Designing Cognitive 3D Immersive CAPTCHA for Enhancing Security of Virtual Reality Systems

Jeongeun Shim
Korea University
Dongyun Joo
Korea University
Hyemin Shin
Korea University
Gerard Jounghyun Kim
Korea University
Hanseob Kim
khseob0715@konyang.ac.kr

Konyang University

Research Article

Keywords: Usable Security, Cybersecurity, Authentication, CAPTCHA, Security in VR, User Experience, Immersive Tool, VR CAPTCHA

Posted Date: October 3rd, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6840630/v1

Additional Declarations: No competing interests reported.

Designing Cognitive 3D Immersive CAPTCHA for Enhancing Security of Virtual Reality Systems

Jeongeun Shim[1], Dongyun Joo[1†], Hyemin Shin[1†], Gerard Jounghyun Kim[1], Hanseob Kim[2*]

[1]Department of Computer Science and Engineering, Korea University, Seoul, 02841, Republic of Korea.
[2]Department of Artificial Intelligence, Konyang University, Daejeon, 35365, Republic of Korea.

*Corresponding author(s). E-mail(s): khseob0715@konyang.ac.kr;
Contributing authors: jemonshim@korea.ac.kr; dyjoo123@korea.ac.kr; mini9974@korea.ac.kr; gjkim@korea.ac.kr;
†These authors contributed equally to this work.

Abstract

With the advancement and maturing levels of related technologies, virtual reality (VR) is finding more uses in the industry, especially as a one-of-a-kind high-value generating system, and even for the general mass (e.g., entertainment and social activities). As such, there is an increasing concern for security for VR (as would be needed for any information system), which has seen only limited attention. One such area is the authentication procedures for VR users. One natural solution is to adopt the CAPTCHA as widely used in the 2D-based systems. However, recent advances in artificial intelligence (AI) have exposed critical vulnerabilities. Furthermore, conventional 2D CAPTCHA systems may be ill-suited for immersive 3D environments, often causing interaction inconvenience, disrupting user experience, and reducing immersion. On the other hand, there may be an opportunity for drastically improving the level of security by incorporating 3D spatial reasoning and interaction, which is generally regarded as more difficult for AI to crack, into the CAPTCHA authentication procedure in VR. In this paper, we propose "Cognitive 3D Immersive CAPTCHA" that leverages spatial interaction and user behaviors unique to virtual environments. As a foundation for this new approach, we set forth ten design principles tailored to CAPTCHA systems in VR. To explore and validate these principles, we design four prototype CAPTCHA systems as illustrative examples. The prototypes are evaluated to assess how well each satisfies the proposed principles, through a user study, expert

review, and robustness analysis. The findings support the potential of these systems, while also identifying certain vulnerabilities and suggesting directions for improvement.

Keywords: Usable Security, Cybersecurity, Authentication, CAPTCHA, Security in VR, User Experience, Immersive Tool, VR CAPTCHA

1 Introduction

Advances in virtual reality (VR) technology have introduced a new dimension to digital content, being able to immerse users in a 3D environment and offer experiences that are more realistic, natural, and engaging than conventional 2D working spaces. In particular, users are able to interact in 3D as they usually do in the real world. This means that users are able to carry out tasks in a fundamentally different manner to the usual 2D-oriented WIMP (window, icon, mouse, pointer) based interaction [1]. All of these have brought about a new breed of "serious" applications and experiences that are unique, one-of-a-kind, and high-value generating.

Most naturally, the seriousness calls for security needs. In fact, VR systems are more prone to security threats because of their loosely integrated multi-component system configuration. There are sensors, displays, controllers, and interface devices, on-headset and back-end computers, connected in a various fashion, all of which are potential weakest links, possibly exchanging and processing sensitive, confidential, and high-value information [2–4].

To address such a problem and the emerging need, this study investigates, among many possible security measures, the authentication method in the context of VR. Authentication is the process of verifying that a user is who they claim to be, and preventing unauthorized access. It serves as the first step in protecting user data, privacy, and system integrity [5]. Among various authentication methods, this research focuses on extending the CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart [6]), which can control access in non-login environments and act as a supplementary mechanism to traditional login-based authentication methods. CAPTCHA works by presenting challenges that are easy for humans to solve but difficult for machines, effectively distinguishing between humans and automated systems (or bots) [1, 6, 7].

In fact, several studies suggest that traditional CAPTCHA—mostly reliant on text input or 2D interactions—was not well-suited for VR environments, due to inconvenient interaction and disruption of the 3D and immersive experience [8, 9].

The main interaction modalities of VR, such as 3D gestures, manipulations, and spatial navigation [10] necessitate new interaction design for CAPTCHA-based authentication. This research proposes a cognitive 3D immersive CAPTCHA that combines the spatial characteristics of VR with cognitive abilities. The immersive nature and unique modalities of VR can offer users an intuitive problem-solving experience and promote enhanced demonstration of the associated cognitive abilities, such as memory encoding, storage, and retrieval processes [11, 12]. Conversely, bots would

encounter significant challenges due to the increased computational load and complexity of processing and reasoning about the 3D data [11, 13]. Therefore, the "Cognitive 3D Immersive CAPTCHA (also referred to as VR CAPTCHA)" as proposed in this work is expected to strengthen the level of security against bots while maintaining a positive VR experience. A user study and a formal expert review were conducted to evaluate the feasibility, robustness, and usability of four different designs of cognitive 3D immersive CAPTCHA in VR. The key contributions of this research are summarized as follows:

- Introduction of design principles for Cognitive 3D Immersive CAPTCHA in VR
- Implementation and evaluation of four different designs of cognitive 3D immersive CAPTCHAs to demonstrate their potential for superior security and user experience in VR
- Insights derived from a formal expert review by security and AI specialists, focusing on robustness, vulnerabilities, and potential improvements
- AI-driven robustness analysis of the proposed VR CAPTCHA

The paper is organized as follows: Section 2 presents a literature review of security in VR. Section 3 introduces the proposed design principles of CAPTCHA in VR, including the proposed concept of "Cognitive 3D Immersive CAPTCHA". To evaluate the security robustness, vulnerabilities, and user experience of the VR CAPTCHA, three types of validation were conducted. The design, procedure, and results of the experiment with the general public are described in Section 4. Expert interviews on security and AI are presented in Section 5. The robustness analysis, evaluating how well the VR CAPTCHA withstands attacks from AI-based automated systems, is reported in Section 6. Section 7 offers an in-depth discussion of the findings, and finally, Section 8 concludes the paper.

2 Related Works

2.1 Security in Virtual Reality

As VR technology continues to advance and becomes more prevalent and usability much improved, concerns about security and privacy in using VR systems have belatedly grown steadily. VR introduces unique security challenges due to its three-dimensional nature and novel interaction methods [1, 10, 14, 15], which require extensive user data and continuous sensor input to function effectively. This data includes motion tracking, environmental mapping, and sometimes biometric information like eye movements and facial expressions. The unique interaction modalities of VR create novel attack vectors, such as sensory spoofing, virtual environment manipulation, and unauthorized data access [2–4].

One of the primary security techniques is authentication, which verifies that a user is who they claim to be. Authentication is crucial for protecting user data and privacy, and for maintaining the integrity of virtual spaces [2, 4]. Traditional 2D screen-based authentication methods, such as passwords or PINs, are not appropriate (or at least quite inconvenient) for VR due to the inaccessibility to the physical keyboard and

the difficulty of inputting text using the virtual keyboard in 3D spaces [8, 9, 15]. This has led researchers to explore alternative authentication interfaces specifically designed for VR. As a result, eye tracking-based or (3D) behavioral biometrics have gained attention as potential solutions [9, 16–18]. Jiao et al. [18] proposed Medusa3D, which utilized reflexive eye responses to visual stimuli. Rupp et al. [9] developed a gesture-based authentication that utilizes ten gestures and virtual agents as interaction partners.

While biometrics may offer convenience and continuous authentication with some congruence to the interaction style of VR, they present significant weaknesses. They are vulnerable to spoofing or imitation attacks, where attackers use replicas or recordings to deceive the system. Once compromised, the individual has no recourse, is at heightened risk for identity theft [9, 14, 18, 19]. In addition, challenges with accuracy due to hardware limitations and implementing biometric systems can increase the complexity and (the already high) cost of VR devices [8, 18, 20]. These weaknesses indicate that there is still a need for more robust and more effective authentication methods tailored to the unique characteristics of immersive environments.

2.2 Traditional CAPTCHA and VR-Specific CAPTCHA

CAPTCHA has been widely used as an authentication method that verifies that a user is a human. It is designed to be easy for humans but difficult for machines to distinguish between users and bots [1, 6, 7]. Traditional CAPTCHAs involve tasks such as recognizing distorted text, identifying objects in images, or solving simple puzzles on 2D screens. However, advancements in artificial intelligence (AI) and machine learning (ML) have reduced the effectiveness of traditional CAPTCHAs. Modern AI algorithms, especially deep learning models, have achieved high accuracy in tasks like image and text recognition, enabling bots to solve the CAPTCHA tests [21–24].

One recent development to address the threats posed by AI is Google's reCAPTCHA version 3, a so-called invisible CAPTCHA. reCAPTCHA analyzes user behaviors—such as mouse movements, click times, browsing history, cookies—in the background to assess whether the interaction is human [1, 23, 25, 26]. While this approach improves user experience by eliminating the need for explicit challenges, it paradoxically fails to gain trust from many users. A study revealed that most users perceived invisible reCAPTCHA as insecure, due to the absence of visible security mechanisms [25]. Additionally, some argue that reCAPTCHA is unfair. Users gain only minimal and indirect benefits, such as a reduction in spam or fraud during their browsing experience while Google gains significant benefits from reCAPTCHA by collecting data from millions of users during their day-to-day browsing activities. Furthermore, ethical concerns are raised about using user data to generate training data for AI and ML [26].

Since traditional CAPTCHAs are designed for 2D computing environments (e.g., desktop, mobile touchscreen), they are not directly transferable to VR. There are also differences in interaction modalities and user experience expectations between 2D and VR environments. In a comparative study, users tended to prefer VR-specific CAPTCHAs that used mid-air interactions over traditional CAPTCHAs in VR environments [15]. Additionally, text-based CAPTCHAs require keyboard input,

which is impractical in VR. The difficulty of inputting text using a virtual keyboard in 3D spaces has been reported [8, 9, 15]. This necessitates the exploration of new CAPTCHA interfaces that are more suited to immersive VR. Some studies have begun to explore VR-specific CAPTCHA implementations. Li et al. [15] proposed vrCAPTCHA, which includes task-driven and bodily motion-based challenges using mid-air interactions. Hosaka et al. [21] introduced a stereoscopic text-based CAPTCHA for head-mounted displays, leveraging binocular disparity. Doken et al. [1] developed a gamified mixed reality CAPTCHA using occluded 3D objects, combining security tasks with interactive gaming elements.

These studies demonstrate innovative approaches to security in VR by leveraging unique 3D interactions and visual properties, not just to align CAPTCHA interaction to that of VR, but also to strengthen the level of security. However, most such approaches lack robustness evaluation against bots and may not fully address the balance between security and VR user experience.

2.3 Cognitive-Based CAPTCHA

Cognitive CAPTCHA involves tasks that require higher-level human cognitive processes such as reasoning, semantic interpretation, problem-solving, and decision-making [20]. Unlike traditional CAPTCHA, which focuses on perceptual challenges like recognizing distorted text or images, cognitive CAPTCHA presents challenges that exploit advanced cognitive abilities unique to humans, making it more difficult for bots to solve them. These include tasks involving semantic interpretation of texts or images, mathematical calculations, puzzle-solving, and other cognitive challenges [27].

For example, the Dynamic Cognitive Game (DCG) CAPTCHA, which requires users to play a simple moving object matching game. The dynamic and interactive nature of these games offers increased resistance to relay attacks [28]. Arkose Labs developed CAPTCHAs that present users with 3D challenges, such as rotating objects to their correct orientation or identifying anomalies within a scene, tapping into human cognitive abilities [29]. These CAPTCHAs necessitate semantic understanding, and furthermore, spatial reasoning/manipulation, which are challenging for bots to replicate due to the intricacies of mimicking both human cognitive and spatial functions [30].

