Int. J. Human–Computer Studies 205 (2025) 103650

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

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

Effects of interface layouts on cognitive performance for pedicle screw placement simulator in immersive environments

Lang Qin [a], Kadek Ananta Satriadi [b], Jiazhou Liu [b], Yuhan Zhan [a], Jiang Shao [a,*], Peimeng Liu [a], Zhiyong Chen [a], Yongtao Liu [c,*]

[a] School of Architecture & Design, China University of Mining and Technology, Xuzhou, China
[b] Faculty of Information Technology, Monash University, Melbourne, Australia
[c] Department of Spine Surgery, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

ARTICLE INFO

Keywords:
Virtual reality training
Pedicle screw placement
Interface layout
Cognitive performance
Visual search

ABSTRACT

Many medical procedures, including pedicle screw placement, require intricate hand-sight coordination. In recent years, immersive virtual reality (VR) technologies have gained traction in supporting training of such complex tasks. To effectively perform the task, the ability to see the screw position from different angles during the procedure is crucial, as it meets the user's need for comprehensive spatial information to guide their actions. Yet, current literature lacks guidelines for designing view layouts for VR simulators in this context. We conducted a repeated measure experiment investigating various layout parameters (8 layouts and 2 view sizes). We gathered behavioral metrics, eye-tracking data, and subjective ratings from 27 participants. We found that layout design significantly impacts task performance, with placing views on the left of the visual field in a vertical arrangement reducing task response time. Furthermore, we found the effects of view arrangements on the flow of visual search patterns. Our study provides design guidelines to inform future design of VR pedicle screw placement simulators and other types of simulators requiring the combination of manual tasks and multiple-perspective views.

1. Introduction

Virtual Reality (VR) and Augmented Reality (AR) techniques in medical procedures training and applications have gained traction and offered promising benefits. VR medical training enables surgeons and trainees to improve surgical proficiency, minimizes practical errors, and provides opportunities to address high-risk cases (Jensen et al., 2018; Abich et al., 2021; Lohre et al., 2020a; Frisk et al., 2022). This approach is safe, reproducible, and characterized by a low learning curve and cost, which explains the growing interest in adopting VR medical procedure training (Godzik et al., 2021; Morimoto et al., 2022). The use of AR, on the other hand, typically involves overlying information on the patient's body during the medical procedures. This information includes three-dimensional (3D) data of the patient such as spine model, surgical plan (Ma et al., 2017), or multi-perspective views which are useful in guiding the insertion of medical instruments (Jud et al., 2020; Elmi-Terander et al., 2019). In contrast to the traditional surgery approach, AR enables surgeons' focus to remain on the patient without having to switch their attention to screens to observe and read the information. Specifically, with a heads-up display, surgeons can maintain their focus on the patient while simultaneously assessing the anatomical structures and surgical trajectories (Frisk et al., 2022). Recent studies highlight the trends in AR-assisted surgical techniques, as well as evaluation measures in practice, human factors, challenges, and potential directions for future research (Birlo et al., 2022; Yoo et al., 2022). These immersive interactive experiences make learning or operations more convenient and efficient. Interactive interfaces assist with navigation, provide valuable visual information, and offer a more user-friendly experience.

Despite the rapid development of the domain, there remain open questions when it comes to the design of interfaces and the effects of surgical task performance, as well as cognitive patterns. Collaborating with physicians in the spine surgery domain, we use the pedicle screw placement procedure as a case study to address this design challenge. Their valuable suggestions ensuring the tasks in our user study were directly relevant to the human-computer interaction in this surgical application. The pedicle screw placement is a common spinal surgery

---

* Corresponding authors.
E-mail addresses: shaojiang@cumt.edu.cn (J. Shao), tmt@xzhmu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.ijhcs.2025.103650
Received 27 April 2025; Received in revised form 18 September 2025; Accepted 26 September 2025

technique requiring precise alignment and insertion to avoid neurovascular damage and ensure stability. This procedure requires intricate, often prolonged, hand-sight coordination. To effectively perform the task, the ability to see the screw position from different angles during the procedure is crucial, as it meets the user's need for comprehensive spatial information to guide their actions. Adequate image guidance is necessary during spine procedures, such as multiple views showing the top, front, and side views of the spine. (Ramirez Manuel et al., 2024; Sakai et al., 2020; Elmi-Terander et al., 2019). However, designing such views involves adjustment of several parameters including the spatial arrangement and size of the views. How such parameters affect task performance and attention patterns of the surgeon in immersive-interface-supported applications remains largely unknown.

In our study, we investigated several different layout designs by taking into account the spatial arrangement of three views (top, front, and side) and the size of the view (1.0, 0.8). Using VR environments as the study setup, we ran an experiment comparing these designs and gathered behavioral metrics, eye-tracking data, and subjective feedback. Our research demonstrated that the arrangement of visual elements has a notable impact on performance. Specifically, vertically aligned views positioned on the left side of the visual field led to a longer task response time. Notably, 89 % of the participants (24 out of 27) were right-handed, with the remainder being ambidextrous; all participants used their right hand as the dominant hand for the surgical tasks. We also observed how different view arrangements influenced users' visual search patterns.

These findings contribute practical design guidelines for developing better immersive medical procedure applications, particularly those used for pedicle screw placement and other applications requiring users to coordinate manual tasks while monitoring multiple views. Based on our study's findings, we propose a list of design implications provided at the end of this paper.

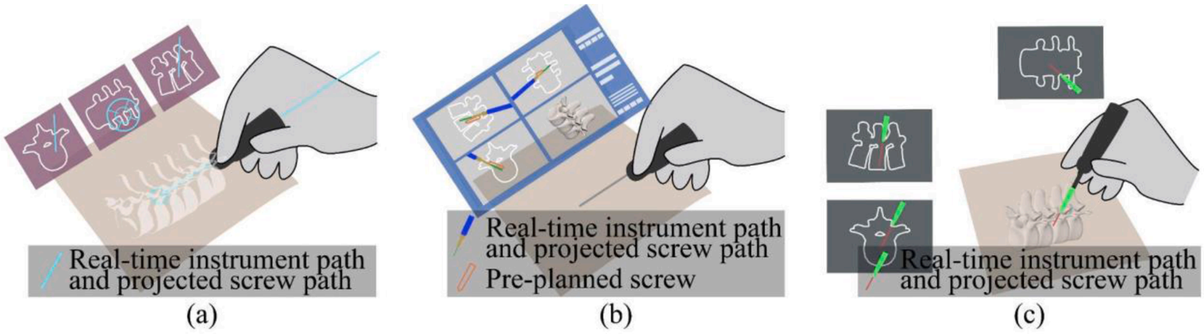
2. Related work

Our work is related to interfaces used in pedicle screw placement applications, Human-Computer Interaction studies on medical VR interfaces, and layout design in immersive environments, which we will review in this section.

2.1. Interfaces used in pedicle screw placement applications

To ensure safe and effective screw placement, surgeons must determine the insertion point and direction based on the conditions of the spine. The auxiliary interfaces provide real-time navigation by displaying both preoperative and intraoperative image data. The data preprocessing methods include the creation of surface rendering models and planning information, among others (Doughty et al., 2022). Interfaces in screw placement applications can vary depending on the layout and the type of presented information. In this section we highlight three different interfaces from existing work and provide examples, as shown in Fig. 1. The first example comes from a case in the review by Manuel et al., which describes an AR-assisted spinal surgery using the XVision system conducted by researchers at Johns Hopkins University (Ramirez Manuel et al., 2024). The second interface is from a study by Sakai et al. that utilized the Stealth Navigation system of Medtronic (Sakai et al., 2020). The third interface is the previous version of the AR spine surgery navigation system developed by Minliang Wang and his team (Caduceus S System, Previous versions of the Caduceus system, accessed 25 April 2025). These AR interfaces represent three common types of virtual interfaces based on the information presented.

In the linear layout interface (Fig. 1, left), the cross-sectional, frontal, and lateral views are horizontally arranged above the surgical site. This layout situates the information adjacent to the primary workspace, so users only need to move their eyes a short distance to view it while operating. In the grid layout interface (Fig. 1, middle), auxiliary views are also displayed above the surgical site but in a grid pattern. Compared to the 1 × 3 arrangement in the linear layout, the 2 × 2 grid increases complexity to some extent. In the around layout interface (Fig. 1, right), unlike the linear layout, the multiple views are not placed on a single side; instead, two views are vertically arranged on the left side of the



Interface / Content / Reference	Layout			Information					Display Modality	
				In auxiliary views			At the surgical site			
	Linear	Grid	Around	Spine	Pre-planned screw	Real-time instrument path	Virtual spine model	Projected screw path	Headset	Monitor
(a) Manuel et al.	☑	☐	☐	☑	☐	☑	☑	☑	☑	☑
(b) Sakai et al.	☐	☑	☐	☑	☑	☑	☐	☐	☑	☑
(c) Minliang Wang et al.	☐	☐	☑	☑	☐	☑	☑	☑	☑	☑

Fig. 1. Examples of AR spine surgery navigation systems and their interfaces: (a) XVision system (Manuel et al., 2024), (b) Medtronic Stealth Navigation system (Sakai et al., 2020), and (c) Caduceus S System (Wang et al., accessed 25 April 2025).

interface, forming a wrapped distribution around the surgical site for greater flexibility. Although the spatial distributions of these interfaces differ, they share the common characteristic of being positioned near the surgical site, primarily above or to the left of the field of view. We therefore discuss improved layout designs based on these features.

Regarding interface information in Fig. 1, both interfaces (a) and (c) display the spinal model, real-time instrument and screw path, and (b) shows pre-planned screws, enabling clear visualization and adjustment both in the auxiliary views and the surgical site. To ensure comparability with existing applications, our study references these three examples and implements these elements within an immersive environment.

2.2. Human-Computer interaction studies on medical VR interfaces

Human-Computer Interaction (HCI) studies of medical VR interfaces focus on how users interact with immersive systems, as well as assessing user performance and treatment outcomes (Mäkinen et al., 2022; Rose et al., 2018). Studies aim to improve usability, accessibility, and effectiveness by leveraging natural interaction methods and multimodal feedback (Heinrich et al., 2023; Li et al., 2021). Researchers combine gesture-based control with eye-tracking to explore complex data and create intuitive experiences that improve accessibility (Orlosky et al., 2017). These studies are centered around physical and visual interaction to optimize the user experience. Therefore, our research also addresses the key component of hand-eye interaction.

Studies related to the surgical navigation applications mainly test program feasibility, conduct comparative experiments, and evaluate whether these systems are more effective than traditional methods (Lohre et al., 2020b; Sakai et al., 2020). Surgeons need to interact with anatomical 3D models and view 2D medical imaging from a variety of perspectives (Wirth et al., 2018; Coffey et al., 2011). As a result, HCI researchers have focused on preferred interaction techniques for medical imaging, such as using VR to perform windowing tasks on 2D slices from three different viewpoints (Wirth et al., 2018). How to address information perception and ease of use in the cognitive processing stage when using the system is a common problem that has yet to be explored for AR-assisted surgical navigation applications (Doughty et al., 2022).

Previous studies have indicated that visual processing involves multiple forms of guidance, such as bottom-up, stimulus-driven guidance; top-down, user-driven guidance; scene guidance; guidance based on the perceived value of certain items or features; and guidance based on prior search history (Wolfe et al., 2017). These factors collectively modulate search behavior, influencing both accuracy and efficiency. This insight suggests that, in the visual search task of our study, visual attention may be affected by factors such as the features of spatial layout, the value of each image, the relationships between images, and previously viewed images.

