, trajectories, or orientations in three-dimensional space to guide robot movement. Although AR has been applied to support this input process, the effects of different sensory cues on operational performance during robotic pointing tasks under varying interaction configurations remain insufficiently studied. Therefore, this research systematically investigated the effects of visual, auditory, and haptic cues on 3D pointing performance under four real-virtual interaction configurations in an AR-based human-robot interface. Both objective measures (accuracy and completion time) and subjective workload were used to provide a comprehensive evaluation. The experimental results indicate that the effectiveness of sensory cues is highly dependent on the real-virtual configuration. Under aligned conditions (real-real and virtual-virtual), no significant differences in accuracy were observed among sensory cues, suggesting that sufficient spatial information is inherently available and additional assistance provides limited benefit. Notably, the virtual-virtual condition achieved accuracy comparable to the real-real condition, which may be attributed to steady visual rendering and continuous depth cues provided during the task. In contrast, under misaligned configurations (real-virtual and virtual-real), visual cues significantly outperformed auditory and haptic cues in pointing accuracy. This advantage is likely due to the more comprehensive spatial information conveyed by the visual modality design, including directional guidance and alignment confirmation. Meanwhile, haptic cues were associated with shorter completion times, indicating their effectiveness in guiding rapid movement, albeit with lower spatial precision. Subjective evaluations further revealed that visual cues tend to increase perceived mental workload and task complexity, suggesting a trade-off between performance and cognitive demand.

This study makes a theoretical contribution by proposing a cue design principle that emphasizes selecting appropriate sensory modalities based on the interaction context and by identifying the performance degradation caused by misaligned interaction configurations. The experimental findings provide insights into AR-assisted 3D input interfaces and support the development of more effective human-robot interaction systems. Future work may be extended in two directions. First, the scope of 3D input tasks can be expanded to include more complex operations, such as rotation and trajectory tracing. Second, as multimodal cues can complement environmental information, their combined effects on user performance warrant further investigation. Future studies may explore

the integration of different sensory modalities to examine whether their combination can provide enhanced assistance under various real-virtual interaction configurations.

Author Contributions

C.-H.C.: conceptualization and design; data interpretation; drafting of the manuscript; critical revision of the manuscript; final approval; agreement to be accountable for all aspects of the research; I.-F.H.: conceptualization and design; data acquisition; data analysis and interpretation; agreement to be accountable for all aspects of the research; K.-J.C.: conceptualization and design; data acquisition; data analysis and interpretation; drafting of the manuscript; critical revision of the manuscript; agreement to be accountable for all aspects of the research. All authors have read and agreed to the published version of the manuscript.

Funding

This work was financially supported by the National Science and Technology Council, Taiwan, under the grant number 112-2221-E-007-098 MY3.