Building upon these advancements, this research aims to apply cognitive CAPTCHA concepts to VR environments. The immersive nature of VR, the availability of 3D interactions, and VR-specific modalities—such as spatial navigation, handheld controllers, hand gestures, and head movements—can enhance user engagement and offer an intuitive problem-solving experience [8, 11, 15]. Humans can leverage spatial reasoning and motor skills in VR environments to solve a CAPTCHA naturally, while bots would face significant challenges due to the complexity of interpreting a 3D space and interacting within the space [13]. This increased complexity can improve security robustness.

By integrating cognitive challenges with immersive VR interactions, the proposed approach aims to explore an engaging and robust solution to enhance both user experience and security effectiveness in VR environments.

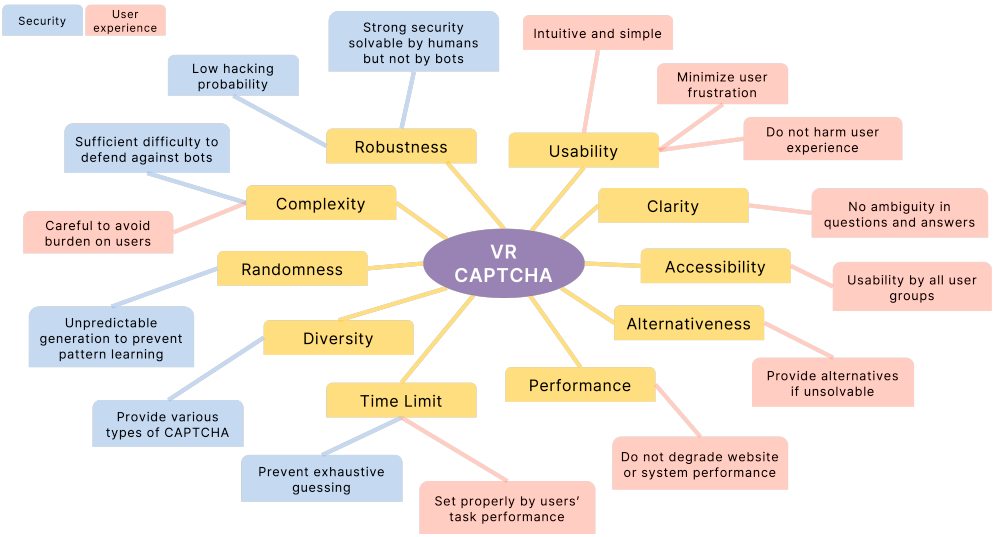


Fig. 1: Ten design principles of CAPTCHA to be used in VR systems (the yellow boxes). The blue boxes highlight security aspects, while the red boxes focus on user experience.

3 Cognitive 3D Immersive CAPTCHA

Given the current needs, our motivation and the current status as reviewed in the previous sections, we reiterate and assert ten design principles of CAPTCHA in VR along with the concept of cognitive 3D immersive CAPTCHA. Considering these principles, four VR CAPTCHAs were devised and evaluated as illustrative examples. In particular, the four designs involved particular attention to the principles related to user convenience and experience. The following subsections describe the principles and the development process of the four designs.

3.1 Design Principles of CAPTCHA in VR

Ten design principles of CAPTCHA in VR are shown in Figure 1 (the yellow boxes). The blue boxes refer to aspects of the extent of the security level - that is, the complexity in preventing patterns from being recognized and making it difficult for bots to solve the challenges. The red boxes emphasize the aspect of user experience, focused on making it easier and natural for users to "use" the system.

Some principles particularly go well with the cognitive 3D immersive concept. For example, Robustness may be enhanced by requiring high-level cognitive and 3D spatial abilities [20], while Complexity and Diversity (e.g., necessitating multiple cognitive and multi-sensory capabilities) increase the needed effort for attack bots by introducing higher computational demands and greater model complexity for processing 3D data [11, 13]. On the other hand, it is especially important to pay attention to the interaction design as VR systems, despite offering immersive and multimodal experiences

[11, 12], do have Usability concerns (e.g., unfamiliarity, cybersickness, manipulability, cumbersomeness). Accessibility may also be a concern for users with visual or auditory impairments, and provide options for those who may face challenges operating handheld controllers due to physical limitations.

The following outlines key design principles, highlighting how each contributes to ensuring security and usability in VR CAPTCHA.

- **Robustness—Ensuring Strong Security Solvable by Humans but Not by Bots:** The fundamental purpose of CAPTCHA is to distinguish genuine user access from those by bots. With recent advancements in DL algorithms, traditional CAPTCHA tests—such as distorted text recognition or image selection—can be easily decoded [1, 21, 22, 24]. Challenges should leverage high-level cognitive abilities and sensory interpretation skills, which AI or automated algorithms cannot easily replicate [30].
- **Usability—Easy to Use and Learn:** Designing CAPTCHA challenges to minimize inconvenience or stress during the solving process is essential [26]. Note that 2D interfaces are mostly standardized and familiar; VR/3D interfaces are generally not. VR devices generally have much less accuracy than 2D counterparts, such as the mouse and touchscreen [31]. Key interaction issues, such as interface familiarity, cybersickness, manipulability, and cumbersomeness, must be taken into account. Usability plays a critical role in shaping the overall user experience.
- **Complexity—Providing Tasks with Right and Sufficient Difficulty to Deter Bots without Burdening Users:** CAPTCHA should possess enough complexity to prevent automated scripts or bots from pattern learning. However, this complexity must be balanced between security and usability so as not to impose an excessive burden on users [21, 26, 32]. For instance, adding background noise or unique textures can make it difficult for ML-based decoders while still being recognizable to humans [33–35]. Complexity is a key factor in enhancing security, but it requires careful design to avoid complicating the usability aspect [1]. For example, unsolvable CAPTCHA can increase website abandonment rates and degrade user experience [27, 33]. Therefore, challenges should be intuitive, simple, and designed to be solved quickly [8, 25, 36].
- **Clarity—Minimizing Confusion with Clear Questions and Answers:** CAPTCHA challenges should be clearly communicated to users, leaving no ambiguity regarding the correct answer. Ambiguous or open-to-interpretation challenges can increase failure rates and decrease service satisfaction [22, 24]. Therefore, questions and instructions should be provided in a clear and consistent manner, and answers should be unambiguous. Examples or additional explanations can be included if necessary.
- **Randomness—Introducing Randomness in Challenge Generation to Prevent Pattern Learning:** Randomly generating CAPTCHA challenges is essential to prevent automated attacks from learning and predicting patterns [20, 24, 35]. Challenges should be distributed uniformly, ensuring they are independent of individual users and their responses. Similarly, answers must be randomized and uniformly distributed to eliminate predictability. To further enhance security, there

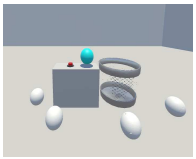
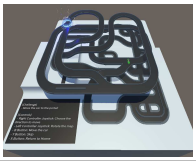
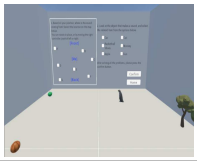
should be no statistical correlation between challenges and their answers, making it difficult for automated attacks to infer relationships or patterns.

- **Diversity—Requiring Multiple Human Abilities to Make Bots Difficult:** Requiring various human cognitive abilities and multi-sensory capabilities in each challenge raises the difficulty for bots due to increased computational load [24, 30]. Tasks that involve complex visual perception, auditory interpretation, or spatial reasoning are inherently easier for humans but computationally intensive for bots. Additionally, incorporating randomness is an effective way against pre-trained bots [33, 35].
- **Accessibility—Designing for Usability Across All User Groups:** CAPTCHA should be designed to ensure accessibility for all users [6, 8, 20, 22, 26]. For instance, providing hearing-based CAPTCHA for visually impaired users, vision-based CAPTCHA for hearing-impaired users, and adjusting color contrast for color-blind users can enhance service inclusivity. Moreover, offering interfaces that enable one-handed operation or simple button can improve usability for users with physical limitations in manipulating controllers.
- **Alternativeness—Providing Alternatives When Users Struggle with Challenges:** A design pattern that expects multiple attempts from users is a problem. Alternative options or different types of CAPTCHAs should be provided if users cannot solve the initial challenge [26]. The alternative CAPTCHA should match the difficulty level of the previous challenge [20]. This approach helps reduce user frustration and encourages continued use of the service. Moreover, connecting users to customer support after a certain number of failed attempts can be considered.
- **Performance—Optimizing to Avoid Negative Impact on System Performance:** CAPTCHA should not degrade the loading time or responsiveness of websites or systems. Avoiding heavy scripts or large resources and minimizing performance impact through optimized code and rendering protocol are essential [22, 24, 35, 37]. This strategy is important for preventing user abandonment and maintaining service quality.
- **Time Limit—Enhancing Security by Encouraging Timely Challenge Resolution:** Setting a time limit for solving CAPTCHA prevents brute-force attacks or random guess attempts by limiting the time available for exhaustive guessing. Bots face significant challenges in achieving high accuracy within a reasonable time due to the computational load and the complexity of processing data [7, 13, 24]. Time limits should be set considering the difficulty of the challenge and the average solving time.

3.2 Proposed Immersive CAPTCHA Prototypes

Prototypes of the proposed CAPTCHAs were developed in four distinct types, complying with ten design principles (as much as possible, but to be re-evaluated in later sections) and applying VR-specific interactions. Table 1 shows the operational screen, success conditions, immersive attributes, 3-dimensional attributes, and cognitive abilities of the proposed CAPTCHAs. The CAPTCHAs were implemented using Unity 3D (*Ver.* 2022.3.17f1) with C#.

Table 1: Synopsis of the proposed Cognitive 3D immersive CAPTCHA for VR system — 3D Shape Match and Gathering, 3D Wayfinding, 3D Assembling, and 3D Hearing.

	3D Shape Match and Gathering	**3D Wayfinding**	**3D Assembling**	**3D Hearing**
Task scene				
Success conditions	Finding two objects most similar to the target object (at least one object is identical to the target object) - evaluate by no. of objects found	Moving the car from the starting point to the destination along the pathway - completion of the route through the destitation in time	Completing an assembly from a pool of parts with time limit - no. of correct parts in right poses	Identifying the location of the object emitting given sound and selecting a word related to that object; no. of correctly identified sound source objects and associated words
Immersive and 3D attributes	3D selection, Viewpoint change, Small movement with handheld controllers	3D selection, Viewpoint change, Remote object manipulation with handheld controllers	3D selection, Viewpoint change, Small movement with handheld controllers	Perceiving 3D sound direction, Viewpoint change, Small movement with handheld controllers
Cognitive abilities	Perceptual similarity judgment, Visual memory, Spatial perception ability, Motor control ability	Spatial perception ability, Path planning ability	Shape and pattern recognition ability, Spatial perception ability, Motor control ability	Auditory localization ability, Spatial perception ability, Multi-sensory (auditory and visual) ability, Visual-verbal reasoning ability

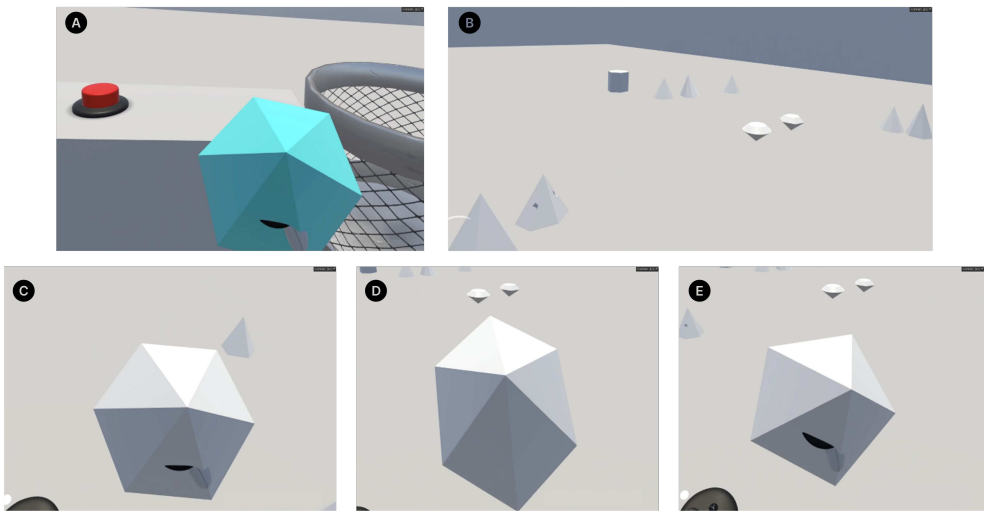


Fig. 2: 3D Shape Match and Gathering— Given a target object (in blue), the user must navigate and find the two objects with the same/similar shape/size. The top right shows the virtual space where the candidate objects are scattered. The user has to approach, examine/manipulate objects at angles (bottom row), and finally fetch the matching objects. The bottom row illustrates the user examining the candidate objects for a shape match.