2.3. Layout design in immersive environments

Designing interface layouts in immersive environments have been a quite established topic since the emergence of VR and AR applications. Layout refers to the arrangement of elements and can be quantitatively evaluated using aesthetic factors such as proportion, order, density, and equilibrium (Deng et al., 2020). Existing research has illustrated that the spatial layout of multiple views in immersive environments pose important roles in user behaviors and use of space (Satriadi et al., 2020) and spatial memory (Liu et al., 2024). Furthermore, design dimensions such as curvature (horizontal, vertical, or both), aspect ratio (e.g., 2 × 3, 3 × 2), and orientation (views facing the user, or facing the same direction) were found to affect the user's task completion time, accuracy, travelled distance, gaze change, and preferences (Liu et al., 2020). Most of these studies highlight the importance of layout design in a room-scale immersive environment. Despite these findings, how view layout impacts the effectiveness measures in smaller scale workspace such as pedicle screw placement remains under-explored.

In addition, research has shown that among the interface design dimensions, color and transparency affect users' information comprehension and cognitive performance (Hussain et al., 2024; Davis et al., 2023; Zhou et al., 2021), so other factors must be set as appropriate in layout studies. The color and transparency components are additional factors we are exploring in our study, differentiating our work to existing work in immersive layouts.

3. Design factors

The design of experimental tasks needs to consider the materials and procedures involved in the operation. Our user study was conducted based on the workflow of traditional surgeries, the associated visual information, and the interfaces of AR-assisted surgeries and VR surgical training, as shown in Fig. 2. In traditional surgery, surgeons analyze the screw placement path based on preoperative CT scans and 3D models, which involve the spine's top view, front view, and cross-sectional view, all of which are essential. During the surgery they determined the placement position and orientation with clinical experience. After the screw is inserted, they use intraoperative fluoroscopic instruments to capture X-ray images for inspection, involving the top view and front view. While this method of result inspection provides a general indication of the outcome, it does not reveal the cross-sectional view. In some AR-assisted clinical applications, virtual information is overlaid on the physical environment. This information includes 3D data such as spine models (Ma et al., 2017), or multi-perspective views which are useful in guiding the insertion of medical instruments (Jud et al., 2020; Elmi-Terander et al., 2019). The VR simulator provides virtual assistive information and operates within a virtual environment.

As the original context may no longer be accessible during the evaluation of an AR interaction, this requires the use of immersive VR replay that simulates the interaction context. Therefore, this paper uses VR technology for user studies, which can provide an immersive environment and manipulation experience.

3.1. Layout and size

We designed a within-subjects two-factor repeated measures experiment. The factors are layout schemes and view sizes, with a total of 16 combinations (8 layouts × 2 sizes).

Layout — the layout factor defines 1) how AR views are distributed on the user's view and 2) which side of the spine model is shown on these views. We refer to the first part as the Spatial Distribution Scheme and the Sequence Scheme for the second part. For the spatial distribution, we consider the left side, upper side, and bottom side of the user's view. For the sequence, we consider the arrangement order of the top-down, right, and front view of the spine. We then selected 8 unique combinations that form the layout factor of our study. To make referencing them easier, we label each layout instance based on the combination of spatial distribution and sequence, as shown in Fig. 3(a). L3 is three views arranged vertically on the left side; U1L2 is one view in the upper part and two views arranged vertically on the left side; L1U2 is one view on the left side and two views arranged horizontally in the upper part; U3 is three views arranged horizontally in the upper part; and B3 is three views arranged horizontally at the bottom. TRF is top view, right view, and front view; RTF is right view, top view, and front view.

Using this definition we label the final layouts as L3-TRF, L3-RTF, U1L2-TRF, U1L2-RTF, L1U2-TRF, L1U2-RTF, U3-TRF, and B3-TRF. Among these layout designs, U1L2-RTF is based on an existing AR application, specifically the Caduceus system (2019), with U1L2-TRF modifying its view sequence. U3-TRF references the AR system interface used by Ramirez Manuel et al. (2024), while B3-TRF places them at the bottom. The remaining four layouts are our original designs.

Size — the size factor defines the size of the spine view relative to the background. There are two size schemes: Full (100 %) and Reduced (80

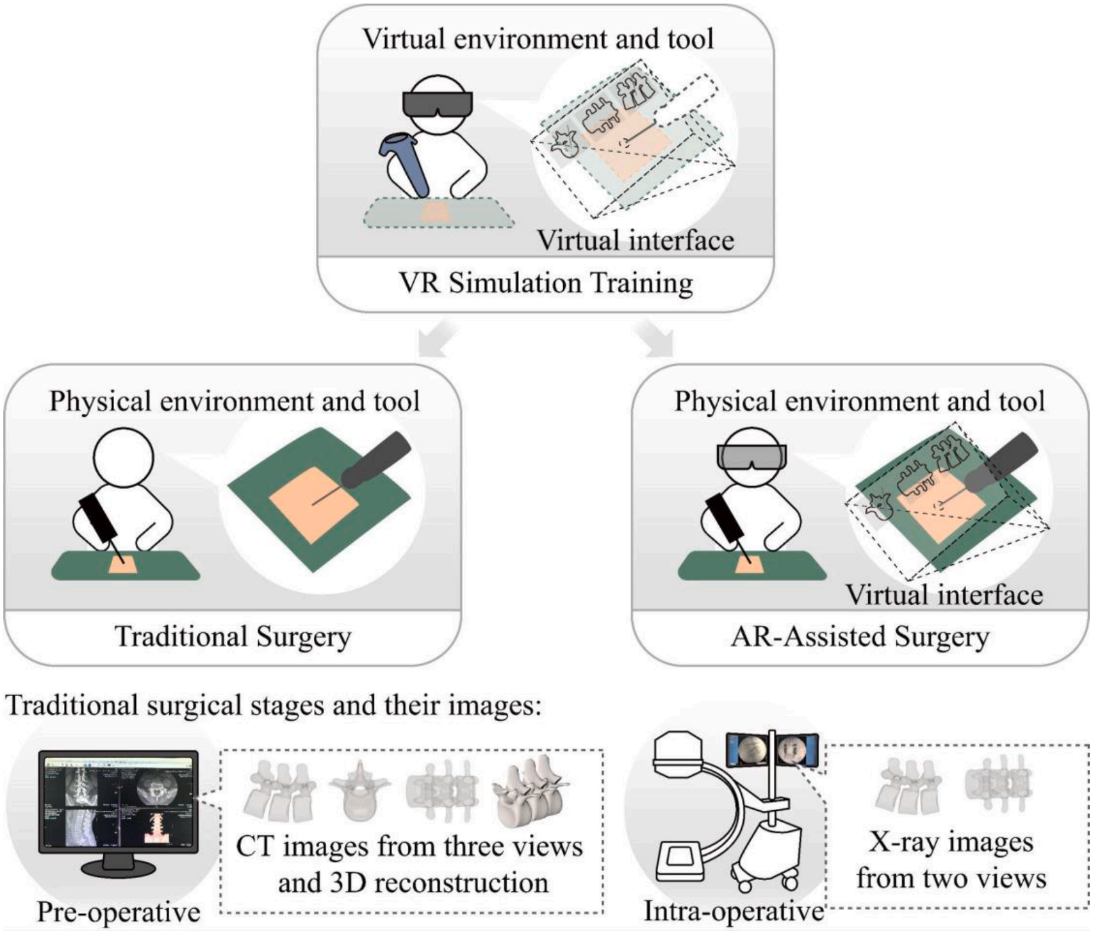


Fig. 2. The traditional process of pedicle screw placement surgery and the image-supported information. Human-computer interaction scenarios and interfaces in traditional, AR-assisted, and VR surgeries.

%), as shown in Fig. 3(b). The background of the views is a rectangle of the same size, with the view centered within it. The overall ratio of the two views is 1:0.8. The vertical angular size of 100 % is 31.9 degrees, while 80 % is 25.7 degrees. We selected these two sizes based on the pilot study. Both allow users to clearly view the content, but they differ in the spacing between the views.

3.2. Materials

We set the properties of the experimental materials based on existing studies and a pilot study. A detailed description is provided below.

Spine Model — The digital spine model used was the L3-L5 vertebrae of the lumbar spine. We created the model using 3D modeling software, replicating the structure and dimensions of the teaching mold.

Planning Screw — The planning screw model was set up as a red cylinder with a diameter of 6.0 mm and a length of 45.0 mm, which corresponds to the dimensions of a standard screw for the lumbar vertebrae L3-L5. The planning protocols used in the experiments, which were supported by clinically experienced spine surgeons, used conventional planning methods, with minor variations in the three views, and were similarly difficult to manipulate.

Handheld Tool — We replaced the default VR controller digital model with the surgical tool digital model. The anterior rod portion of the tool has a diameter of 6.0 mm. The physical controller in the participant's hand was assembled with a rod module to simulate a real tool and increase the realism of the operation.

Views — The front view and right view of the 3D screw model are displayed as a red rectangle measuring 6.0 mm × 45.0 mm. The top view of the screw is displayed as a red circle with a diameter of 6.0 mm, representing the insertion point. This caters to the practical technique of determining the entry point of the screw through the top view and adjusting the angle of entry through the front and right views.

As existing studies have shown that the transparency of interface elements may affect visual perception (Shen et al., 2021), we kept them consistent in the experiments. The spine model, planning screw, and handheld tool have no added transparency. View transparency was uniformly set at 0.6, a value determined after combining the practicalities of this procedure and the recommendations of extant studies (Hussain et al., 2024). This configuration was tested in a pilot study to ensure its effectiveness and to confirm that participants could clearly perceive the relevant visual elements while maintaining appropriate visual clarity.

4. Experiment

4.1. Task and interactions

Participants were asked to confirm the screw position and adjust the tool using the views. In the virtual environment, the digital spine model was fixed approximately 45 cm below the user's line of sight, corresponding to a height near the waist, mimicking the height at which a surgeon would operate during surgery. When the user brings the tool close to the spine, the tool's views from three perspectives will be displayed in real time in the three views.

Hand Operation — Participants are required to place the tip of the tool on the screw placement point, keep it stationary, and adjust the

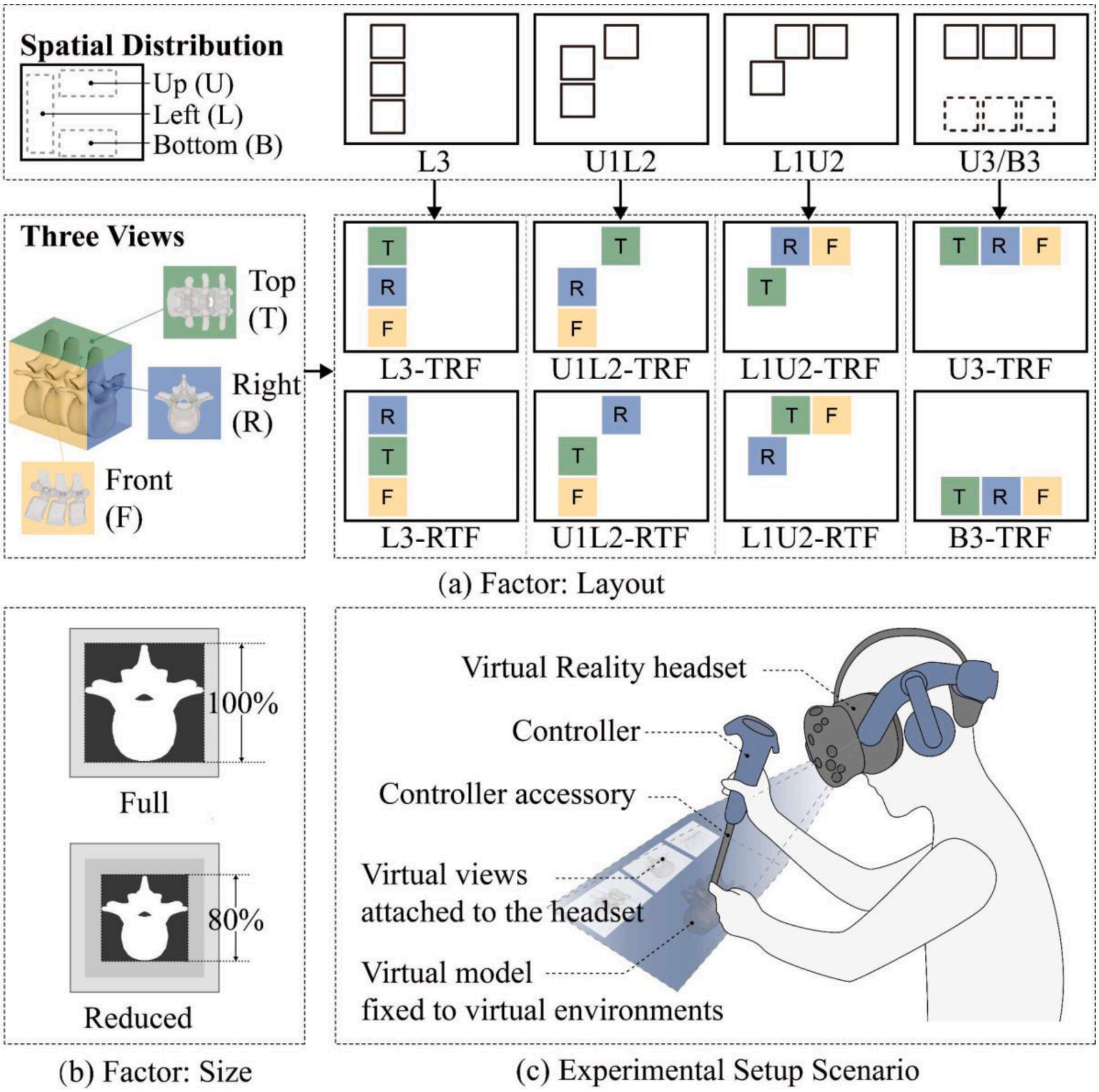


Fig. 3. Our experiment explored the (a) Layout and (b) Size factors of the AR views and was run in a virtual reality environment simulating the surgical procedure (c).