Institutional Review Board Statement

The study was approved by the Research Ethics Review Committee of National Tsing Hua University (No. 11404HE064; approved on 12 May 2025).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, C.H. Chu, upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used OpenAI's ChatGPT solely for English language polishing and improving the clarity and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Chu, C.H.; Weng, C.Y. Experimental Analysis of Augmented Reality Interfaces for Robot Programming by Demonstration in Manufacturing. *J. Manuf. Syst.* **2024**, *74*, 463–476. <https://doi.org/10.1016/j.jmsy.2024.03.016>.
2. Tien, K.W.; Lu, Y.J.; Chu, C.H. Task Allocation Based on Single-Limb Actions for Augmented Reality-Assisted Human-Robot Collaboration. *J. Manuf. Syst.* **2025**, *82*, 1000–1019. <https://doi.org/10.1016/j.jmsy.2025.08.006>.
3. Chu, C.H.; Lee, I.Y. An Experimental Study on the Design of Interaction Modalities in Augmented Reality for Industrial Robot Programming. *J. Comput. Inf. Sci. Eng.* **2025**, *25*, 051005. <https://doi.org/10.1115/1.4068084>.
4. Chu, C.H.; Liu, Y.L. Augmented Reality User Interface Design and Experimental Evaluation for Human-Robot Collaborative Assembly. *J. Manuf. Syst.* **2023**, *68*, 313–324. <https://doi.org/10.1016/j.jmsy.2023.04.007>.
5. Ong, S.K.; Yew, A.W.W.; Thanigaivel, N.K.; et al. Augmented Reality-Assisted Robot Programming System for Industrial Applications. *Robot. Comput.-Integr. Manuf.* **2020**, *61*, 101820. <https://doi.org/10.1016/j.rcim.2019.101820>.
6. Chen, K.J.; Chu, C.H. An Experimental Study on the Effects of Control Frame and View in Robot Teleoperation. *Int. J. Hum.-Comput. Interact.* **2026**, 1–23. <https://doi.org/10.1080/10447318.2026.2651379>.
7. Qin, Y.; He, W.; Zhang, X.; et al. Usability of Pseudo-Haptic Feedback for Manual Precise Manipulation of Objects in Augmented Reality. *Int. J. Hum.-Comput. Interact.* **2025**, *41*, 7923–7938.
8. Wang, I.J.; Yeh, L.; Chu, C.H.; et al. Assistive Sensory Feedback for Trajectory Tracking in Augmented Reality. *J. Comput. Inf. Sci. Eng.* **2024**, *24*, 031001. <https://doi.org/10.1115/1.4062543>.
9. Hancock, P.A.; Mercado, J.E.; Merlo, J.; et al. Improving Target Detection in Visual Search through the Augmenting Multi-Sensory Cues. *Ergonomics* **2013**, *56*, 729–738. <https://doi.org/10.1080/00140139.2013.771219>.