3.2.1 3D Shape Match and Gathering

In this task, the user is situated in a small room with many 3D objects scattered in it, and is also given a target 3D object and a basket. Then the user is asked to navigate through the room, find the two objects most similar to the target, grab them, and bring them back to insert into the basket. The user uses 6DoF handheld controllers in both hands. The joystick on the controller is used to navigate the virtual environment, and an object can be grabbed/released by the grip button. Thus, an object can be held in each hand simultaneously.

This task leverages the human ability to judge 3D shape similarity not merely based on "geometry", but on "perceptual" similarity. While geometric similarity refers to measurable, often mathematically defined features (e.g., vertex matches, surface distances), perceptual similarity is a broader, cognitively informed notion that subsumes geometric similarity and extends to subjective dimensions (e.g., usage context, cultural background, and functional features) grounded in human visual and conceptual systems [38, 39]. For instance, humans may perceive two chairs with different leg shapes as similar due to their shared function and typical pose, even though their geometries differ substantially.

In our implementation, the 15 candidate 3D objects were designed with subtle variations in shape and size to different extents (see Appendix A.1). Small geometric

differences that are easily accepted by humans may be misclassified by AI systems that rely on feature embeddings lacking perceptual grounding [38, 39].

A pilot study was conducted via Amazon Mechanical Turk to assess perceptual similarity among the candidate objects. Based on participants' similarity ratings, the top two highest-ranked objects were selected as the most perceptually similar matches and subsequently used as CAPTCHA challenge items. Further details are provided in Appendix A.2. Note that in all cases, the correct answers would subsume geometrically identical candidates and also other qualitatively similar ones. Success is achieved if the participant finds and fetches the correct objects. The number of attempts per challenge is limited to one to prevent attacks such as brute-force attacks and random guess attacks caused by multiple tries.

3.2.2 3D Wayfinding

In this task, the user is given a complex 3D miniature roadway in front of him, and is asked to guide the car from the starting to finishing position in a designated time (see Figure 3). The road is composed of multiple layers; thus, some parts are obscured depending on the viewing angle, making it difficult to grasp the entire road structure and connections at once.

As was with the 3D Shape Match and Gathering, the user uses 6DoF handheld controllers in both hands. The joystick on the controller is used to navigate and move about in the virtual environment, and the car can be controlled to change its direction (turn in 4 directions clockwise or counter) with the joystick, and move by a discrete amount (to the next "waypoint" which are not visibly marked - see Figure 3) along the road by a button press. The car is not allowed to move off the track. The other second controller can be used to hold/manipulate the track itself (to set oneself in the right viewpoint). By repeating these steps and arriving at the destination, the participant succeeds. The number of attempts per challenge is limited to one to prevent multiple tries.

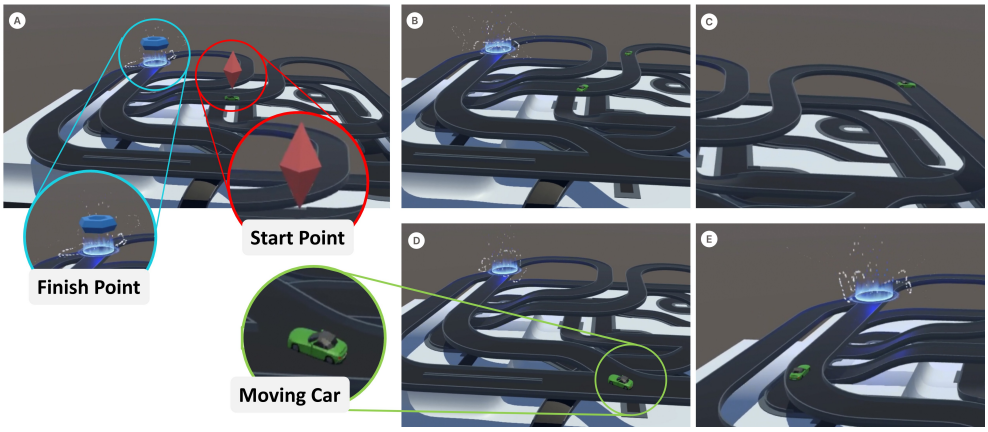


Fig. 3: 3D Wayfinding—Given a complex 3D miniature roadway, the user is asked to guide the car from the starting to the finishing position within a designated time.

3.2.3 3D Assembling

In this task, the user is given a number of parts and asked to complete a 3D assembly by choosing the correct parts and placing them in the correct position and pose within a designated time. The process is guided by a half-transparent 3D template (that shows the look of the completed assembly).

The interface is the same as with the 3D Shape Match and Gathering task. Users hold 6DoF handheld controllers in both hands. The joystick is used for navigation within the virtual environment, while objects can be grabbed and released using the grip button. Once grabbed, objects can be moved and rotated relative to the controller. Note that all objects are within the arm's reach. Navigation within the task scene is mostly needed to change the perspectives to look for the correct parts.

A number of parts, from which the correct parts must be selected and fetched, are scattered on the table (see Figure 4). Incorrect parts differ from the right ones in their shapes, sizes, colors, and textures (e.g., all parts must have the same color or texture). The participant selects the blocks that compose the target assembly by positioning and rotating them into the 3D template. After completion, upon pressing the finish button, the assembly completion score (the part is deemed in the correct pose with an arbitrarily set error threshold) is displayed in front, allowing participants to check the value. The number of attempts per challenge is set up to three times.

3.2.4 3D Hearing

In this task, the user is situated in a 3D virtual space, then given a number of 3D sound bytes and asked to point and designate the object from which the sound is heard to come from (e.g., find the sound source). This challenge involves multiple tasks requiring spatial sound localization and visual reasoning. Participants can solve it by rotating their head or adjusting their position in place using the joystick on the right controller. Seven objects are placed all around the user, and one of them repeatedly plays a sound. The positions of the seven objects are given as choices, and

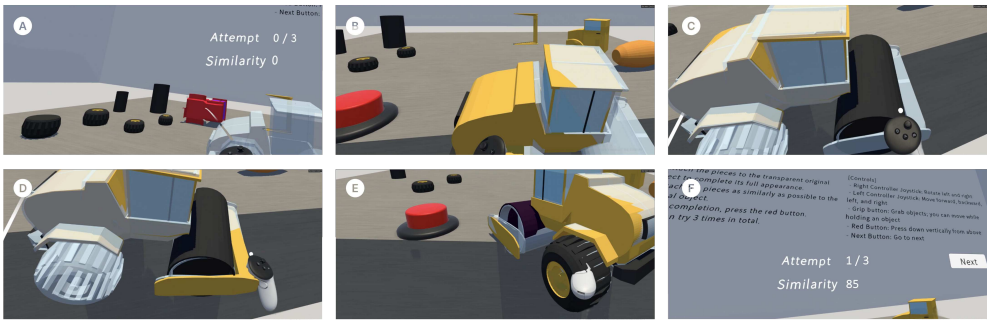


Fig. 4: 3D Assembling—The user is given a number of candidate parts and asked to complete a 3D assembly by choosing the right parts and placing them in the right position and pose within a designated time. The process is guided by a half-transparent 3D template (that shows the look of the completed assembly).

the participant selects one position of the object emitting the sound based on their position. Note that the object emitting the sound has nothing to do with the sound itself. For instance, a cat's meowing sound may be coming from a carrot. Despite the situation, the user is asked to just locate the sound source with no regards to the sound-to-object relationship.

Then another task, which combines the auditory, visual, and cognitive modalities, solvable only through in-place rotation interaction, is added to increase/adjust the difficulty. Once the sound source object is identified, this task requires selecting a word related to the object from given six choices. Thus, in the previous example of a carrot making the meowing sound, a proper choice might be a "rabbit". Success is achieved if the participant correctly answers both tasks. The number of attempts per challenge is limited to one to prevent multiple tries.

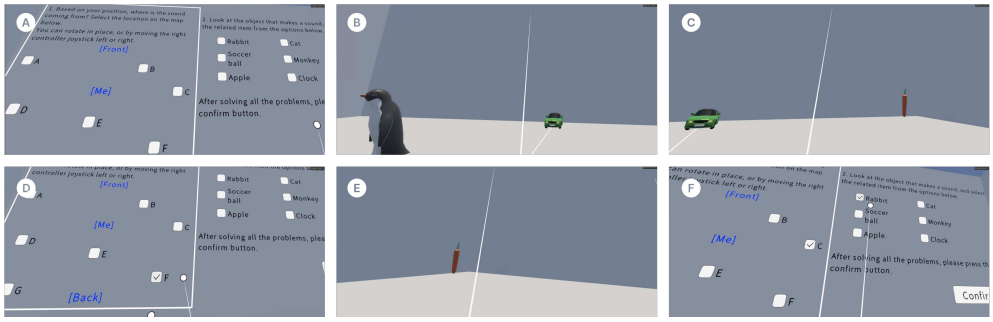


Fig. 5: 3D Hearing—The user, situated in a 3D virtual space, is given a number of 3D sound bytes and asked to identify the location of the object emitting the sound. Subsequently, the user selects a semantically related word associated with the identified object.

4 Experiment and User Evaluation

This section presents the human-subject study conducted to evaluate and validate through comparisons of the four proposed VR CAPTCHAs, especially with respect to the ten aforementioned design principles. The experiment follows a 1×4 within-subject design (the factor being the CAPTCHA type). It should be noted that the aim of this experiment was not to determine which CAPTCHA type is superior, but rather to assess the relative usability and pros/cons among the four.

4.1 Dependent Variables

To assess the effectiveness of the proposed CAPTCHAs, the following objective and subjective dependent variables were measured: task performance, effectiveness based on design principles, and overall VR user experience.

4.1.1 Task Performance

While participants solve CAPTCHA challenges during the given 5-minute period for each treatment (detailed in Section 4.3), the experimental system recorded the following task performance data.

- **Solving Time** is measured as the average time to complete each trial, recorded in seconds.
- **No. of Trials** is the total count of trials within the given 5-minute time limit, regardless of whether the trials are successful or not.
- **Accuracy** is calculated as the percentage of correct answers relative to the total number of trials. However, for 3D Assembling, accuracy is defined as the assembly completion score.

4.1.2 Effectiveness Based on Design Principles

This evaluates whether the proposed VR CAPTCHA effectively fulfills its role of distinguishing humans from bots and whether it can be reliably used for secure authentication. A total of ten questions, each based on one of the ten CAPTCHA principles outlined in Section 3.1, are administered via a questionnaire at the end of each treatment. The questions include: *Robustness, Usability, Complexity, Clarity, Randomness, Diversity, Accessibility, Alternativeness, Performance, and Time Limit.* Responses were collected using a 7-point Likert scale (1: not at all; 7: very much). The detailed questionnaire can be found in the Appendix B.1.