tool's tilt until it aligns with the planned direction, as shown in Fig. 4(a) and (b). After completing the calibration, press the button on the VR controller to confirm. It is worth noting that in actual surgery, the next step would be drilling. However, our task only requires calibration, without proceeding with the drilling operation. This is because drilling no longer relies on the interface view and involves applying force to the handheld tool, which goes beyond the scope of the VR experience.

Views Display — The interface of three views was fixed to the head-mounted display, which meant that the views were fixed in the user's field of view even if the user moved his or her head, as shown in Fig. 3 (c). We adopted this head anchoring mode because it allows the participants' viewing angles to remain consistent during the procedure. We did not use the world anchoring mode (where the image is fixed at a specific location in the environment) because it would result in unnecessary small bodily movements. Relevant studies also suggest that no specific anchoring mode for surgical images is considered superior, as it depends on the specific context (Deib et al., 2018).

4.2. Research hypotheses

The pedicle screw placement procedure discussed in this study requires hand-eye coordination on the part of the surgeon. Multiple views of the spine are commonly used as reference information. In order to perform the task effectively, the surgeon's ability to visualize the screw position from different angles is critical. How the layout of immersive affects the surgeon's task performance and attention patterns remains largely unknown. To address this gap, we propose three experimental hypotheses:

1. The spatial distribution of the layout affects task performance, with the position of images in the field of view influencing task response time and results. This is based on existing work (Cao et al., 2019; Nakatani et al., 2011) suggesting that when gazing at a screen, the human eye tends to focus more on the area above the line of sight than below it, and users are more likely to notice the left side of the screen than the right side. Horizontally and vertically distributed layouts lead to different user performance (Grobelny et al., 2005).
2. The specific arrangement of the three views (top, front, and right) influences task performance. The relationship between targets can be horizontally or vertically adjacent, spaced apart, or angularly deviated. For example, in layout U1L2-TRF, the right view and front view are vertically adjacent, the right view and top view are adjacent with

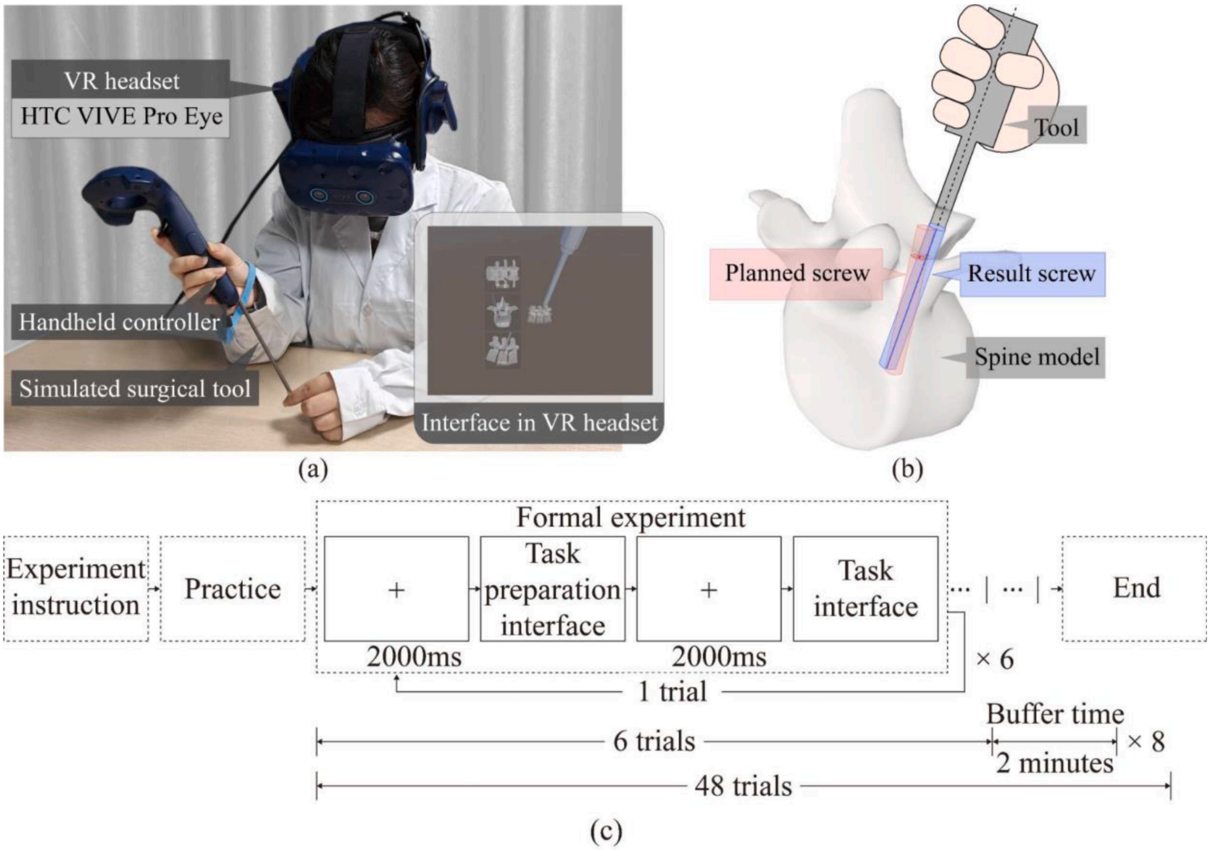


Fig. 4. Experimental setup: (a) Participant performing the task with interface in VR headset, (b) Hand operation and screw placement result, (c) Experimental procedure.

angular deviation, and the top view and front view are spaced with angular deviation. Existing work (Chen et al., 2011) suggesting that during visual search, the distance or relative position between elements affects user performance.

3. The size of the image affects task performance. The size parameters defined in this study correlate with the graphic area ratio and inter-element spacing, which means that larger sizes increase the area ratio while reducing the inter-element spacing and smaller sizes exhibit the inverse pattern, thereby influencing recognition efficiency. This is based on existing work (Shen et al., 2021; Niu et al., 2021) suggesting that proper area ratios with adequate image-outline spacing enhance the distinction between individual images and improve user recognition.

4.3. Participants and apparatus

A total of 27 participants (10 females, 17 males) were recruited for this experiment, with ages ranging from 20 to 28 years ($M = 23.41$, SD = 1.67). They are medical graduate students who are beginners in clinical practice. All had a normal or corrected-to-normal vision, with no color blindness or color weakness. Among them, 24 were right-handed, 3 were ambidextrous, and their dominant hand for operating tools was the right hand. Prior to using the VR surgical training simulator, participants had used VR devices between 0 and 5 times ($M = 1.04$, SD = 1.14), and were therefore considered VR novices with limited experience. The experiment was approved by our local Medical Ethics Committee, and all participants provided informed consent.

We developed the experiment software using Unity (2019.4.31f1c1) and ran the experiment software on Steam VR. We used the HTC VIVE Pro Eye as the VR headset. The headset features two 3.5-inch AMOLED screens, with a single-eye resolution of 1440 × 1600 and a dual-eye resolution of 3 K (2880 × 1600). It has a refresh rate of 90 Hz and a field of view of 110 degrees. We used the ErgoLAB software, an eye-tracking data collection platform, in combination with the VR Plugin to collect eye-tracking data.

This study was conducted at the Intelligent Interaction Laboratory of China University of Mining and Technology, in collaboration with the Embodied Visualisation Group at Monash University, Australia. The experiment was approved by the Medical Ethics Committee of the First People's Hospital of Xuzhou (Affiliated Hospital of China University of Mining and Technology) (approval document: No. xyy11 [2024] 002).

4.4. Procedure

The general procedure consisted of three phases: 1) introduction, 2) experiment. 3) questionnaire.

Introduction — After completing the demographic questionnaire and confirming informed consent, participants underwent task training. We explained the procedure, task requirements, and operational details to each participant. Before the formal experiment, participants had ample time to familiarize themselves with the VR tools and practice. After participants fully understood the process and tasks, and successfully completed more than six consecutive trials, they proceeded to the formal experiment.

Experiment — The experimental procedure is shown in Fig. 4(c). The formal experiment comprised 48 trials, consisting of all 16 combinations of 8 layouts and 2 sizes, each repeated 3 times. Each trial was divided into two phases: preparation and operation targeting. First, a "+" symbol appeared for 2000 ms to correct gaze. Then, in the preparation phase, a spine model and its three views were displayed. Participants could observe the views' positions, adjust their body posture, and bring the surgical tool closer to the operation area. Once ready, they pressed the trigger on the controller to continue. Next, a "+" symbol appeared again for 2000 ms, followed by the operation targeting phase. Based on the preparation view, red planning screws were displayed. After completing the adjustments, they pressed the trigger to confirm, ending the trial. The experiment followed this loop. The trials were conducted in a completely random order. After every six trials, participants took a 2-minute break. They need to wear the headset the entire time. The total experiment lasted about 40 min.

Questionnaire — After completing the task, participants filled out a questionnaire regarding the conditions.

4.5. Measures

Learning from existing interface evaluation studies (Molina et al. 2021; Renata et al. 2018), we collect the following data.

Response Time — Response time reflects the total time spent on cognition and operation. It is defined as the time taken to complete the targeting task. The response time for each trial is defined as the interval between the appearance time of the interface and the button press time for confirmation during the operation targeting phase.

Success Rate — Success rate indicates the quality of task completion. In pedicle screw placement surgery, if the screw breach is less than 2 mm after placement, it is defined as being within the safety zone (Aoude et al., 2015). In our study, tasks within the safe zone were defined as successful (score=1), and those outside as failures (score=0). We examine the task results through VR replay.

Gaze-data — Eye tracking techniques can be used for physiological measurements; specifically, fixation duration and saccade count are key metrics for assessing visual task performance (Ibbotson et al. 2011; Unema et al. 2005; Li et al. 2017; Ahlstrom et al. 2006). Fixation duration reflects the ease of information extraction; saccade count indicates the fluency of visual search. By analyzing eye movement behavior, we can investigate users' attention switching, visual perception and performance in immersive environments (Guo et al. 2016; Wang et al. 2021; Ariansyah et al. 2022; Li et al. 2021).

Workload — We used the NASA Task Load Index (NASA-TLX) questionnaire which is commonly used to measure users' subjective workload (Hart et al., 1988).