10. Lin, T.C.; Krishnan, A.U.; Li, Z. Comparison of Haptic and Augmented Reality Visual Cues for Assisting Tele-Manipulation. In Proceedings of the 2022 IEEE International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 9309–9316. <https://doi.org/10.1109/icra46639.2022.9811669>.
11. Huttmacher, F. Why Is There so Much More Research on Vision than on Any Other Sensory Modality? *Front. Psychol.* **2019**, *10*, 2246. <https://doi.org/10.3389/fpsyg.2019.02246>.
12. Suzuki, R.; Karim, A.; Xia, T.; et al. Augmented Reality and Robotics: A Survey and Taxonomy for AR-Enhanced Human-Robot Interaction and Robotic Interfaces. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022; pp. 1–33. <https://doi.org/10.1145/3491102.3517719>.
13. Krishnan, A.U.; Lin, T.C.; Li, Z. Human Preferred Augmented Reality Visual Cues for Remote Robot Manipulation Assistance: From Direct to Supervisory Control. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023; pp. 7034–7039. <https://doi.org/10.1109/iros55552.2023.10341969>.
14. Teather, R.J.; Stuerzlinger, W. Visual Aids in 3D Point Selection Experiments. In Proceedings of the 2nd ACM Symposium on Spatial User Interaction, Honolulu, Hawaii, USA, 4–5 October 2014; pp. 127–136. <https://doi.org/10.1145/2659766.2659770>.
15. Siew, C.Y.; Ong, S.K.; Nee, A.Y.C. A Practical Augmented Reality-Assisted Maintenance System Framework for Adaptive User Support. *Robot. Comput.-Integr. Manuf.* **2019**, *59*, 115–129. <https://doi.org/10.1016/j.rcim.2019.03.010>.
16. Wang, X.; Ong, S.K.; Nee, A.Y.C. Multi-Modal Augmented-Reality Assembly Guidance Based on Bare-Hand Interface. *Adv. Eng. Informatics* **2016**, *30*, 406–421. <https://doi.org/10.1016/j.aei.2016.05.004>.
17. Lin, T.C.; Krishnan, A.U.; Li, Z. Perception and Action Augmentation for Teleoperation Assistance in Freeform Telemanipulation. *ACM Trans. Hum.-Robot. Interact.* **2024**, *13*, 1–40. <https://doi.org/10.1145/3643804>.
18. Teather, R.J.; Stuerzlinger, W. Pointing at 3D Targets in a Stereo Head-Tracker Virtual Environment. In Proceedings of the 2011 IEEE Symposium on 3D User Interfaces (3DUI), Singapore, Singapore, 19–20 March 2011; pp. 87–94. <https://doi.org/10.1109/3dui.2011.5759222>.
19. Cutting, J.E.; Vishton, P.M. Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth. In *Perception of Space and Motion*; Academic Press: San Diego, CA, 1995; pp. 69–117.
20. Bhowmik, A.K. Virtual and Augmented Reality: Human Sensory-Perceptual Requirements and Trends for Immersive Spatial Computing Experiences. *J. Soc. Inf. Disp.* **2024**, *32*, 605–646. <https://doi.org/10.1002/jsid.2001>.
21. Diaz, C.; Walker, M.; Szafir, D.A.; et al. Designing for Depth Perceptions in Augmented Reality. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Nantes, France, 9–13 October 2017; pp. 111–122. <https://doi.org/10.1109/ismar.2017.28>.
22. Liu, L.; van Liere, R.; Nieuwenhuizen, C.; et al. Comparing Aimed Movements in the Real World and in Virtual Reality. In Proceedings of the 2009 IEEE Virtual Reality Conference, Lafayette, LA, USA, 14–18 March 2009; pp. 219–222. <https://doi.org/10.1109/vr.2009.4811026>.
23. Batmaz, A.U.; Machuca, M.D.B.; Pham, D.M.; et al. Do Head-Mounted Display Stereo Deficiencies Affect 3D Pointing Tasks in AR and VR? In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 585–592. <https://doi.org/10.1109/vr.2019.8797975>.
24. Bashar, M.R.; Barrera Machuca, M.D.B.; Stuerzlinger, W.; et al. The Effect of Visual Depth on the Vergence–Accommodation Conflict on 3D Selection Performance within Virtual Reality Headsets. *Vis. Comput.* **2025**, *41*, 9645–9661. <https://doi.org/10.1007/s00371-025-03990-x>.
25. Batmaz, A.U.; Turkmen, R.; Sarac, M.; et al. Re-Investigating the Effect of the Vergence-Accommodation Conflict on 3D Pointing. In Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology, Christchurch, New Zealand, 9–11 October 2023; pp. 1–10. <https://doi.org/10.1145/3611659.3615686>.
26. Lee, J.S.; Ham, Y.; Park, H.; et al. Challenges, Tasks, and Opportunities in Teleoperation of Excavator toward Human-in-the-Loop Construction Automation. *Autom. Constr.* **2022**, *135*, 104119. <https://doi.org/10.1016/j.autcon.2021.104119>.
27. Ganatra, S.; Jeong, H. Enhancing Human-Robot Teleoperation through Depth Sensing and eXtended Reality. In Proceedings of the 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Saint Malo, France, 8–12 March 2025; pp. 1110–1114. <https://doi.org/10.1109/VRW66409.2025.00221>.
28. Bazilinskyy, P.; van Haarlem, W.; Quraishi, H.; et al. Sonifying the Location of an Object: A Comparison of Three Methods. *IFAC-Pap.* **2016**, *49*, 531–536. <https://doi.org/10.1016/j.ifacol.2016.10.614>.
29. Jones, J.A.; Swan, J.E., II; Singh, G.; et al. The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception. In Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization, Los Angeles, CA, USA, 9–10 August 2008; pp. 9–14. <https://doi.org/10.1145/1394281.1394283>.
30. Adam, J.J.; Paas, F.G.W.C.; Buekers, M.J.; et al. Perception-Action Coupling in Choice Reaction Time Tasks. *Hum. Mov. Sci.* **1996**, *15*, 511–519. [https://doi.org/10.1016/0167-9457\(95\)00035-6](https://doi.org/10.1016/0167-9457(95)00035-6).