4.1.3 Overall VR User Experience

Since the proposed CAPTCHAs are used in VR with VR interfaces (requiring 3D spatial reasoning and interaction), albeit relatively short, the usability and experience are still important for the overall VR system user experience. The impact of CAPTCHA use on the VR user experience is measured across several dimensions as follows:

- **Task Workload** is measured by assessing the physical and mental effort required to solve the CAPTCHA in a virtual environment using VR devices. The SIM-TLX [40] questionnaire is used, which includes ten items—*Mental Demand, Physical Demand, Temporal Demand, Frustration, Complexity, Stress, Distraction, Strain, Difficulty, and Presence*—rated on a 21-point Likert scale ranging from 0 to 100, with intervals of 5 points.
- **System Usability** refers to the overall usability of the VR system, which presents a CAPTCHA challenge and provides functionality for solving it in the VR environment. This is measured using the System Usability Scale (SUS) questionnaire [41], which consists of ten items rated on a 5-point Likert scale. This evaluation yields a total usability score ranging from 0 to 100.
- **Flow/Interrupt** refers to the continuity of the main VR content, even after completing CAPTCHA challenges. In practical applications, CAPTCHA tasks are generally unrelated to the main content, and excessive focus on them can disrupt the user experience. Thus, CAPTCHA should be designed to minimize interference with

the flow of the main content. Drawing on related work [42–44], four questions are developed and evaluated using a 7-point Likert scale, where higher scores indicate greater disruption to the main content. To clearly simulate realistic conditions, the questionnaire explicitly instructs participants to imagine using a VR service with immersive content, where they are occasionally required to complete CAPTCHA for authentication. The detailed questionnaire can be found in the Appendix B.2.

- **Cybersickness** is a critical factor in evaluating the VR user experience. Users may also experience cybersickness while engaging in CAPTCHA in VR. However, since the tasks did not involve extensive or prolonged navigation (which is the main cause of cybersickness), we expect that the negative impact of any cybersickness would have been rather minimal. As such, administering a full questionnaire is avoided to reduce participants' fatigue. Instead, participants rated four primary symptoms—*Discomfort, Nausea, Oculomotor, and Disorientation*—using a 7-point Likert scale (1: not at all, 7: very much) [45].

4.2 Participants

Thirty-two participants (15 females and 17 males) were recruited from the university community. Their ages ranged from 19 to 34 ($M = 24.78$; $SD = 3.36$). All participants had normal hearing and normal/corrected vision, with 6 participants wearing glasses. Their VR familiarity was checked using a 7-point Likert scale (1: not at all, 7: very much), resulting in an average of 3.34 ($SD = 0.97$).

Using G*Power [46], it was determined that the minimum sample size required was 24 for the four within-factors in a repeated measures design, with an effect size of 0.25 (medium), a significance level of 0.05, and a power of 0.81 (critical $F = 2.737$). A post hoc analysis with the actual sample of 32 participants yielded a power of 0.92, indicating sufficient statistical reliability.

This study was approved by the university's institutional review board (IRB No. KUIRB-2025-0054-01). Participants consented to the recording of behavioral data, survey completion, and participation in a semi-structured interview. Upon completion of all experiments, they received a monetary compensation of about $20.

4.3 Procedure

Upon arriving at the experimental site, participants first received an introduction to the experiment and completed the written consent form. A pre-survey was given to record and collect the participants' current physical conditions, level of familiarity with VR, device usage experience, and demographic information.

With the experimenter's assistance, participants wore the VR headset (Meta Quest 3) while seated. Each condition was preceded by a short practice session for the assigned CAPTCHA. Participants were instructed on how to use the handheld controllers to interact with and solve the CAPTCHA. After participants became familiar with the VR device and task procedures, the main session began.

The main session lasted a 5-minute, during which participants completed as many CAPTCHA problems as possible. Since only one problem was presented at a time,

after solving each one, a new problem was randomly selected from a pool and presented. At the end of the 5-minute session, the virtual objects disappeared, marking the conclusion of that session. Participants then removed the VR headset and completed follow-up surveys on a PC. After a 5-minute break, they proceeded to the next treatment, with the presented order of treatments being counter-balanced.

Although the difficulty of each CAPTCHA varied, participants spent an equal 5 minutes on each type, ensuring equal exposure across all conditions. After completing all treatments, participants filled out a post-experiment survey and took part in a brief interview. The entire experiment took approximately one hour.

4.4 Results

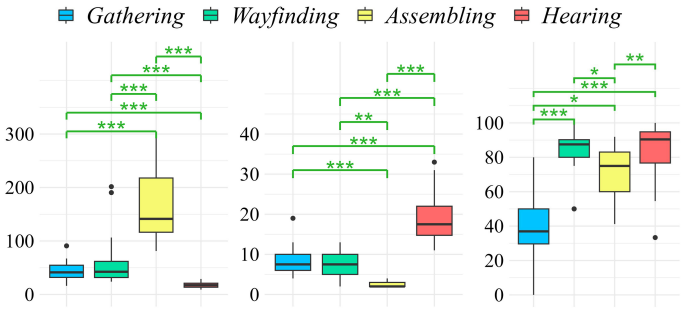
This experiment followed a within-subjects design with four treatments. To determine whether the data met the parametric assumptions, the Shapiro-Wilk test for normality and Mauchly's test for sphericity were used. If both tests showed p-values above the significance threshold, one-way repeated measure ANOVA was used, followed by Tukey's HSD for post-hoc analysis. Conversely, if either test failed, the Friedman test was applied instead, and pairwise comparisons, i.e., post-hoc analysis, were conducted using the Conover test with Bonferroni correction. The significance level for all statistical analyses was set at 5%.

In this subsection, CAPTCHA types are abbreviated as follows: *Gathering, Wayfinding, Assembling, and Hearing.*

4.4.1 Task Performance

Since none of the variables met the parametric assumptions, the Friedman test was applied. The statistical results are summarized in Figure 6 and Table 2.

- **Solving Time per (Successful) Trial:** *Assembling* required significantly more time to complete than the other types, indicating a higher level of complexity or difficulty. In contrast, *Hearing* was completed in the shortest amount of time among



(a) *Solving Time* (b) *No. of Trials* (c) *Accuracy*

Fig. 6: Relative task performance for the four proposed CAPTCHAs (* $p < .05$; ** $p < .01$; and *** $p < .001$).

Table 2: Statistics for the task performance of the proposed CAPTCHAs—*Gathering* (GA), *Wayfinding* (WA), *Assembling* (AS), and *Hearing* (HE) (* $p < .05$; ** $p < .01$; and *** $p < .001$).

Measure	Mean (SD)				Normality				Sphericity		Omnibus			Post-hoc
	GA	WA	AS	HE	GA	WA	AS	HE	W	p	χ^2	ES	p	Groups (p-value)
Solving Time	43.3 (15.9)	57.1 (41.9)	162.0 (60.2)	17.9 (5.71)	.135	< .001	.021	.056	0.173	< .001	83.1	0.865	< .001***	HE < GA (< .001***); HE < WA (< .001***); HE < AS (< .001***); GA < AS (< .001***); WA < AS (< .001***)
No. of Trials	8.31 (3.12)	7.22 (3.06)	2.62 (0.75)	18.6 (5.68)	.005	.185	< .001	.090	0.492	< .001	58.8	0.861	< .001***	HE > GA (< .001***); HE > WA (< .001***); HE > AS (< .001***); GA > AS (< .001***); WA > AS (.0018**)
Accuracy	38.7 (18.6)	83.2 (14.3)	71.1 (15.4)	84.3 (15.4)	.914	< .001	.013	< .001	0.750	.128	59.3	0.618	< .001***	WA > GA (< .001***); AS > GA (.011*); HE > GA (< .001***); AS < WA (.013*); HE > AS (.003**)

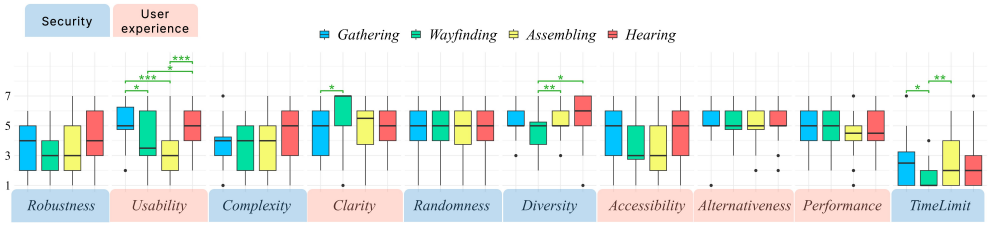


Fig. 7: Statistical results on the effectiveness of the four proposed CAPTCHAs based on the ten design principles for CAPTCHA (* $p < .05$; ** $p < .01$; and *** $p < .001$).

the four. *Gathering* also exhibited relatively short solving times, coming in at the second to *Hearing*. However, the task performance time alone is not sufficient to determine the "inherent" difficulty of the task, as the difficulty can be easily adjusted (e.g., number of objects to find, subtlety in the shape difference, length of the track, number of pieces to assemble).

- **No. of Trials:** An inverse relationship was found between Solving Time and No. of Trials. Participants attempted significantly fewer *Assembling* compared to the other types, possibly due to the longer time required per attempt discouraging multiple tries.
- **Accuracy:** This varied among the CAPTCHA types. *Gathering* showed notably lower accuracy compared to the other types, indicating that there was individual variation in the perceptual similarity correctness. Previous literature suggests that human accuracy for CAPTCHA is typically over 90% [33]. In this study, only *Wayfinding* and *Hearing* approached this level, suggesting that participants found these types more manageable. The accuracy data serves as a more reliable indicator of the inherent difficulty of the task than completion time.

4.4.2 Effectiveness Based on Design Principles

Since none of the ten items met parametric assumptions, the Friedman test was applied. The statistical results are summarized in Figure 7 and Table 3. In Figure 7,

Table 3: Relative sense of security for the proposed CAPTCHAs—*Gathering* (GA), *Wayfinding* (WA), *Assembling* (AS), and *Hearing* (HE)—based on the ten design principles for CAPTCHA (* $p < .05$; ** $p < .01$; and *** $p < .001$).

Measure	Mean (SD)				Normality (p)				Sphericity		Omnibus			Post-hoc
	GA	WA	AS	HE	GA	WA	AS	HE	W	p	χ^2	ES	p	
Robustness	3.56 (1.58)	3.09 (1.44)	3.47 (1.70)	4.09 (1.91)	.012	.015	.038	.024	0.747	.124	5.42	.056	.144	–
Usability	5.16 (1.57)	4.09 (1.78)	3.47 (1.57)	5.22 (1.43)	.003	.010	.070	.012	0.954	.925	26.3	.274	<.001	GA > WA (.019*); GA > AS (<.001***); HE > WA (.019*); HE > AS (< .001***)
Complexity	3.72 (1.53)	3.50 (1.70)	3.66 (1.58)	4.41 (1.74)	.046	.010	.100	.048	0.765	.160	6.53	.068	.089	–
Clarity	4.47 (1.54)	5.69 (1.84)	5.00 (1.70)	5.00 (1.72)	.010	<.001	.006	.001	0.682	.045	10.8	.112	.013	WA > GA (.010*)
Randomness	4.84 (1.78)	4.72 (1.90)	4.56 (1.79)	4.88 (1.68)	.011	<.001	.016	.012	0.953	.921	1.76	.018	.624	–
Diversity	5.28 (0.96)	4.69 (1.42)	5.50 (0.92)	5.47 (1.41)	.010	.019	<.001	.001	0.467	<.001	13.8	.144	.003	AS > WA (.007**); HE > WA (.013*)
Accessibility	4.59 (1.79)	3.81 (1.67)	3.69 (1.79)	4.31 (1.84)	.008	.016	.039	.049	0.814	.297	10.3	.108	.016	–
Alternativeness	5.22 (1.18)	5.19 (1.15)	5.03 (1.06)	5.16 (1.14)	<.001	.016	.008	.009	0.870	.528	2.14	.022	.544	–
Performance	4.88 (1.39)	4.75 (1.39)	4.34 (1.41)	4.69 (1.40)	.028	.071	.060	.033	0.911	.735	3.83	.040	.281	–
Time Limit	2.53 (1.54)	1.66 (0.83)	2.78 (1.68)	2.50 (1.72)	<.001	<.001	.001	<.001	0.640	.021	13.2	.138	.004	AS > WA (.007**); GA > WA (.039*)

the items marked in blue represent the metrics used to evaluate security, while those marked in red assess user experience. Higher values in the blue items signify enhanced security, indicating greater resistance to hacking or bot attacks. Conversely, higher values in the red items imply a more positive user experience.