5. Results

5.1. Response time

In the data preprocessing, 26 valid participants were included, resulting in a total of 1248 data points. One participant was excluded due to missing 6 data points (missing rate of 12.5 %). There are 54 (4.3 % of the data) outliers that were caused by accidental button presses or excessively long response times. Accidental button presses refer to instances where participants unintentionally pressed the handle button, rather than confirming the task completion voluntarily. Excessively long response times refer to trials where the task time exceeded 50 s due to objective interference or participant distractions and lack of focus during task completion. After excluding these outliers, 1194 valid data points remained, which we used for the response time analysis.

Descriptive statistics of response times across all layouts and sizes are shown in Fig. 5. It indicates that the L3-RTF layout has the shortest mean response time of 15.46 s (SD=8.03), followed by L3-TRF (mean=16.05, SD=8.20). B3-TRF has the longest mean response time (mean=19.42, SD=10.71).

For all the data, the mean response time for the full size (100 %) is shorter than for the reduced size (80 %). Additionally, in six out of the eight layouts, the full size (100 %) consistently resulted in shorter response times than the reduced size (80 %). In U1L2-RTF trials, the mean response time between the two sizes differs greatly, whereas in other layouts, the difference between the two sizes' mean response times is minimal. To validate these observations, we ran parametric statistical tests.

We conducted a two-way repeated measures ANOVA to analyze the effects of layout and size on response time. The raw data was not normally distributed (Shapiro-Wilk tests: all $p < 0.001$), prompting a Box-Cox transformation ($\lambda = -0.1$). The transformed data met normality (Shapiro-Wilk tests: all $p > 0.05$) and sphericity assumptions (Mauchly's

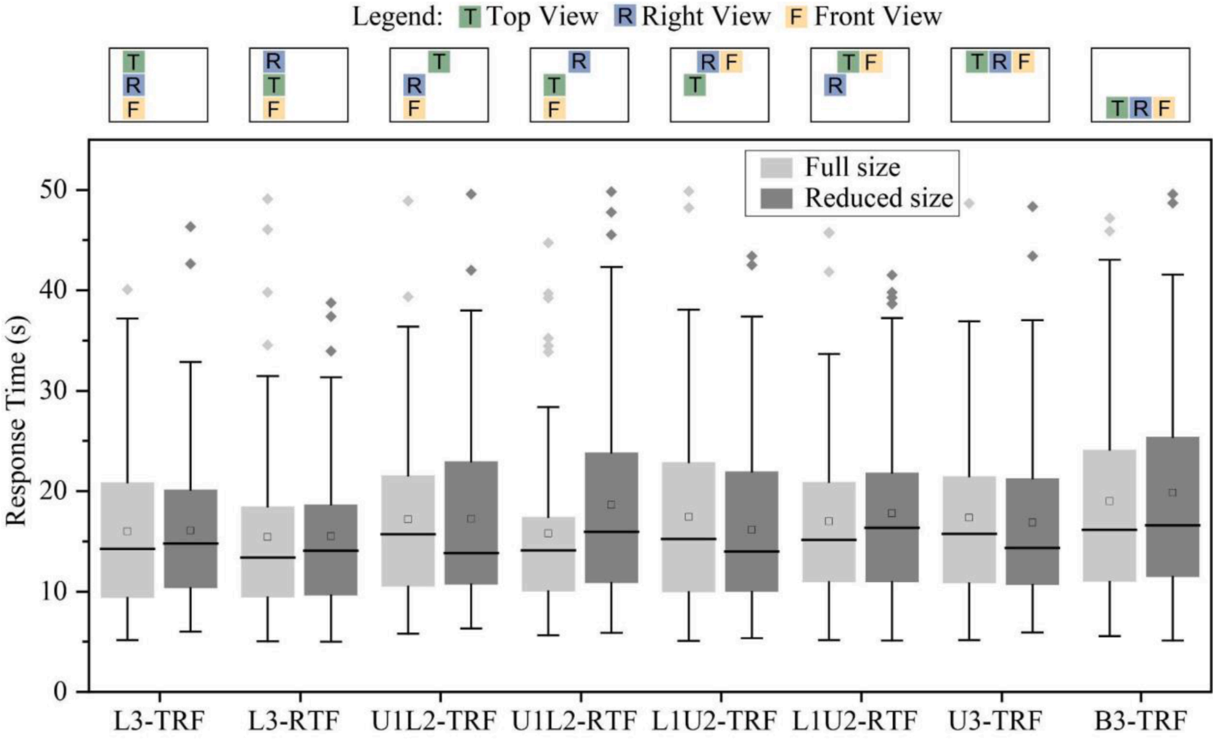


Fig. 5. Distribution of response times for all data across all layouts and sizes.

test: $W = 0.236$, $p = 0.281$) and were used for analysis.

The ANOVA revealed a significant main effect of layout ($F = 2.739$, $p = 0.010 < 0.05$, $\eta_p^2 = 0.118$), while the main effect of size ($F = 0.631$, $p = 0.435 > 0.05$, $\eta_p^2 = 0.040$) and the interaction effect between layout and size ($F = 0.277$, $p = 0.962 > 0.05$, $\eta_p^2 = 0.013$) were both non-significant, indicating no statistically significant influence. These results suggest that the layout significantly affects response time, while size does not have a significant impact on response time. This allowed for further analysis of layouts without interference from interaction effects with the size factor.

We further conducted pairwise comparisons (using LSD test) on the Layout factor. Mean response times and significance results of the pairwise test for all layouts are presented in Fig. 6. We found significant differences between L3-TRF and B3-TRF ($p = 0.024 < 0.05$), L3-RTF and U1L2-TRF ($p = 0.040 < 0.05$), L3-RTF and U1L2-RTF ($p = 0.029 < 0.05$), L3-RTF and L1U2-RTF ($p = 0.025 < 0.05$), L3-RTF and U3-TRF ($p = 0.044 < 0.05$), L3-RTF and B3-TRF ($p < 0.001$), U1L2-TRF and B3-TRF ($p = 0.022 < 0.05$), and L1U2-TRF and B3-TRF ($p = 0.033 < 0.05$).

The layout scheme analysis of the existing system is as follows:

For the U1L2-RTF (used in the Caduceus System, 2019): Compared to L3-RTF, which has the same view sequence, pairwise comparison (LDS test) showed a significant result ($p = 0.029 < 0.05$). The layout with all views on the left (L3-RTF) resulted in shorter response times (L3-RTF: mean = 15.46, SD = 8.03; U1L2-RTF: mean = 17.20, SD = 9.89). Compared to U1L2-TRF, which shares the same spatial distribution but with the positions of the right and top views swapped, the pairwise comparison did not show a significant result ($p = 0.855 > 0.05$), and the difference in response times was negligible (U1L2-TRF: mean = 17.20, SD = 9.27). Additionally, when comparing L3-TRF to L3-RTF, which have the same SDS but different SS, the RTF sequence resulted in shorter response times (L3-TRF: mean = 16.05, SD = 8.20). The pairwise comparison did not show a significant result ($p = 0.325 > 0.05$).

For the U3-TRF (used in the existing system, Ramirez Manuel et al., 2024): Compared to L1U2-TRF, which has the same view sequence but positions the top view on the left, L1U2 resulted in shorter response times (L1U2-TRF: mean = 16.78, SD = 8.93; U3-TRF: mean = 17.11, SD = 8.78). The pairwise comparison did not show a significant result ($p = 0.418 > 0.05$). Compared to B3-TRF, which shares the same view sequence but arranges all views on the upper side, U3 also resulted in shorter response times (B3-TRF: mean = 19.42, SD = 10.71). The pairwise comparison did not show a significant result ($p = 0.147 > 0.05$).

Overall, the data indicate that the layout factor has a significant impact on task response times. L3, the SDS with all views located on the left side of the field of view, had the shortest mean response time among all layouts. B3-TRF required the longest mean response time. L1U2, with one view on the left side and two views on the upper side, resulted in slightly shorter task times compared to U3, the SDS where all three views are positioned on the upper side. No significant effect of the sequence of views was found through pairwise comparison.

5.2. *Task success rate*

In the data preprocessing, 26 valid participants were included, resulting in a total of 1248 data points.One participant was excluded due to missing 6 data points (missing rate of 12.5 %). There are 26 (2.1 % of the data) outliers that were caused by accidental button presses. After excluding these outliers, 1222 valid data points remained. Trials with excessively long response times (over 50 s) were not considered as outliers but were defined as task failures.

The overall success rate demonstrated high consistency ($M = 0.97$, SD=0.169; range=0.96–0.99) across all layouts. Binary logistic regression analysis ($N = 1222$) detected no statistically significant differences in task success between layouts or sizes (all $p > 0.05$). The layout L1U2-RTF exhibited a numerically higher mean success rate (0.99 vs. 0.96)

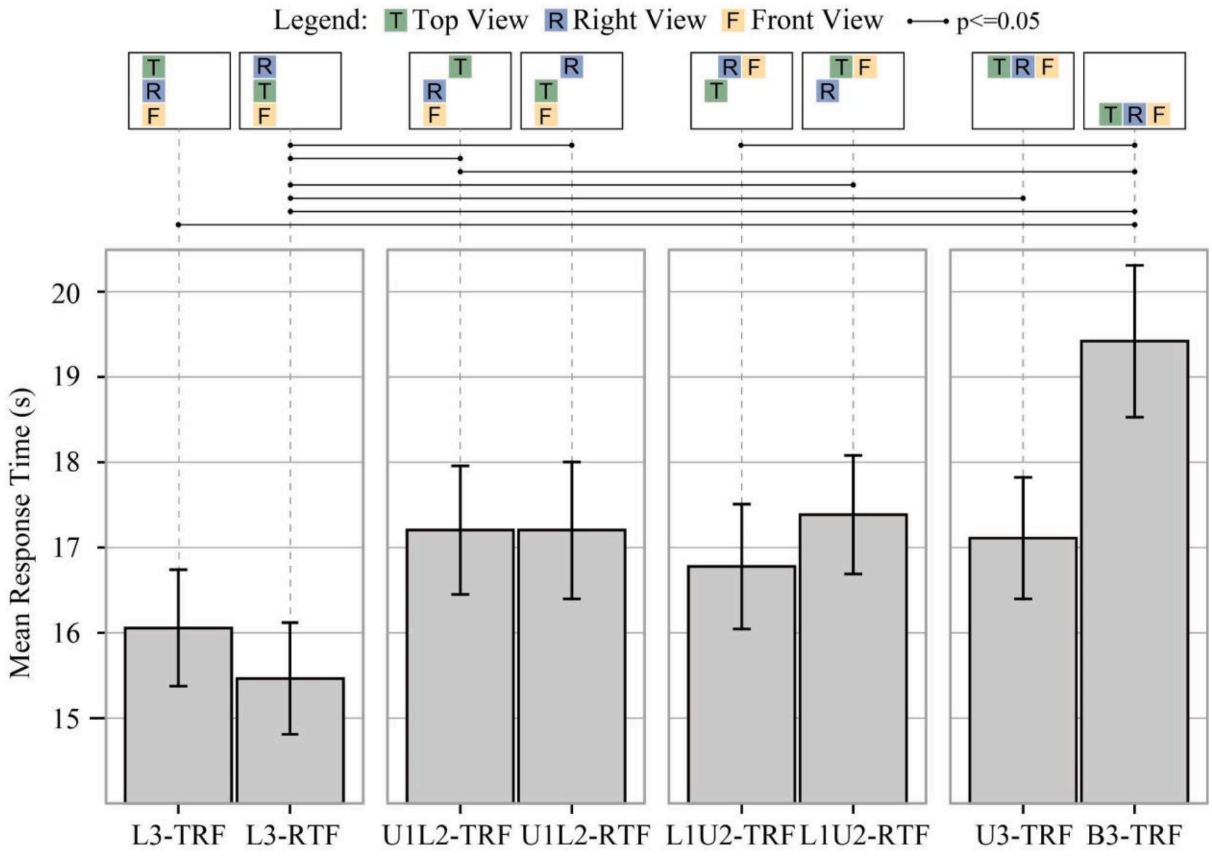


Fig. 6. Mean response times and significance results for all layouts. The error bars represent 95 % confidence intervals.

and a larger odds ratio (OR=3.237, 95 % CI: 0.64–16.30) compared to layout L3-TRF. Although the Hosmer-Lemeshow test demonstrated good model fit (χ^2=0.521, $p = 0.998$), the model's extremely low explanatory power (Nagelkerke R^2=0.01) reveals that the model explains only a very small proportion of the variance in task success.

Overall, the success rate data did not provide valuable insights. Given that the task represents a simplified version of the surgical simulation, these results should be considered for reference only.

5.3. Eye movement

In the data preprocessing, 26 valid participants were included, resulting in a total of 1248 data points. One participant was excluded due to missing 6 data points (missing rate of 12.5 %). After excluding the same outliers (54 outliers, 4.3 % of the data) as those identified in the response time analysis, 1194 valid data points remained.

5.3.1. Fixation duration and saccade count

Fixation event refers to the behavior where the eyes remain steady and focused on a specific point. Saccade event refers to the rapid eye movement from one fixation point to another. In the analysis of this section, fixation duration refers to the total amount of time spent on fixation behavior within a single task. Longer fixation duration under normal circumstances is likely to indicate more observation and cognitive processing time. Saccade count refers to the number of continuous saccade events recorded by the eye-tracking device within a single task.

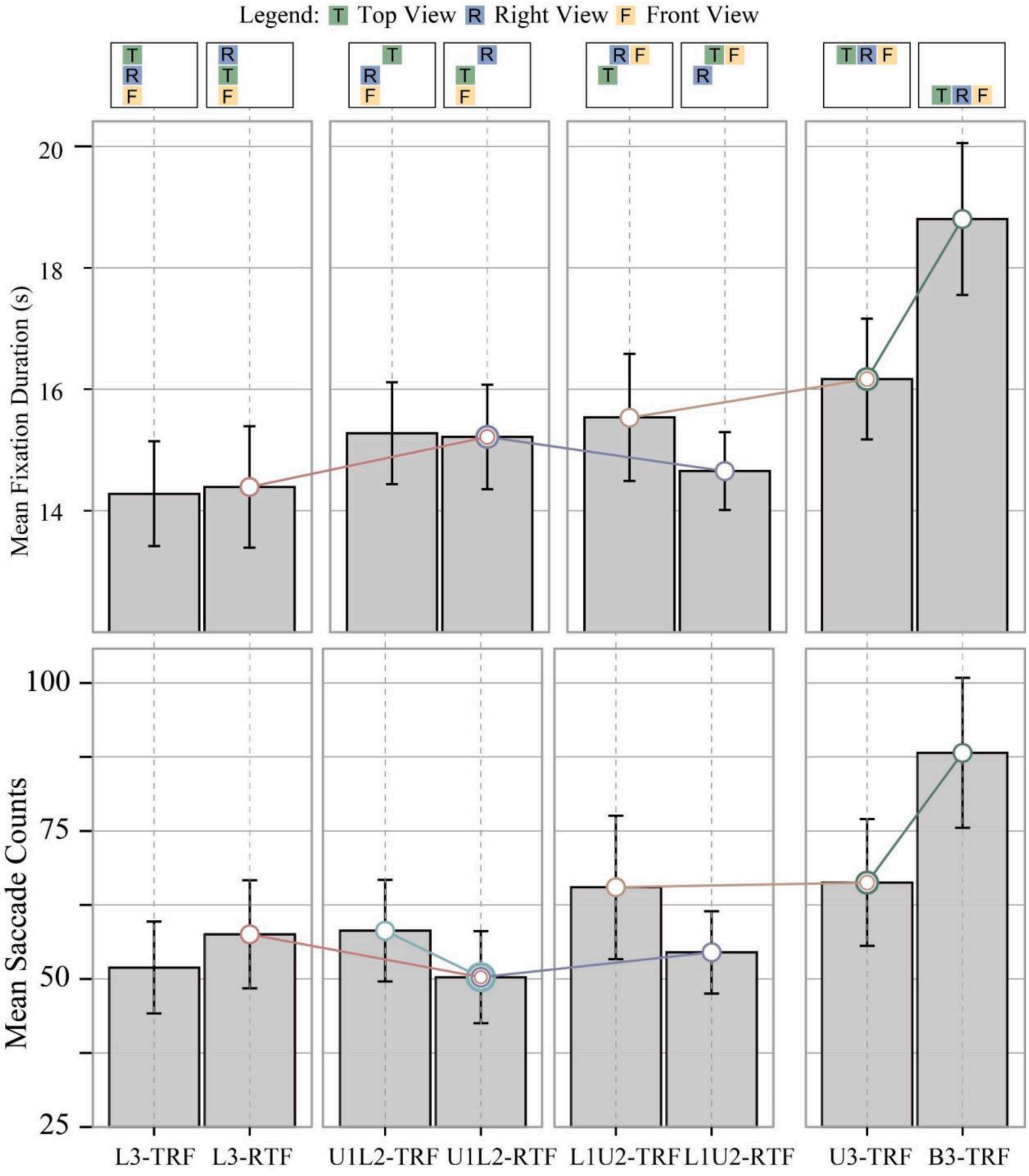


Fig. 7. Eye movement data across all layouts: Mean fixation duration and mean saccade counts. The error bars represent 95 % confidence intervals.

Higher saccade count under normal circumstances is likely to indicate a more complex or difficult visual search, or a greater number of information comparisons.

As shown in Fig. 7, the fixation duration and saccade count corresponding to the 8 layouts are presented. The figure of fixation duration shows that layouts L3-TRF, L3-RTF, and L1U2-RTF have shorter fixation durations. B3-TRF has the longest fixation duration, followed by U3-TRF. The figure of saccade count shows that L3-TRF,

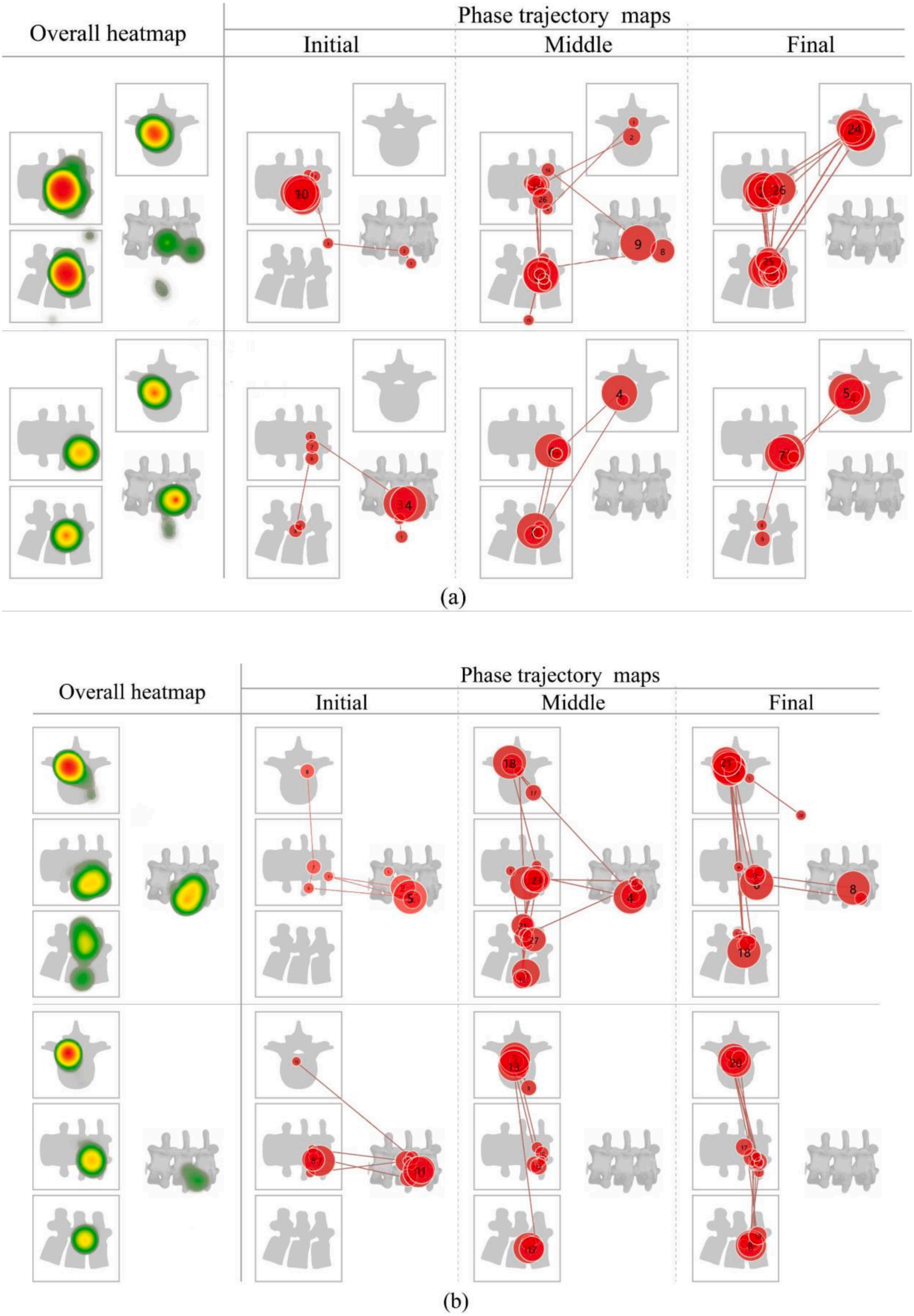


Fig. 8. Eye-tracking heatmaps and trajectory maps for four layouts: (a) U1L2-RTF, (b) L3-RTF, (c) U3-TRF, (d) L1U2-TRF.

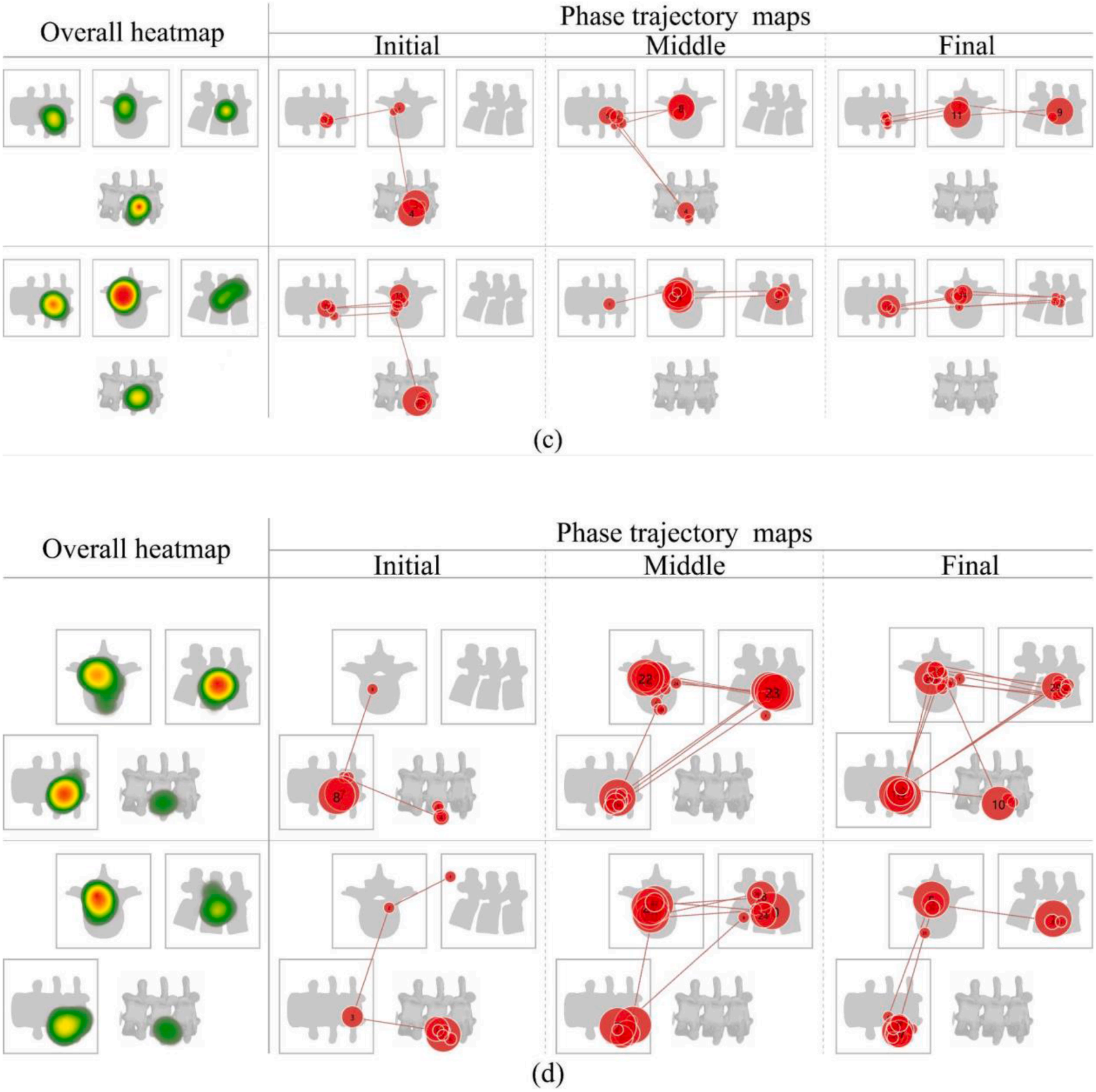


Fig. 8. (*continued*).