A few significant differences were observed among the CAPTCHA types. *Hearing* demonstrated significantly better usability than *Assembling*. Similarly, *Gathering* showed significantly better usability compared to both *Wayfinding* and *Assembling*. In contrast, *Assembling* exhibited significantly lower usability than both *Gathering* and *Hearing*. Moreover, *Assembling* and *Hearing* required users to employ more diverse abilities, such as multi-sensory capabilities and cognitive skills, compared to *Wayfinding*. It implies that it is hard for Bots to break *Assembling* and *Hearing*.

Notably, *Hearing* displayed positive trends in both security metrics and user experience metrics, suggesting it effectively balances security with user experience. On the other hand, *Wayfinding* tended to have generally lower security metrics, indicating it may be less resistant to unauthorized access.

4.4.3 Overall VR User Experience

Since none of the variables met parametric assumptions, the Friedman test was applied. The statistical results are summarized in Figure 8 and Table 4. Significant differences were observed in task workload, system usability, and flow/interrupt among the CAPTCHA types.

- **Task Workload:** Lower scores signify a lesser cognitive and physical burden on the user. *Gathering* was found to be significantly less complex than both *Wayfinding* and *Assembling*. *Wayfinding* was felt significantly more complex than *Gathering* and *Hearing*, with higher stress levels and greater difficulty in manipulation compared to *Hearing*. *Assembling* imposed significantly higher physical demands and was perceived as more difficult to manipulate than both *Gathering* and *Hearing*. In

(a) Task Workload

(b) System Usability

(c) Flow/Interrupt

(d) Cybersickness

Fig. 8: Statistical results on the user experience in VR for the four proposed CAPTCHAs (* $p < .05$; ** $p < .01$; and *** $p < .001$).

contrast, *Hearing* was the easiest among the four, and associated with lower stress levels than *Wayfinding* and *Assembling*. This perception is consistent with the task performance measures.

- **System Usability:** Higher scores reflect better usability and user experience. Both *Gathering* and *Hearing* demonstrated significantly better system usability than *Wayfinding* and *Assembling*. *Assembling* received the lowest scores, indicating significantly poorer system usability compared to *Gathering* and *Hearing*. While *Wayfinding* was significantly less usable than *Gathering* and *Hearing*, it tended to perform better than *Assembling*. VR interaction involving 3D manipulation still seems to be relatively difficult. *Hearing* requires the least amount of such effort, resulting in the highest usability.
- **Flow/Interrupt:** Lower scores are desirable as they indicate less disruption to the main task flow. *Assembling* had the highest interruption scores, making it significantly more disruptive to the main task flow compared to *Gathering* and *Hearing*. This suggests that *Assembling* negatively impacted the user experience by interfering with the primary task. Despite all CAPTCHAs employing the usual VR interfaces, the flow seems to be affected by the stress level or difficulty of the task.
- **Cybersickness:** Lower scores are preferred as they indicate a lower level of sickness. Although the CAPTCHA task lasted a relatively short time, cybersickness has been reported. Repeated interactions with CAPTCHA tasks during the 5-minute experiment may have exacerbated this aspect. No significant differences were observed among the CAPTCHA types. However, *Hearing* tended to have lower scores, indicating a potential trend toward causing less cybersickness compared to the others, due to its minimal VR interaction involving 3D manipulation.

Table 4: Relative VR user experience for the proposed CAPTCHAs—*Gathering* (GA), *Wayfinding* (WA), *Assembling* (AS), and *Hearing* (HE) (* $p < .05$; ** $p < .01$; and *** $p < .001$).

	Mean (SD)				Normality				Sphericity		Omnibus			Post-hoc
	GA	**WA**	**AS**	**HE**	**GA**	**WA**	**AS**	**HE**	*W*	*p*	*statistic*	*ES*	*p*	Groups (p-value)
Mental Demand	30.0 (29.2)	36.1 (29.0)	42.0 (30.4)	21.6 (23.7)	.001	.029	.006	<.001	0.760	.149	13.1	.137	.004**	AS > HE (.008**)
Physical Demand	26.4 (29.0)	29.7 (27.5)	47.2 (30.5)	18.0 (20.6)	<.001	.004	.018	<.001	0.878	.571	37.0	.386	<.001***	AS > GA (< .001***); AS > WA (.030*); WA > HE (.040*); AS > HE (< .001***)
Temporal Demand	25.6 (24.8)	39.1 (28.9)	37.8 (30.0)	22.2 (25.6)	.002	.063	.051	<.001	0.751	.130	9.70	.101	.021*	–
Frustration	22.3 (27.4)	37.8 (29.9)	42.7 (31.8)	17.5 (22.5)	<.001	.014	<.001	<.001	0.906	.712	20.2	.210	<.001***	AS > GA (.008**); WA > HE (.023*); AS > HE (.002**)
Complex	26.7 (25.5)	46.9 (30.7)	64.7 (26.1)	17.8 (20.8)	<.001	.098	.011	<.001	0.613	.013	60.1	.626	<.001***	AS > GA (<.001***); WA > GA (.010*); AS > HE (<.001***); WA > HE (<.001***)
Stress	31.4 (29.2)	45.6 (27.7)	49.1 (32.3)	19.5 (22.0)	.001	.172	.010	<.001	0.865	.508	36.4	.379	<.001***	WA > GA (.028*); AS > GA (.013*); WA > HE (<.001***); AS > HE (<.001***)
Distraction	18.3 (24.4)	36.9 (28.6)	39.7 (33.1)	22.8 (21.9)	<.001	.045	.006	.001	0.813	.290	17.5	.183	<.001***	WA > GA (.007**); AS > GA (.003**)
Strain	27.3 (29.4)	40.8 (31.0)	40.2 (32.9)	22.8 (24.5)	<.001	.051	.010	<.001	0.883	.596	14.5	.151	.002**	AS > GA (.025*)
Difficulty	23.8 (23.0)	47.2 (31.3)	64.4 (25.1)	21.2 (21.6)	<.001	.028	.020	<.001	0.821	.320	44.8	.467	<.001***	WA > GA (.025*); AS > GA (<.001***); HE < WA (.007**); HE < AS (<.001***)
Break in Presence	29.8 (21.4)	43.3 (23.1)	27.0 (19.0)	33.8 (22.7)	.010	.081	.158	.037	0.886	.608	12.5	.130	.006**	WA > GA (.040*); WA > AS (.010*)
System Usability	67.1 (16.9)	52.0 (17.8)	46.2 (14.5)	69.1 (11.8)	.204	.594	.275	.478	0.865	.504	47.7	.476	<.001***	GA > WA (<.001***); GA > AS (<.001***); HE > WA (<.001***); HE > AS (<.001***)
Flow/Interrupt	3.60 (1.58)	4.12 (1.54)	4.59 (1.41)	3.45 (1.25)	.157	.238	.171	.444	0.867	.515	28.1	.294	<.001***	WA > HE (.045*); AS > GA (<.001***); AS > HE (< .001***)
Discomfort	3.03 (1.99)	3.81 (1.67)	4.16 (1.89)	3.09 (1.77)	<.001	.050	.001	.002	0.850	.437	15.4	.160	.001**	AS > GA (.007**)
Nausea	3.06 (1.98)	2.47 (1.92)	2.97 (2.16)	2.19 (1.71)	<.001	<.001	<.001	<.001	0.686	.048	5.66	.059	.129	–
Oculomotor	3.41 (2.01)	2.97 (1.80)	3.12 (2.00)	2.44 (1.70)	.003	.004	<.001	<.001	0.821	.319	5.80	.061	.122	–
Disorientation	2.19 (1.42)	2.94 (1.90)	2.47 (1.67)	1.88 (1.01)	<.001	<.001	<.001	<.001	0.673	.038	8.75	.091	.033	–

5 Expert Review

To further evaluate the security aspects of the proposed CAPTCHAs, four university professors were invited for a comprehensive review. Two with expertise in AI (AP1 and AP2), and two specializing in security (SP1 and SP2). AP1 and AP2 received their PhDs in 2014 and 2017, respectively, and both specialize in machine learning, deep learning, and computer vision. Each rated their VR familiarity as 2 on a 7-point Likert scale (1 = not at all, 7 = very much). SP1, awarded a PhD in 2000, focuses on software, network, and system security, while SP2, who received a PhD in 2022, specializes in software security and privacy. Their VR familiarity scores were slightly higher, at 4 and 3, respectively.

5.1 Interview Procedure

Unlike a general user study, we visited experts' offices for them (with the needed facility—i.e., the Meta Quest 3 headset/controller and a high-performance laptop with NVIDIA RTX4090) to use and experience the four proposed CAPTCHAs first hand. Before the interview began, we obtained consent from the experts for their participation in the interview and session recording. The experts were first briefed on the purpose and background of this study. We then explained the design principles of the CAPTCHA, and gathered feedback on whether they were well-formulated and

which ones should be prioritized. We also mentioned that the CAPTCHA types we implemented adhered to these principles as closely as possible.

Next, we explained each CAPTCHA type in a counter-balanced order. The explanation covered how the CAPTCHA design principles were reflected and how to use the handheld controllers to interact with and solve the CAPTCHA. After the explanation, the experts, with the assistance of the experimenter, wore the VR headset and used handheld controllers to experience the given CAPTCHA.

After each experience, we conducted discussions based on a semi-structured interview format, focusing on security robustness, vulnerability of these by AI technologies, as well as needed improvements. Once all the types were discussed, the interview concluded. The interview lasted approximately one hour. Experts were compensated approximately $75. The recorded interviews were transcribed by the experimenter, and the findings for each type are summarized in the following sections.

5.2 Summary of the Expert Reviews

5.2.1 3D Shape Match and Gathering

Experts AP1 and AP2 noted that while this CAPTCHA type aimed to leverage human cognitive abilities, it may be susceptible to AI attacks through 3D reconstruction and similarity comparison. AP2 asserted that AI systems might reconstruct 3D models to assess similarities, potentially compromising security. However, AP2 also mentioned that the criteria for perceptual similarity judgments could differ between humans and AI, providing a layer of defense. AP1 observed that human judgments of perceptual similarity are subjective and can be ambiguous, which might affect usability (i.e., humans too would be prone to making mistakes). To enhance security, AP2 suggested using textures that are difficult for AI to reconstruct in 3D, thereby hindering similarity assessments.

5.2.2 3D Wayfinding

All experts agreed that this CAPTCHA type could be vulnerable to hacking via reinforcement learning. They pointed out that straightforward pathfinding without obstacles would allow AI to easily figure out the right pathway to the destination. AP1 proposed that increasing or altering the intervals of the waypoints (discrete positions on the track that the car makes each movement by) to increase the computational load of the AI attack, enhancing security. AP2 recommended increasing the complexity by adding additional other interactive missions before reaching the destination. SP1 and SP2 similarly suggested introducing variables like obstacles or falls to eliminate predictability and reduce the potential for AI exploitation.

5.2.3 3D Assembling

Experts AP1 and AP2 expressed concerns that AI could mimic user actions or attempt to assemble objects by recognizing and manipulating individual parts. To counter this, AP1 suggested applying complex textures, such as images of animals or objects, to the blocks, making them harder for AI to recognize. Additionally, AP1 recommended

adding grid patterns or intricate lines to the background or incorporating visual effects like snow or rain to disrupt the AI's visual processing. While humans can effectively filter out such distractions, these measures could significantly increase the workload of an AI attack.