U1L2-RTF, and L1U2-RTF have shorter saccade counts. B3-TRF has the highest saccade count, followed by L1U2-TRF and U3-TRF.

The layout scheme analysis of the existing system is as follows. For the U1L2-RTF (used in the Caduceus System, 2019):

- Compared to L3-RTF: With the same SS, the spatial distribution L3 (with all three views on the left) resulted in a shorter fixation duration, which aligns with the response time analysis. U1L2 (with one view on the upper side and two views on the left) showed fewer saccade counts.
- Compared to U1L2-TRF: With the same SDS, the RTF sequence (with the right view on the upper side) resulted in fewer saccade counts.
- Compared to L1U2-RTF: Both of these layouts place the right view either separately on the upper side or on the left. The spatial distribution L1U2 resulted in a shorter fixation duration, while U1L2 showed fewer saccade counts.

For the U3-TRF (used in the existing system, Manuel et al., 2024):

- Compared to L1U2-TRF: With the same SS, the spatial distribution L1U2 (with one view on the left and two views on the upper side) resulted in a shorter fixation duration compared to U3 (with all three views on the upper side).
- Compared to B3-TRF: With the same SS, B3-TRF showed a longer fixation duration and more saccade counts.

Additionally, comparing L3-TRF and L3-RTF, with the same SDS, the fixation durations were very similar, which contrasts with the response time data in Section 4.1 (where L3-RTF has a shorter mean response time). L3-RTF showed more saccade counts than L3-TRF.

Although the above data provides some interesting findings, the relationship between fixation duration, saccade counts, and visual cognitive performance remains to be discussed. Further analysis will be conducted using visual search heatmaps and trajectory charts.

5.3.2. *Heat maps and trajectory maps*

Heat maps and trajectory maps were utilized to represent the participants' visual processing during the task. The color in the heat map corresponds to fixation duration, with longer fixations represented in red and shorter ones in green, providing a visual representation of attention distribution. The trajectory maps across initial, middle, and final phases show the paths and sequential patterns of eye movements during search stages.

As shown in Fig. 8, the heatmaps and trajectory maps are presented for four representative cases: (a) U1L2-RTF, (b) L3-RTF, (c) U3-TRF, and (d) L1U2-TRF. Among them, (a) and (c) are based on the layouts of the existing system. The view sequences in (a) and (b) are identical, as are those in (c) and (d). These search processes exhibit both similarities and differences, and have been selected as illustrative examples for comparative analysis.

We summarize the generalizable visual search patterns as follows: first, users reference the 3D model and 2D top view to target the screw placement position, then adjust the tool angle based on the front and right views, making refinements by referencing the 3D model or top view, and finally confirm the task after reviewing the views and model.

We observed differences in the search processes. In some trials, one view receives more attention; for example, the right view in (b) has the highest heat. In other trials, the time spent on each view is similar, as seen in the three views in (a) and the first row of (d) (abbreviated as (d)-1). In the initial phase when calibrating the screw, search behavior is divided into: primarily referencing the 3D model, primarily referencing the 2D top view, and combining both the model and the top view. In the middle phase when adjusting the tool angle, typical behaviors include: reviewing the top view or 3D model (e.g., Fig. 8(a)-1 and (b)-1), moving back and forth between the right and front views (e.g., (c)-2 and (d)-2), and quickly switching between the three views (e.g., (a)-2 and (d)-1). In the final phase, when the task is nearing completion, typical behaviors include: quickly moving between the three views (e.g., (a)-1, (b)-2, and (c)-2), and quickly moving between the three views and the model (e.g., (b)-1 and (d)-1).

Overall, the characteristics and patterns of visual search behavior provide intuitive evidence for the next discussion.

5.4. *Subjective evaluation*

We collected 27 valid questionnaires, where participants self-assessed task loads across six dimensions. The scores for these dimensions demonstrated good internal consistency (Cronbach's $\alpha = 0.786$). Fig. 9 shows the distribution of scores. Participants generally rated their performance highly, with high effort and low frustration levels. Physical demand scores were slightly lower than mental demand scores.

The questionnaire also gathered results from all participants comparing the dimensions (indicating which factor had a greater impact on task load), which were used to calculate the weight for each dimension. W_i represents the weight of the i th dimension, C_i represents the number of times the i th dimension won in all comparisons, and C_{total} is the total number of comparisons. W_i is calculated as the ratio of C_i to C_{total}. The number of wins and weights for the six dimensions are shown in Table 1. The data indicate that participants ranked the dimension weights in descending order as: physical demand, temporal demand, and mental demand, though the values were very close (Physical Demand = 0.180, Temporal Demand = 0.175, Mental Demand = 0.172).

In summary, the questionnaire results suggest that participants rated their performance positively and were actively engaged in the experiment. The evaluations for mental, physical, and temporal demands were similar, with physical demand slightly higher, indicating that the task's operational difficulty may have been slightly higher than its conceptual difficulty.

6. Discussion

After analyzing the results, we can revisit the research hypotheses and discuss them further. The test results of the hypotheses are as follows:

1. The spatial distribution of the layout influences task performance: The spatial arrangement of images in the field of view affects task response time. Specifically, placing all images on the left side of the visual field in a vertical arrangement significantly reduces response times.
2. The specific arrangement of the three views have an impact on task performance: The arrangement of the top, front, and right views affects eye-tracking metrics such as fixations and saccades.
3. Image size does not significantly affect task performance: Different image sizes did not result in significant differences in task response time or results.

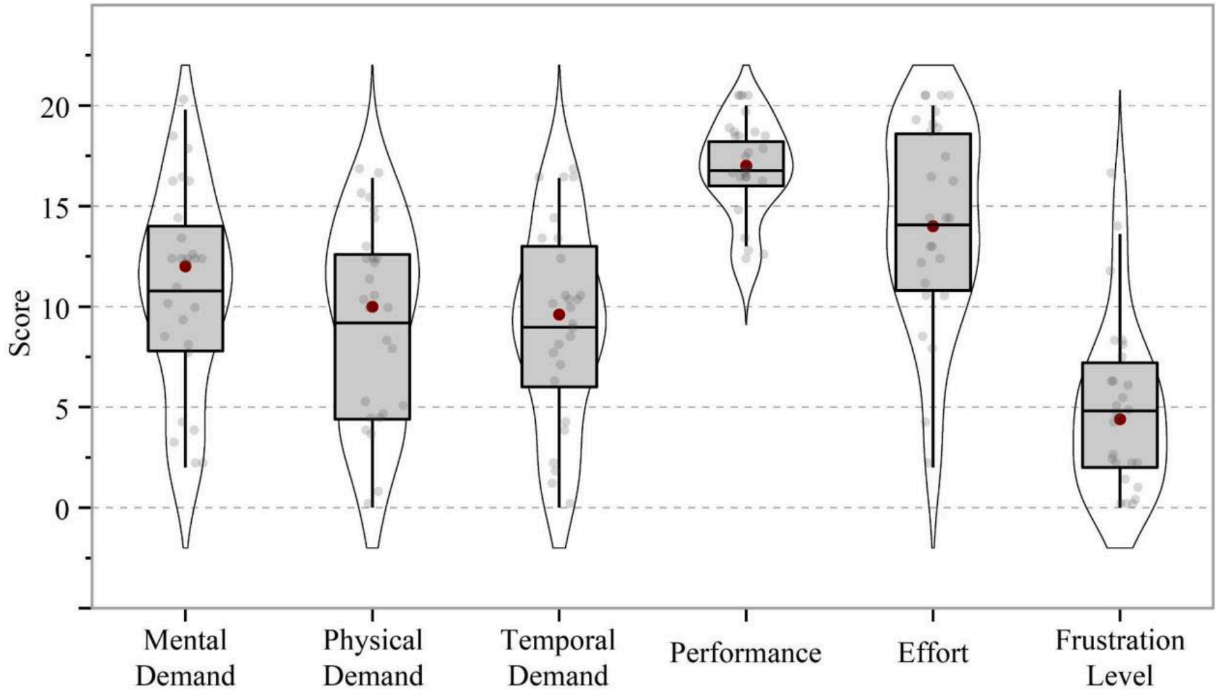


Fig. 9. Distribution plot of NASA-TLX scores across six dimensions, showing the median (red dot), mean (line), interquartile range (box), and the outlier range defined by 1.5 times the IQR.

Table 1
Number of pairwise wins for each dimension of the NASA-TLX scale across all participants.

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration Level
Count	70	73	71	84	72	36
Weight	0.172	0.180	0.175	0.207	0.177	0.089

6.1. *Suggested spatial distribution scheme*

We can interpret the results of the data analysis through visual cognition patterns, which shows that the scheme L3 significantly reduces response time. Studies of traditional display interfaces have shown that content above the horizontal line of sight and to the left of the screen typically attracts visual attention first (Li et al., 2017). Compared to left-to-right scanning, users tend to browse vertically faster, which is related to their experience with digital screens. In our experimental interface, the schemes L3, U1L2, L1U2, U3 and B3 represent combinations of different weight assignments for the layout sides, i.e., L3: all on the left side, U1L2: more on the left side than on the upper side, L1U2: less on the left side than on the upper side, U3: all on the upper side, and B3: all on the bottom side. Apparently, L3, which is distributed on the left side of the visual field and arranged vertically, has a significant advantage in visual search. As for the cases with poorer task performance, the layouts U3 and B3, which arrange three views horizontally, are less efficient compared to other layouts. B3 places views at the bottom of the visual field, which contradicts the user's natural tendency to focus on the upper side.

In addition, from the perspective of design aesthetics, the L3 spatial distribution aligns with the balance and symmetry of the interface layout, making it a visually appealing and rational design.

6.2. *Task-based multi-view*

To further discuss the relationship between multiple views and visual search, we first interpret selective visual attention and the visual attention phenomena observed in this task. In psychological research, this vision-guided behavior is widely known as selective visual attention (Desimone et al., 1995; Moore et al., 2017), which is one of the basic cognitive functions. One approach divides selective visual attention into top-down and bottom-up processes. Top-down attention is driven by internal signals, such as motivational state, while bottom-up attention is based purely on the physical salience of visual stimuli, such as color and brightness. In human-computer interaction, task-based visual search falls under top-down selective visual attention, as seen in the active inspection behavior of participants in this experiment. Participants were required to perform a selective visual search to identify areas where screws were not yet aligned. The search process includes observation, back-and-forth comparison, and rechecking, which can be observed in the eye-tracking data visualization. Therefore, our study suggests that when designing layouts with multiple views, related views should be placed next to each other to facilitate smoother visual movement. For example, the front and right views, which need to be referenced during the mid-task tool adjustment, should be positioned either horizontally or vertically adjacent.

Additionally, regarding the display continuity of the multiple views, in consultation with experienced surgeons, we agreed that adjusting the angle typically does not affect the placement point, but rechecking the point is necessary before drilling. Therefore, our study recommends retaining the top view during the procedure for real-time checks.

6.3. *Support for challenging views*

The ability to manipulate objects in 3D space is crucial for many clinical medical practices, although individual spatial abilities vary (Lufler et al., 2012; Rahmani et al., 2024; Drey et al., 2023). Perception and manipulation of 3D objects involve spatial cognitive abilities and mental rotation processes. The presentation of 3D objects can affect cognitive difficulty, with larger rotation angles resulting in longer cognitive processing times (Guillot et al., 2007; Meneghetti et al., 2016; Azarby et al., 2022; Hattab et al., 2021). Immersive VR environments can help students enhance their spatial understanding of 3D volumes (Alatta et al., 2017; Wang et al., 2018; Shen et al., 2019; Spiteri et al., 2018; Gittinger et al., 2023). Optimizing 3D object presentations and facilitating users' mastery are essential areas that merit further investigation.

We further discuss the user performance using eye-tracking metrics. In some trials, the eye-tracking heatmap shows that the fixation duration on the right view was the longest, indicating that participants spent more time processing the right view. We observe that RTF sequences in U1L2 and L1U2 layouts have fewer saccade counts than TRFs (U1L2-RTF $<$ U1L2-TRF, L1U2-RTF $<$ L1U2-TRF). This is an interesting finding because, from an aesthetic perspective, neither U1L2 nor L1U2 is symmetric or compact, whereas the advantage of RTF is that it places the right view alone on one side. For example, U1L2-RTF places the right view at the top, while L1U2-RTF places the right view on the left. These make the right view more spatially distinctive and visually prominent. This can be explained by the spatially oriented attention mechanism, where users tend to make biased selections of information based on specific locations (Wang et al., 2023; Hoffmann et al., 2019).

Considering the specific content of the right view, this pattern has important design implications. The right view displays a cross-sectional image of the spine, where lateral movement in the 2D view corresponds to forward-and-backward motion in the 3D manipulation, which makes hand-eye coordination more challenging and requires greater attention from the operator. Therefore, placing the right view in a visually prominent position is a better design choice from the perspective of selective visual attention mechanisms.

This design guidance facilitates the development of assistive interfaces for 3D object manipulation in immersive environments. Prioritizing complex spatial viewpoints within 3D objects can improve user task performance.

7. Conclusion

This study investigates HCI in medical VR interfaces by evaluating user performance and analyzing the effect of auxiliary image layouts on task efficiency and cognitive load. By analyzing response times, task success rates, and eye-tracking data, combined with descriptive statistics, ANOVA, and pairwise comparisons of layout schemes, we conclude that layout affects task performance and visual search. We propose the following design guidelines:

- Spatial Distribution of Layout: For a setup with three views and right-handed users, vertically arrange the views on the left side of the interface. Avoid excessive horizontal distribution of views. It is not recommended to place views below the operation object.
- Arrangement of Views: Ensure that all three views are always displayed, allowing users to view them at any time. Focus on content that is difficult to understand and operate, such as emphasizing the right view on the interface to address hand-eye coordination challenges. Place the views with related auxiliary information in spatially associated positions, such as setting the front and right views, which are used to assist in adjusting the tool angle, adjacent to each other for quick comparison.

In conclusion, this study focuses on the design of auxiliary image layouts in virtual environments, exploring the interface of a VR-based pedicle screw placement simulator. It provides detailed explanations and recommendations, thereby filling a gap in the research of digital interfaces for such systems. The research offers design guidance for creating assistive interfaces in immersive interactions, particularly for hand-eye coordination. However, the study also has limitations. For example, it did not investigate the performance of left-handed users, who may have unique interface layout needs. This remains a potential direction for future research.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (GPT-3.5) to improve readability and language in a few parts of the paper. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Lang Qin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kadek Ananta Satriadi:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Jiazhou Liu:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Yuhan Zhan:** Validation, Methodology, Investigation, Formal analysis. **Jiang Shao:** Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Peimeng Liu:** Visualization, Project administration, Data curation. **Zhiyong Chen:** Visualization, Project administration. **Yongtao Liu:** Supervision, Resources, Conceptualization.

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 paper was financially supported by the National Natural Science Foundation of China (No. 72001202) and the National Key R&D Program of China (No. 2024YFC3015005).

Data availability

Data will be made available on request.

References

Abich IV, J., Parker, J., Murphy, J.S., Eudy, M., 2021. A review of the evidence for training effectiveness with virtual reality technology. Virtual. Real. 25 (4), 919–933. https://doi.org/10.1007/s10055-020-00498-8.

Ahlstrom, U., Friedman-Berg, F.J., 2006. Using eye movement activity as a correlate of cognitive workload. Int. J. Ind. Erg. 36 (7), 623–636. https://doi.org/10.1016/j.ergon.2006.04.002.

Alatta, R.A., Freewan, A., 2017. Investigating the effect of employing immersive virtual environment on enhancing spatial perception within design process. ArchNet-IJAR Int. J. Archit. Res. 11 (2), 219. https://doi.org/10.26687/archnet-ijar.v11i2.1258.

Aoude, A.A., Fortin, M., Figueiredo, R., Jarzem, P., Ouellet, J., Weber, M.H., 2015. Methods to determine pedicle screw placement accuracy in spine surgery: a systematic review. Eur. Spine J. 24, 990–1004. https://doi.org/10.1007/s00586-015-3853-x.

Ariansyah, D., Erkoyuncu, J.A., Eimontaite, I., Johnson, T., Oostveen, A.M., Fletcher, S., Sharples, S., 2022. A head mounted augmented reality design practice for maintenance assembly: toward meeting perceptual and cognitive needs of AR users. Appl. Erg. 98, 103597. https://doi.org/10.1016/j.apergo.2021.103597.

Azarby, S., Rice, A., 2022. Understanding the effects of virtual reality system usage on spatial perception: the potential impacts of immersive virtual reality on spatial design decisions. Sustainability 14 (16), 10326. https://doi.org/10.3390/su141610326.

Birlo, M., Edwards, P.E., Clarkson, M., Stoyanov, D., 2022. Utility of optical see-through head mounted displays in augmented reality-assisted surgery: a systematic review. Med. Image Anal. 77, 102361. https://doi.org/10.1016/j.media.2022.102361.

Caduceus S System. Las Vegas (NV): SURGLASSES LLC, 2025. Available online: https://surglasses.com/en/surgery/caduceus-s-ar/ (accessed 25 April 2025).

Cao, Y., Qu, Q., Duffy, V.G., Ding, Y., 2019. Attention for web directory advertisements: a top-down or bottom-up process? Int. J. Hum.–Comput. Interact. 35 (1), 89–98. https://doi.org/10.1080/10447318.2018.1432162.

Chen, C.H., Chiang, S.Y., 2011. The effects of panel arrangement on search performance. Displays 32 (5), 254–260. https://doi.org/10.1016/j.displa.2011.05.003.

Coffey, D., Malbraaten, N., Le, T.B., Borazjani, I., Sotiropoulos, F., Erdman, A.G., Keefe, D.F., 2011. Interactive slice WIM: navigating and interrogating volume data sets using a multisurface, multitouch VR interface. IEEE Trans. Vis. Comput. Graph. 18 (10), 1614–1626. https://doi.org/10.1109/TVCG.2011.283.

Davis, C., Yoo, S., Clarkson, M.J., Thompson, S., 2023. Enhanced surgeons: understanding the design of augmented reality instructions for keyhole surgery. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2023. IEEE, pp. 123–127. https://doi.org/10.1109/VRW58643.2023.00031.

Deib, G., Johnson, A., Unberath, M., Yu, K., Andress, S., Qian, L., Osgood, G., Navab, N., Hui, F., Gailloud, P., 2018. Image guided percutaneous spine procedures using an optical see-through head mounted display: proof of concept and rationale. J. Neurointerv. Surg. 10 (12), 1187–1191. https://doi.org/10.1136/neurintsurg-2017-013649.

Deng, L., Wang, G., 2020. Quantitative evaluation of visual aesthetics of human-machine interaction interface layout. Comput. Intell. Neurosci. 2020. https://doi.org/10.1155/2020/9815937.

Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18 (1), 193–222. https://doi.org/10.1146/annurev-psych-122414-033400.

Doughty, M., Ghugre, N.R., Wright, G.A., 2022. Augmenting performance: a systematic review of optical see-through head-mounted displays in surgery. J. Imaging 8 (7), 203. https://doi.org/10.3390/jimaging8070203.

Drey, T., Montag, M., Vogt, A., Rixen, N., Seufert, T., Zander, S., Rietzler, M., Rukzio, E., 2023. Investigating the effects of individual spatial abilities on virtual reality object manipulation. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3581004.

Elmi-Terander, A., Burström, G., Nachabe, R., Skulason, H., Pedersen, K., Fagerlund, M., Gerdhem, P., 2019. Pedicle screw placement using augmented reality surgical navigation with intraoperative 3D imaging: a first in-human prospective cohort study. Spine 44 (7), 517–525. https://doi.org/10.1097/BRS.0000000000002876.

Frisk, H., Lindqvist, E., Persson, O., Weinzierl, J., Bruetzel, L.K., Cewe, P., Elmi-Terander, A., 2022. Feasibility and accuracy of thoracolumbar pedicle screw placement using an augmented reality head mounted device. Sensors 22 (2), 522. https://doi.org/10.3390/s22020522.

Gittinger, M., Wiesche, D., 2023. Systematic review of spatial abilities and virtual reality: the role of interaction. J. Eng. Educ. https://doi.org/10.1002/jee.20568.

Godzik, J., Farber, S.H., Urakov, T., Steinberger, J., Knipscher, L.J., Ehredt, R.B., Tumialán, L.M., Uribe, J.S., 2021. Disruptive Technology" in Spine Surgery and Education: virtual and Augmented Reality. In: Oper. neurosurg., 21, pp. S85–S93. https://doi.org/10.1093/ons/opab114.

Grobelny, J., Karwowski, W., Drury, C., 2005. Usability of graphical icons in the design of human-computer interfaces. Int. J. Hum. Comput. Interact. 18 (2), 167–182. https://doi.org/10.1207/s15327590ijhc1802_3.

Guillot, A., Champely, S., Batier, C., Thiriet, P., Collet, C., 2007. Relationship Between Spatial Abilities, Mental Rotation and Functional Anatomy Learning. Adv. Health Sci. Educ. 12, 491–507. https://doi.org/10.1007/s10459-006-9021-7.

Guo, F., Ding, Y., Liu, W., Liu, C., Zhang, X., 2016. Can eye-tracking data be measured to assess product design?: visual attention mechanism should be considered. Int J Ind Erg. 53, 229–235. https://doi.org/10.1016/j.ergon.2015.12.001.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: *Advances in* Psychology. North-Holland, Vol. 52, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hattab, G., Hatzipanayioti, A., Klimova, A., Pfeiffer, M., Klausing, P., Breucha, M., Bechtolsheim, F.V., Helmert, J.R., Weitz, J., Pannasch, S., Speidel, S., 2021. Investigating the utility of VR for spatial understanding in surgical planning: evaluation of head-mounted to desktop display. Sci. Rep. 11 (1), 13440. https://doi.org/10.1038/s41598-021-92536-x.

Heinrich, F., Bornemann, K., Polenz, L., Lawonn, K., Hansen, C., 2023. Clutch & Grasp: activation gestures and grip styles for device-based interaction in medical spatial augmented reality. Int. J. Hum. Comput. Stud. 180, 103117. https://doi.org/10.1016/j.ijhcs.2023.103117.