5.2.4 3D Hearing

SP2 acknowledged that tasks simultaneously requiring multi-sensory abilities and multi-cognitive abilities, such as auditory, visual, spatial perception, and language reasoning, present a high level of difficulty for AI, thereby enhancing security. All experts agreed that this CAPTCHA type offered high usability for human users, yet a relatively difficult enough task. AP2 mentioned that challenges similar to the "cocktail party effect," where individuals selectively focus on specific sounds amid background noise, would be another intuitive and similar task for humans, but difficult for AI to replicate. AP1 also suggested altering object textures to make them harder for AI to recognize, further improving security. For example, if a cat object has an elephant texture on it, AI may prioritize the texture and think it is an elephant.

5.2.5 Other Comments

SP1 cautioned that even with random problem generation, repeated exposure could make the system vulnerable to rainbow table attacks — a method where attackers use precomputed tables of hash values and their corresponding plaintexts to efficiently bypass security measures; thus, a sufficient number of unique problems is essential to prevent repetition. AP2 noted that tasks with multiple steps are currently challenging for AI to mimic, which could be leveraged for security.

All experts emphasized the importance of usability, advocating for simple actions and minimal movements to operate the CAPTCHA effectively. Additionally, they suggested considering CAPTCHA methods based on behavioral characteristics such as head or arm movements, or manipulation patterns that distinguish humans from bots. SP1 recommended implementing CAPTCHA systems that intervene only when unusual patterns are detected during normal usage.

AP1 and SP1 observed that compared to traditional CAPTCHAs, the proposed types involve excessive movement, complex manipulation, and longer solving times, which may be impractical during time-sensitive tasks like logging in, voting, or making purchases. SP1 stressed that sacrificing usability for security is counterproductive, as users might abandon the system entirely; maintaining usability while ensuring security is crucial. In this sense, all the experts echoed the results of the user experiment in which 3D Hearing, with minimal 3D manual interaction, was most preferred and rated to be the most effective among the four.

6 Robustness Analysis

6.1 3D Wayfinding

Based on expert reviews and the user study, 3D Wayfinding was identified as the most vulnerable in terms of robustness. To further investigate this aspect, we conducted

robustness testing to determine whether AI could successfully break this CAPTCHA using reinforcement learning.

In this CAPTCHA, the success criterion requires the user to navigate a car to the destination by passing through designated discrete waypoints. The waypoint information—including IDs and absolute coordinates—is invisible to the user and cannot be directly accessed ahead of time. To compromise the CAPTCHA, an attacker would need to reconstruct the 3D environment by analyzing video recordings of the CAPTCHA being performed.

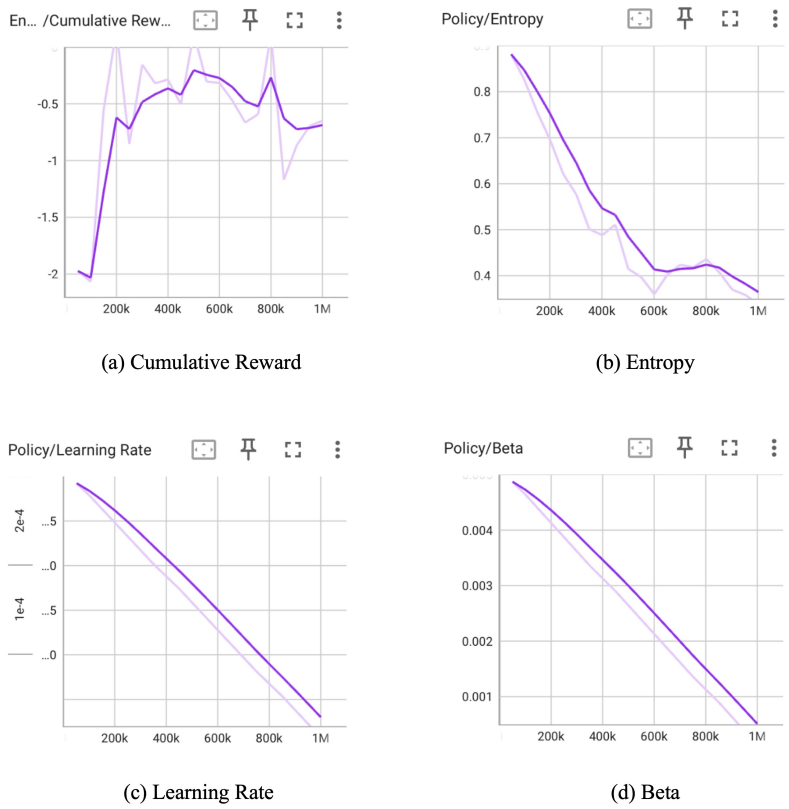
Reconstructing the 3D environment from 2D video presents significant challenges. Without depth information, and with roads overlapping in multiple layers causing blind spots, achieving a complete and accurate reconstruction is difficult. Even if the roads are fully reconstructed and the positions of the waypoints are identified, these coordinates would be relative to an arbitrary reference coordinate system. Successfully solving the CAPTCHA requires knowledge of the exact waypoint positions in absolute (i.e. with respect to the hidden true reference frame) coordinates and navigating through those specific locations.

Determining the absolute coordinates is challenging because the origin point and scale information of the VR space are unknown. While one could theoretically assign a random origin in 3D space to compute absolute coordinates, this approach would demand enormous training time and computational resources, making it impractical for quick attacks.

Assuming that 3D reconstruction is achievable, we experimented with reinforcement learning to assess whether AI could break the CAPTCHA. Without access to the waypoint information, the AI failed to learn effectively. It moved to arbitrary positions from its current location, but did not know where the exact waypoints were.

Conversely, when the waypoint information was provided, successful learning became possible. We employed the Proximal Policy Optimization (PPO) algorithm and conducted training over 1,000,000 steps. Rewards were structured as follows: reaching the destination yielded a reward of +1, moving outside the designated path incurred a penalty of -1, and a step penalty of -0.01 was applied to encourage minimal path length. Figure 9 shows the training statistics of reinforcement learning. The cumulative reward increases over time, indicating that the agent successfully learned the navigation policy. Meanwhile, the entropy, learning rate, and beta values gradually decrease, suggesting that the agent's actions became more deterministic and the training process remained stable throughout. As a result, AI was able to break the CAPTCHA with a success rate of 97.86%. When we increased the difficulty by changing the waypoint IDs, the success rate decreased to 14%.

However, the provision of waypoint information implies that the source code would have been compromised. In such a scenario, methods to bypass the CAPTCHA without AI technology would likely be available. Thus, while AI could breach the CAPTCHA under these conditions, the requirement of internal waypoint data significantly limits the practicality of this attack vector.



(a) Cumulative Reward

(b) Entropy

(c) Learning Rate

(d) Beta

Fig. 9: Training statistics of reinforcement learning: Cumulative reward increases while entropy, learning rate, and beta gradually decrease, indicating stable policy convergence.

6.2 3D Shape Match and Gathering

Two requirements are put forth: (1) human performance of shape match (and gathering) must be above the typical 90%, while (2) the accuracy of AI attack bot should be significantly lower. A convolutional neural network (CNN) as an AI attacker was put to the test. The success criterion for this CAPTCHA requires the user to find the two objects most similar to the target object.

Since 3D models used in the CAPTCHA matching task were not directly available, a Multi-View CNN (MVCNN) was employed to recognize 3D shapes from 2D images captured from multiple angles [47]. As illustrated in Figure 10, cameras were arranged in a cylindrical configuration and tilted toward the object center to photograph the 3D objects [48]. Datasets were generated by capturing images from 6, 12, 22, 42, and 82 different angles using orthographic and perspective projection [47, 49].

The original MVCNN model (as published in [47]) demonstrated poor classification performance with our datasets. To enhance accuracy, we replaced CNN_1 component

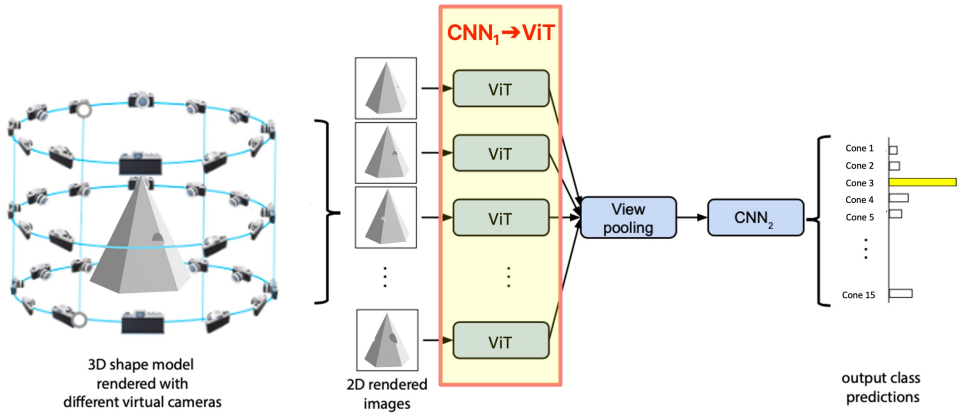


Fig. 10: Modified Multi-View Convolutional Neural Network (MVCNN) architecture.

in Figure 10 with the Vision Transformer (ViT) [48, 50] and trained the model for 30 epochs using 5-fold cross-validation with our datasets.

Figure 11 indicates that the classification accuracy for identifying the one most similar object to the target did not reach the participants' accuracy of 73.91% when using 42 views or fewer. Only when utilizing 82 orthographic views did the model achieve a test accuracy of 80.89% (F1-score 0.8075), comparable to human-level performance.

To identify the second most similar object to the target, we calculated cosine similarity using features extracted from the vision transformer. Figure 12 presents the similarity matrix computed from the Pyramid dataset, which consists of 82 orthographic views. Higher values indicate stronger similarity between class pairs. The matrix is symmetric, and self-similarity values (diagonal) are all 1.

Analysis of the matrix revealed that AI did not produce similarity scores consistent with human judgments. Similar trends were observed across other datasets as well. This suggests that even when the AI correctly identified the most similar object, its second choice often differed from human selection, confirming the robustness of this CAPTCHA against AI attacks.

However, participants' accuracy in finding the second most similar object in this task was only 38.67%, far below the typical CAPTCHA criterion of a 90% human success rate. Moreover, alternative methods not explored in this study may enable AI to achieve perceptual similarity judgments more closely aligned with those of humans.

7 Discussion

7.1 Design Guidelines for VR CAPTCHA

Based on this study's findings and analyses, several recommendations have emerged for designing cognitive 3D immersive CAPTCHA systems for both security and user experience in VR environments:

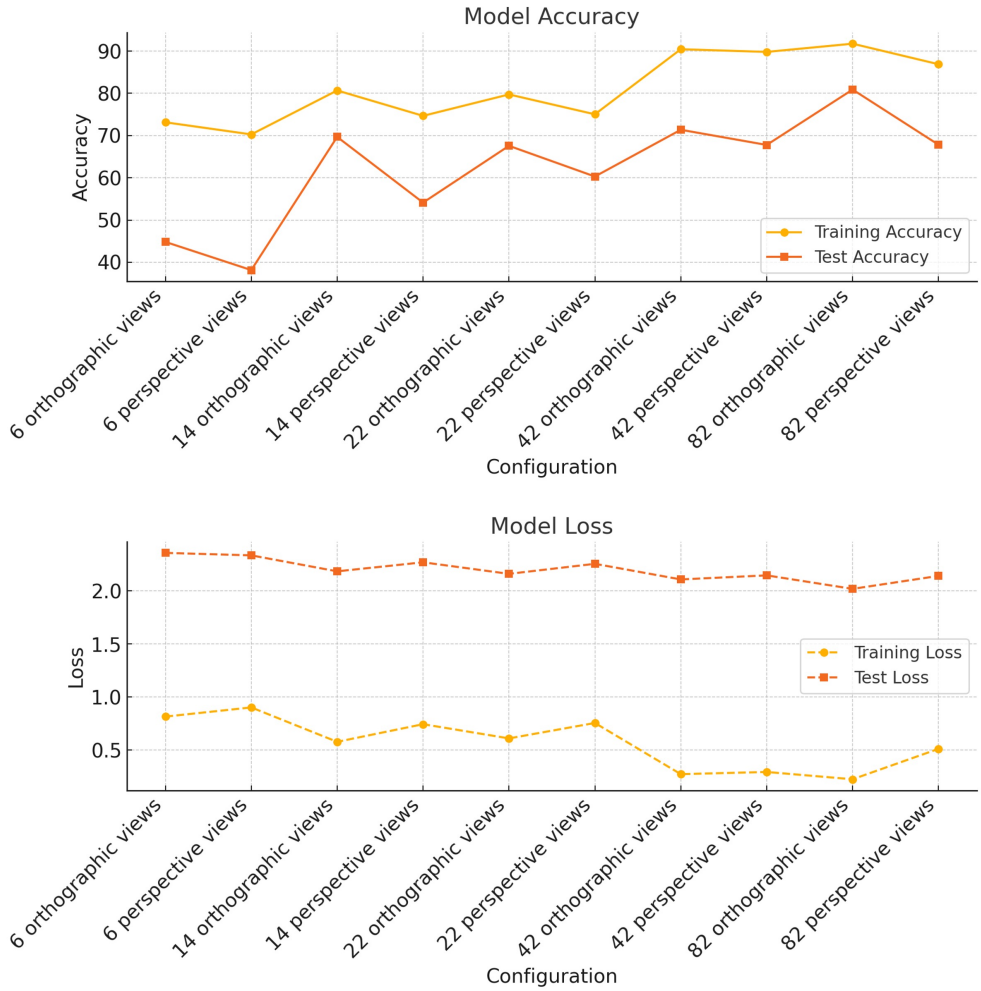


Fig. 11: Training and test accuracy (Top) and loss (Bottom) of the modified MVCNN model for identifying the one most similar object. The test accuracy remained below the participants' accuracy of 73.91% when using up to 42 views. Only with 82 orthographic views exceeded human-level performance, achieving 80.89% (F1-score 0.8075) test accuracy.

- **Balance between Interaction Complexity and Usability**: Involving 3D reasoning and interaction, thus making the CAPTCHA task more complex, is one natural way of raising its security level, i.e., making it more difficult for AI's breach attempts. However, while one may regard 3D interaction to be something natural and easy for humans to carry out, at least with the current 3D VR interfaces, this may not be entirely true. In fact, there may be significant differences in 3D

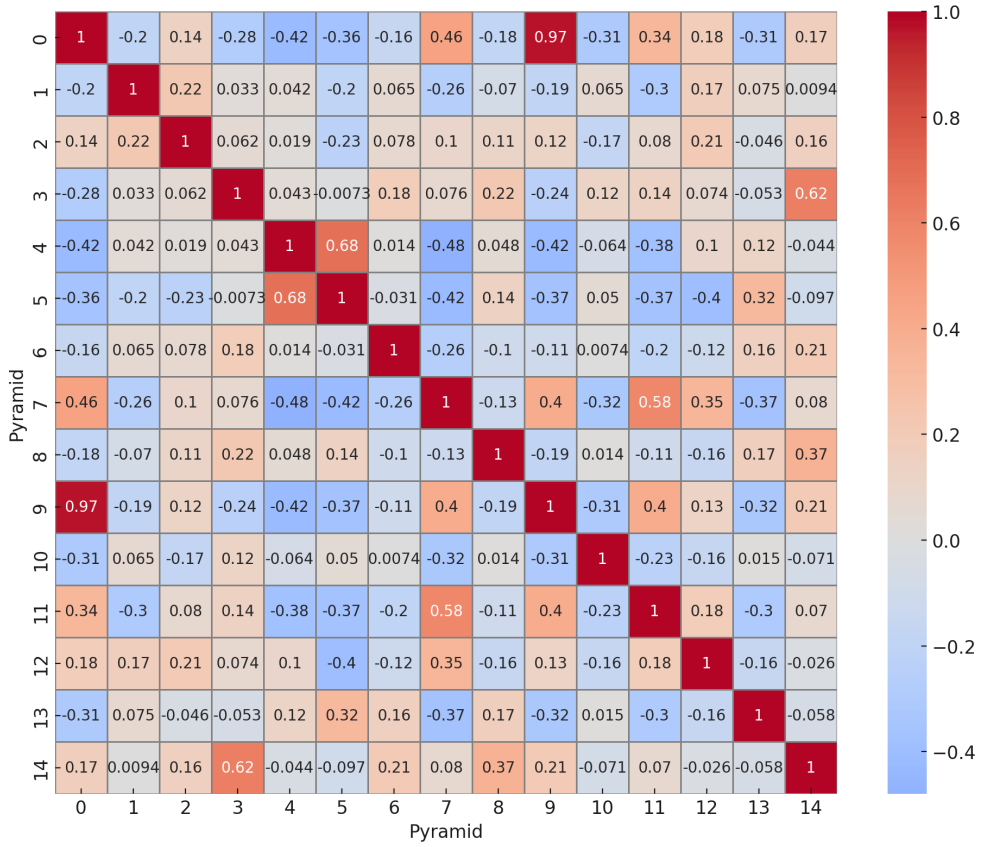


Fig. 12: Cosine similarity matrix of the Pyramid dataset, consisting of 82 orthographic views, showing the second most similar object based on features extracted by the vision transformer. The results exhibit inconsistencies with human judgments, highlighting the robustness of the CAPTCHA against AI attacks.

interaction between real life and synthetic VR spaces due to a lack of physicality, inaccuracies of the sensors, less than realistic looking graphics, lack of full multimodality, etc. Thus, one recommendation is to incorporate 3D spatial reasoning but limit the actual physical 3D interaction to a limited extent (especially manipulation). In our study, e.g., the 3D Wayfinding involves complex 3D spatial reasoning but also requires the user to enact it physically, significantly affecting the usability. The lowered usability may negatively impact the performance as well, making it ultimately unsuitable for an effective CAPTCHA task. On the other hand, the 3D Hearing involves non-trivial 3D cognitive or perceptual capability, with only minimal physical interaction (with no object manipulation). This contributed to producing a much higher, relatively positive user experience and task performance.

One can posit that the 3D Shape Match and Gathering could be much improved by limiting the task to 3D Match only as well.

- **Ensuring Clarity and Reducing Ambiguity**: Minimizing ambiguity in questions and answers is crucial for achieving high user accuracy without confusion. 3D Shape Match and Gathering highlights this aspect, as its perceptual similarity-based tasks, despite positive usability metrics, showed a low human accuracy rate of 38.67%. This discrepancy is caused by subjective and ambiguous judgments. Such variability compromises the CAPTCHA's reliability in distinguishing humans from bots.

  Ensuring clear-cut answers for 3D tasks is generally likely to introduce more ambiguities (if any) due to the increased spatial dimension and associated complexity. Thus, it would be adamant to identify the right task that is difficult, particularly from the perspective of the current AI state of the art. For instance, expert reviews found a common opinion to increase the difficulty of AI only by applying complex textures and dynamic backgrounds. This can disrupt AI's ability to interpret 3D environments, while humans can naturally filter out such distractions.
- **Improving Accessibility and Minimizing Cybersickness**: Employing immersive 3D interaction may reduce overall accessibility for users with diverse abilities, especially physical. Such an issue has been a serious concern [51–53] and was also often observed in our study. There were cases where the buttons on the controller were not operated smoothly due to physical limitations, and an alternative method could have facilitated the CAPTCHA performance. 3D manipulation as required in 3D Shape Match and Gathering and 3D Assembling relies much on one's dexterity, which can degrade with age [31, 54]. Additionally, in cases where navigation is needed, albeit to a limited extent duration and distance-wise, a substantial portion of the population had high sensitivity to cybersickness.

7.2 Limitations

One limitation of this study is the varying difficulty levels among the CAPTCHA types, which may have influenced the experimental results. For example, 3D Assembling involves assembling multiple pieces, inherently demanding more effort compared to simpler tasks like 3D Hearing. Standardizing the difficulty levels across all CAPTCHA types would have facilitated a more equitable comparison. However, achieving uniform difficulty is inherently challenging, as reducing the complexity of tasks like 3D Assembling (e.g., by using only three pieces) would improve usability and ease of operation but compromise robustness against automated attacks. Another limitation lies in the inability to fully analyze the robustness of certain CAPTCHA types, such as 3D Assembling and 3D Hearing, due to constraints in expertise and resources. Despite such an aspect, the suggested guideline has little to do with the task-specific difficulty, and addresses the general characteristics of VR interaction and comparative capability between 2D and 3D oriented tasks.

Future work should address these gaps by conducting controlled studies where task difficulty is treated as an independent variable and systematically evaluating the trade-offs between usability and security. Additionally, expanding robustness testing

methodologies to include tasks like 3D Assembling and 3D Hearing would provide a more complete understanding of their security implications in VR environments.

8 Conclusion

This study explored the feasibility of cognitive 3D immersive CAPTCHA to enhance security in VR environments. The focus was on designing CAPTCHA capable of effectively distinguishing between humans and bots while balancing security with user experience. Four distinct CAPTCHA types were proposed, developed, and evaluated through user studies and expert reviews.

The findings demonstrate that leveraging VR-specific interactions and immersive features, such as mid-air gestures, spatial reasoning, and multi-sensory capabilities, can significantly enhance security by presenting tasks inherently difficult for AI to solve. However, certain CAPTCHA types revealed usability limitations, highlighting the need for further optimization to improve user experience without undermining security.

Challenges related to cybersickness, a common limitation in VR environments, were observed even during tasks with minimal solving time. It was noted that repeated interactions with CAPTCHA tasks could exacerbate this issue. It is therefore recommended that future CAPTCHA systems in VR be designed with mechanisms to mitigate such side effects.

Although this research focused on the VR domain, the concept of cognitive-based 3D CAPTCHA offers opportunities for broader application in 2D environments. By incorporating 3D graphics, texture deformations, and cognitive tasks, the design principles developed in this study could be adapted to create robust CAPTCHA systems for traditional 2D platforms without VR headsets. Such generalizable designs can contribute to enhancing security across diverse digital platforms.

In conclusion, this study introduced novel approaches to cognitive-based 3D CAPTCHA design for immersive environments, laying a foundation for future exploration and application. It also emphasized the need to balance usability and security while prioritizing user comfort. Future work should focus on controlled studies that treat task difficulty as an independent variable to better understand the trade-offs between usability and security. The insights from this work are anticipated to guide the development of next-generation authentication systems that are secure, user-friendly, and widely applicable across both immersive and conventional interfaces.

Declarations

- **Funding** This work was supported by the IITP/ITRC Program (IITP-2022-RS-2022-00156354) and the National Research Foundation of Korea (RS-2025-00514411).

- **Informed consent** This study was approved by the Institutional Review Board of Korea University (IRB No. KUIRB-2025-0054-01), and informed consent was obtained from all individual participants included in the study.

Appendix A 3D Shape Match and Gathering

A.1 3D Object Sets for Perceptual Similarity Assessment

One of the proposed prototypes for the Cognitive 3D Immersive CAPTCHA is the 3D Shape Match and Gathering task, which leverages the concept of perceptual similarity. To design problem sets that align with human assessments of similarity, we constructed a set of candidate 3D objects with subtle variations. These variations include changes in the number of vertices, geometric distortions, scaling, surface imperfections (e.g., scratches), and shape alterations. Figure A1 illustrates an example set of hexagonal pyramid-shaped 3D objects used as candidates in the task. The final answer sets were refined through an online pilot study, as described in the section below.

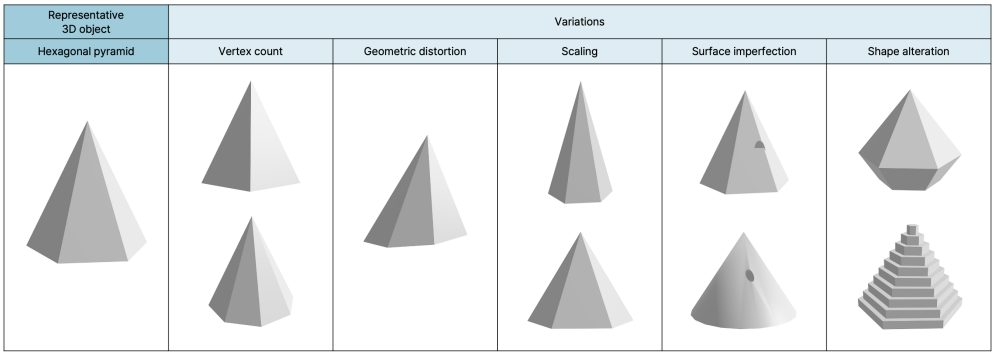


Fig. A1: Examples of shape and size variations applied to a hexagonal pyramid, including changes in the number of vertices, geometric distortions, scaling, surface imperfections (e.g., scratches), and shape alterations. These variants served as candidate objects in 3D Shape Match and Gathering CAPTCHA.