Hoffmann, E.R., Chan, A.H.S., Man, S.S., Chan, L.C.M, 2019. Determining the validity of the visual field principle for designing control/display arrangements. Appl Erg. 81, 102887. https://doi.org/10.1016/j.apergo.2019.102887.

Hussain, M., Park, J., 2024. Effect of transparency levels and real-world backgrounds on the user interface in augmented reality environments. Int. J. Hum.–Comput. Interact. 40 (16), 4265–4274. https://doi.org/10.1080/10447318.2023.2212218.

Ibbotson, M., Krekelberg, B., 2011. Visual perception and saccadic eye movements. Curr. Opin. Neurobiol. 21, 553–558. https://doi.org/10.1016/j.conb.2011.05.012.

Jensen, L., Konradsen, F., 2018. A review of the use of virtual reality head-mounted displays in education and training. Educ. Inf. Technol. 23, 1515–1529. https://doi.org/10.1007/s10639-017-9676-0.

Jud, L., Fotouhi, J., Andronic, O., Aichmair, A., Osgood, G.M., Navab, N., Farshad, M., 2020. Applicability of augmented reality in orthopedic surgery – a systematic review. BMC. Musculoskelet. Disord. 21 (1), 103. https://doi.org/10.1186/s12891-020-3110-2.

Li, F., Lee, C.H., Feng, S., Trappey, A., Gilani, F., 2021b. Prospective on eye-tracking-based studies in immersive virtual reality. In: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp. 861–866. https://doi.org/10.1109/CSCWD49262.2021.9437692.

Li, J., Yu, S.L., Liu, W., 2017. Cognitive characteristic evaluation of CNC interface layout based on eye-tracking. J. Comput.-Aided Des. Comput. Graph. 29 (7), 1334–1342.

Li, Z., Kiiveri, M., Rantala, J., Raisamo, R., 2021a. Evaluation of haptic virtual reality user interfaces for medical marking on 3D models. Int J Hum Comput Stud 147, 102561. https://doi.org/10.1016/j.ijhcs.2020.102561.

Liu, J., Prouzeau, A., Ens, B., Dwyer, T., 2020. Design and evaluation of interactive small multiples data visualisation in immersive spaces. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 588–597. https://doi.org/10.1109/VR46266.2020.00081.

Liu, J., Satriadi, K.A., Ens, B., Dwyer, T., 2024, October. Investigating the effects of physical landmarks on spatial memory for information visualisation in augmented reality. In: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, pp. 289–298. https://doi.org/10.1109/ISMAR62088.2024.00043.

Lohre, R., Bois, A., Athwal, G., Goel, D., 2020b. Improved complex skill acquisition by immersive virtual reality training: a randomized controlled trial. J. bone jt. surg., Am. https://doi.org/10.2106/jbjs.19.00982.

Lohre, R., Wang, J.C., Lewandrowski, K., Goel, D.P., 2020a. Virtual reality in spinal endoscopy: a paradigm shift in education to support spine surgeons. J. spine surg. 6 (Suppl 1), S208–S223. https://doi.org/10.21037/jss.2019.11.16.

Lufler, R.S., Zumwalt, A.C., Romney, C.A., Hoagland, T.M., 2012. Effect of visual–spatial ability on medical students' performance in a gross anatomy course. Anat. Sci. Educ. 5. https://doi.org/10.1002/ase.264.

Ma, L., Zhao, Z., Chen, F., Zhang, B., Fu, L., Liao, H., 2017. Augmented reality surgical navigation with ultrasound-assisted registration for pedicle screw placement: a pilot study. Int. J. Comput. Assist. Radiol. Surg. 12, 2205–2215. https://doi.org/10.1007/s11548-017-1652-z.

Mäkinen, H., Haavisto, E., Havola, S., Koivisto, J.M., 2022. User experiences of virtual reality technologies for healthcare in learning: an integrative review. Behav. Inf. Technol. 41 (1), 1–17. https://doi.org/10.1080/0144929X.2020.1788162.

Manuel De Jesus Encarnacion, R., Chmutin, G., Nurmukhametov, R., Soto, G.R., Kannan, S., Piavchenko, G., Montemurro, N., 2024. Integrating augmented reality in spine surgery: redefining precision with new technologies. Brain Sci. 14 (7), 645. https://doi.org/10.3390/brainsci14070645.

Meneghetti, C., Borella, E., Pazzaglia, F., 2016. Mental rotation training: transfer and maintenance effects on spatial abilities. Psychol. Res. 80, 113–127. https://doi.org/10.1007/s00426-014-0644-7.

Molina, C.A., Sciubba, D.M., Greenberg, J.K., Khan, M., Witham, T., 2021. Clinical accuracy, technical precision, and workflow of the first in human use of an augmented-reality head-mounted display stereotactic navigation system for spine surgery. Oper. Neurosurg. 20 (3), 300–309. https://doi.org/10.1093/ons/opaa398.

Moore, T., Zirnsak, M., 2017. Neural mechanisms of selective visual attention. Annu. Rev. Psychol. 68, 47–72. https://doi.org/10.1146/annurev-psych-122414-033400.

Morimoto, T., Kobayashi, T., Hirata, H., Otani, K., Sugimoto, M., Tsukamoto, M., Yoshihara, T., Ueno, M., Mawatari, M., 2022. XR (Extended reality: virtual reality, augmented reality, mixed reality) technology in spine medicine: status Quo and Quo Vadis. J. Clin. Med. 11. https://doi.org/10.3390/jcm11020470.

Nakatani, M., Ohno, T., Katagiri, Y., Nakane, A., Hashimoto, S., 2011, July. The layout for the user-friendly manual: case study on an internet set-up manual. In: International Conference of Design, User Experience, and Usability. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 40–45. https://doi.org/10.1007/978-3-642-21708-1_5.

Niu, Y.F., Zuo, H.R., Yang, X., Xue, C.Q., Peng, N.Y., Zhou, L., Jin, T., 2021. Improving accuracy of gaze-control tools: design recommendations for optimum position, sizes, and spacing of interactive objects. Hum. Factors Ergon. Manuf. Serv. Ind. 31 (3), 249–269. https://doi.org/10.1002/hfm.20884.

Orlosky, J., Itoh, Y., Ranchet, M., Kiyokawa, K., Morgan, J., Devos, H., 2017. Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease. IEEE Trans. Vis. Comput. Graph. 23, 1302–1311. https://doi.org/10.1109/TVCG.2017.2657018.

Previous versions of the Caduceus system. Interview video, 2019. Available online: https://www.youtube.com/watch?v=sQoQP7ehrLk (accessed 25 April 2025).

Rahmani, M.Z., Bukhari, A., Wiyono, N., Amru, K., Baharuddin, B., Nurhadi, A.A., 2024. Exploring links between visuospatial ability and anatomy learning in education: a bibliometric analysis and scientific mapping. Narra J. https://doi.org/10.52225/narra.v4i3.1095.

Renata, V., Li, F., Lee, C.H., Chen, C.H., 2018, October. Investigation on the correlation between eye movement and reaction time under mental fatigue influence. In: 2018 International Conference on Cyberworlds (CW). IEEE, pp. 207–213. https://doi.org/10.1109/CW.2018.00046.

Rose, T., Nam, C.S., Chen, K.B., 2018. Immersion of virtual reality for rehabilitation-Review. Appl Erg. 69, 153–161. https://doi.org/10.1016/j.apergo.2018.01.009.

Sakai, D., Joyce, K., Sugimoto, M., Horikita, N., Hiyama, A., Sato, M., Watanabe, M., 2020. Augmented, virtual and mixed reality in spinal surgery: a real-world experience. J. Orthop. Surg. 28 (3), 2309499020952698. https://doi.org/10.1177/2309499020952698.

Satriadi, K.A., Ens, B., Cordeil, M., Czauderna, T., Jenny, B., 2020. Maps around me: 3d multiview layouts in immersive spaces. Proc. ACM Hum.-Comput. Interact. 4 (ISS), 1–20. https://doi.org/10.1145/3427329.

Shen, C.W., Ho, J.T., Ly, P.T.M., Kuo, T.C., 2019. Behavioural intentions of using virtual reality in learning: perspectives of acceptance of information technology and learning style. Virtual Real. 23, 313–324. https://doi.org/10.1007/s10055-018-0348-1.

Shen, Z., Zhang, L., Li, R., Hou, J., Liu, C., Hu, W., 2021. The effects of color combinations, luminance contrast, and area ratio on icon visual search performance. Displays 67, 101999. https://doi.org/10.1016/j.displa.2021.101999.

Spiteri, T., McIntyre, F., Specos, C., Myszka, S., 2018. Cognitive training for agility: the integration between perception and action. Strength Cond. J. 40 (1), 39–46. https://doi.org/10.1519/SSC.0000000000000310.

Unema, P., Pannasch, S., Joos, M., Velichkovsky, B., 2005. Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. Vis. Cogn. 12, 473–494. https://doi.org/10.1080/13506280444000409.

Wang, P., Wu, P., Wang, J., Chi, H., Wang, X., 2018. A critical review of the use of virtual reality in construction engineering education and training. Int. J. Env. Res. Public Health 15. https://doi.org/10.3390/ijerph15061204.

Wang, X.S., Guo, F., Li, M.M., Tian, X.H., 2021. Effects of mobile news interface design features on users' gaze behaviours and behavioural performance: evidence from China. Behav. Inf. Technol. 40 (16), 1741–1754. https://doi.org/10.1080/0144929X.2020.1778089.

Wang, Z., Yu, R., Yang, L., 2023. Stimulus Motion: a Basic Influence Factor in Visual Search. In: Proceedings of the 2023 International Conference on Power, Communication, Computing and Networking Technologies, pp. 1–4. https://doi.org/10.1145/3630138.3630432.

Wirth, M., Gradl, S., Sembdner, J., Kuhrt, S., Eskofier, B.M., 2018, October. Evaluation of interaction techniques for a virtual reality reading room in diagnostic radiology. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, pp. 867–876. https://doi.org/10.1145/3242587.3242636.

Wolfe, J.M., Horowitz, T.S., 2017. Five factors that guide attention in visual search. Nat. Hum. Behav. 1 (3), 0058. https://doi.org/10.1038/s41562-017-0058.

Yoo, S., Blandford, A., 2022. Augmented reality and surgery: human factors, challenges, and future steps. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2022, pp. 459–461. https://doi.org/10.1109/VRW55335.2022.00101.

Zhou, X., Yin, G., 2021, April. Research on symbol color of automotive augmented reality head-up display. J. Phys. Conf. Ser. 1875 (1), 012013. https://doi.org/10.1088/1742-6596/1875/1/012013.

Lang Qin is a graduate student majoring in industrial design in China University of Mining and Technology. Her current research focuses on intelligent human-computer interaction, with a specialized emphasis on spatial perception, embodied interaction and cognitive ergonomics.

Kadek Ananta Satriadi, PhD, is a Human-Computer Interaction researcher and educator at Embodied Visualisation Group at Monash University, Australia. His research focus is visualisation and spatial computing technologies.

Jiazhou Liu, PhD, is an assistant lecturer in Human-Computer Interaction at Monash University, Australia. His research focus is exploring the effect of display layout in 3D space on data analysis tasks. His research area also covers spatial memory and embodied interaction.

Yuhan Zhan is a graduate student majoring in design in China University of Mining and Technology. Her current research focuses on intelligent human-computer interaction, encompassing areas such as cognitive ergonomics and natural interfaces.

Jiang Shao, PhD, is an assistant professor in the School of Architecture & Design at China University of Mining and Technology. His primary research fields are Brain-Computer Interface, Human-Computer Interaction, information visualization, and cognitive psychology.

Peimeng Liu is a graduate student majoring in industrial design in China University of Mining and Technology.

Zhiyong Chen is a graduate student majoring in design in China University of Mining and Technology.

Yongtao Liu, MD, is an associate chief physician in the Department of Spine Surgery at the Affiliated Hospital of Xuzhou Medical University. His primary research interests include spinal and spinal cord injuries.