A.2 Online Survey

A pilot study was conducted to develop questions and answers to be used as CAPTCHA challenges for 3D Shape Match and Gathering. The test was carried out over three days using Amazon Mechanical Turk. Participants were limited to individuals with an approval rate of 95% or higher and more than 1,000 approvals. Those who completed the survey correctly received a compensation of $1.20.

The survey was conducted on a PC and utilized a dedicated 3D viewer embedded within the webpage. This viewer allowed participants to interactively examine 3D objects by rotating, zooming in, and zooming out using a mouse. The webpage displayed a representative 3D object on the left and 15 comparison objects on the right. For each of the 15 objects, participants rated its similarity to the representative object using a 7-point Likert scale (1: not at all, 7: very much). The interface for the online survey is illustrated in Figure A2.

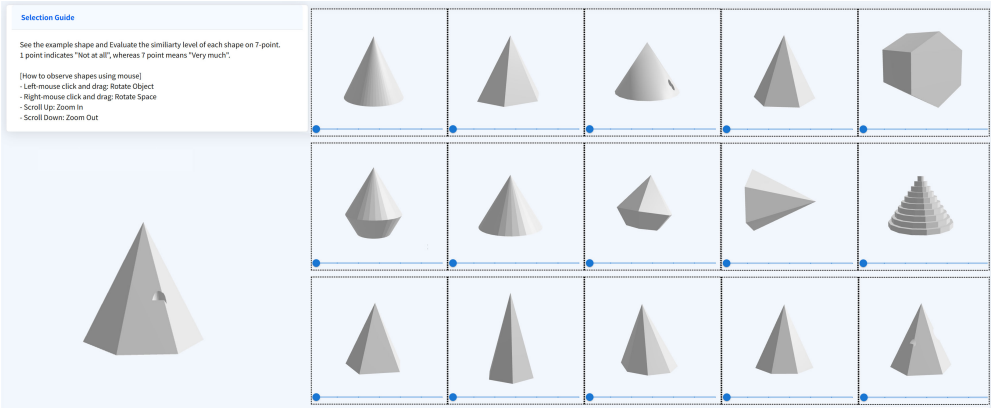


Fig. A2: Interface for the online survey was developed as a pilot study to construct answer sheets for 3D Shape Match and Gathering. On the left, a representative 3D object is presented, and 15 comparison objects are displayed on the right. The objects are shown in an interactive viewer, allowing participants to rotate, zoom, and explore them freely. Participants use the viewer to examine each object and rate its similarity to the reference object using a 7-point slider bar (1: not at all, 7: very much).

In total, 62 valid responses were collected, with an average survey completion time of 35 minutes. Participants' similarity ratings were averaged for each of the 15 comparison objects relative to the representative object. Based on these average scores, the objects were ranked in descending order of perceived similarity. The top two highest-ranked objects were selected as the most similar matches and were subsequently used to generate CAPTCHA challenge items.

By basing the selection of CAPTCHA questions and answers on aggregated human perception data, the process ensured that the challenges reflect human-like reasoning in 3D similarity judgment. This design contributes to the reliability of the CAPTCHA in distinguishing between human users and automated systems.

Appendix B Questionnaires

B.1 Design Principles of CAPTCHA

For each treatment, participants completed a questionnaire that included various items (see Section 4); among them, 10 specific questions were designed to evaluate how well the given CAPTCHA challenges conformed to the design principles of CAPTCHA in VR. These items were rated on a 7-point Likert scale (1: not at all, 7: very much).

- **Robustness**: Humans can solve this challenge, but robots or computers cannot.
- **Usability**: This challenge is intuitive, easy to manipulate, and simple to solve.
- **Complexity**: Humans may find this challenge simple, but robots or computers will perceive it as complex.

- **Clarity**: This challenge and its solution are clear and unambiguous.
- **Randomness**: This challenge and its solution can vary in multiple forms or randomly, making it difficult for robots to infer.
- **Diversity**: This challenge contains a variety of different elements.
- **Accessibility**: This challenge is accessible to everyone, regardless of age or physical disabilities.
- **Alternativeness**: If a human signals they cannot solve this challenge, alternative types of challenges can be presented.
- **Performance**: This challenge does not degrade the system's performance.
- **Time Limit**: If there are no time or attempt limits, this challenge can eventually be solved.

B.2 Flow/Interrupt

Scenario - Consider the following situation: *While engaging with a VR service/content, such as games or virtual tours, you attempt to make a purchase. The screen then transitions and displays a VR CAPTCHA challenge.*
Please review the scenario and respond to the question below, using a 7-point Likert scale (1 = Not at all, 7 = Very much).

- This task would interrupt the virtual reality experience I was previously engaged in.
- After solving this task, I would forget the activity I was doing before.
- After solving this task, I would lose track of how much time has passed.
- After solving this task, I would feel distracted.

References

[1] Doken, S., Lal, D., et al.: Using occluded 3d objects for gamified mixed reality captcha. APSIPA Transactions on Signal and Information Processing **13**(1) (2024)

[2] Tseng, W.-J., Bonnail, E., McGill, M., Khamis, M., Lecolinet, E., Huron, S., Gugenheimer, J.: The dark side of perceptual manipulations in virtual reality. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2022)

[3] Odeleye, B., Loukas, G., Heartfield, R., Spyridonis, F.: Detecting framerate-oriented cyber attacks on user experience in virtual reality (2021)

[4] Giaretta, A.: Security and privacy in virtual reality: a literature survey. Virtual Reality **29**(1), 10 (2024)

[5] Hedaia, O.A., Shawish, A., Houssein, E.H., Zayed, H.: Bio-captcha voice-based authentication technique for better security and usability in cloud computing. International Journal of Service Science, Management, Engineering, and

Technology (IJSSMET) **11**(2), 59–79 (2020)

[6] Von Ahn, L., Blum, M., Langford, J.: Telling humans and computers apart automatically. Communications of the ACM **47**(2), 56–60 (2004)

[7] Von Ahn, L., Blum, M., Hopper, N.J., Langford, J.: Captcha: Using hard ai problems for security. In: Advances in Cryptology—EUROCRYPT 2003: International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4–8, 2003 Proceedings 22, pp. 294–311 (2003). Springer

[8] Stephenson, S., Pal, B., Fan, S., Fernandes, E., Zhao, Y., Chatterjee, R.: Sok: Authentication in augmented and virtual reality. In: 2022 IEEE Symposium on Security and Privacy (SP), pp. 267–284 (2022). IEEE

[9] Rupp, D., GrieBer, P., Bonsch, A., Kuhlen, T.W.: Authentication in immersive virtual environments through gesture-based interaction with a virtual agent. In: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 54–60 (2024). IEEE

[10] Jeung, S., Hilton, C., Berg, T., Gehrke, L., Gramann, K.: Virtual reality for spatial navigation. In: Virtual Reality in Behavioral Neuroscience: New Insights and Methods, pp. 103–129. Springer, ??? (2022)

[11] George, C., Khamis, M., Buschek, D., Hussmann, H.: Investigating the third dimension for authentication in immersive virtual reality and in the real world. In: 2019 Ieee Conference on Virtual Reality and 3d User Interfaces (vr), pp. 277–285 (2019). IEEE

[12] Fişek, S.: Human-computer interaction, and virtual reality applications for memory enhancement. Human Computer Interaction **8**, 15 (2024) https://doi.org/10.62802/a5dj9288

[13] Pansare, P., Tripathi, M., Gupta, A.: An efficient 3d data annotation and object detection pipeline for production line. In: 2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS), pp. 1–6 (2024). IEEE

[14] Kürtünlüoğlu, P., Akdik, B., Duygu, R., Karaarslan, E.: Towards more secure virtual reality authentication for the metaverse: A decentralized method proposal. In: 2023 16th International Conference on Information Security and Cryptology (ISCTürkiye), pp. 1–6 (2023). IEEE

[15] Li, X., Chen, Y., Patibanda, R., Mueller, F.: vrcaptcha: exploring captcha designs in virtual reality. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–4 (2021)

[16] Mathis, F., Fawaz, H.I., Khamis, M.: Knowledge-driven biometric authentication in virtual reality. In: Extended Abstracts of the 2020 CHI Conference on Human

Factors in Computing Systems, pp. 1–10 (2020)

[17] Nair, V., Rack, C., Guo, W., Wang, R., Li, S., Huang, B., Cull, A., O'Brien, J.F., Latoschik, M., Rosenberg, L., et al.: Inferring private personal attributes of virtual reality users from head and hand motion data. arXiv preprint arXiv:2305.19198 (2023)

[18] Jiao, A., Duan, D., Xu, W.: Medusa3d: The watchful eye freezing illegitimate users in virtual reality interactions. Proceedings of the ACM on Human-Computer Interaction **8**(MHCI), 1–21 (2024)

[19] George, C., Buschek, D., Ngao, A., Khamis, M.: Gazeroomlock: Using gaze and head-pose to improve the usability and observation resistance of 3d passwords in virtual reality. In: Augmented Reality, Virtual Reality, and Computer Graphics: 7th International Conference, AVR 2020, Lecce, Italy, September 7–10, 2020, Proceedings, Part I 7, pp. 61–81 (2020). Springer

[20] Dinh, N.T., Hoang, V.T.: Recent advances of captcha security analysis: a short literature review. Procedia Computer Science **218**, 2550–2562 (2023)

[21] Hosaka, T., Furuya, S.: Stereoscopic text-based captcha on head-mounted displays. In: VISIGRAPP (5: VISAPP), pp. 767–774 (2020)

[22] Kumar, M., Jindal, M., Kumar, M.: A systematic survey on captcha recognition: types, creation and breaking techniques. Archives of Computational Methods in Engineering **29**(2), 1107–1136 (2022)

[23] Guerar, M., Verderame, L., Migliardi, M., Palmieri, F., Merlo, A.: Gotta captcha'em all: a survey of 20 years of the human-or-computer dilemma. ACM Computing Surveys (CSUR) **54**(9), 1–33 (2021)

[24] Tariq, N., Khan, F.A., Moqurrab, S.A., Srivastava, G.: Captcha types and breaking techniques: Design issues, challenges, and future research directions. arXiv preprint arXiv:2307.10239 (2023)

[25] Tanthavech, N., Nimkoompai, A.: Captcha: Impact of website security on user experience. In: Proceedings of the 2019 4th International Conference on Intelligent Information Technology, pp. 37–41 (2019)

[26] Pettis, B.T.: recaptcha challenges and the production of the ideal web user. Convergence **29**(4), 886–900 (2023)

[27] Algwil, A.M.: A survey on captcha: Origin, applications and classification. Journal of Basic Sciences **36**(1), 1–37 (2023)

[28] Mohamed, M., Gao, S., Saxena, N., Zhang, C.: Dynamic cognitive game CAPTCHA usability and detection of streaming-based farming. USEC (2014)

[29] Searles, A., Nakatsuka, Y., Ozturk, E., Paverd, A., Tsudik, G., Enkoji, A.: An empirical study & evaluation of modern {CAPTCHAs}. In: 32nd Usenix Security Symposium (usenix Security 23), pp. 3081–3097 (2023)

[30] Gonzalez, C.: Building human-like artificial agents: A general cognitive algorithm for emulating human decision-making in dynamic environments. Perspectives on Psychological Science **19**(5), 860–873 (2024)

[31] Chen, J., Or, C.: Assessing the use of immersive virtual reality, mouse and touchscreen in pointing and dragging-and-dropping tasks among young, middle-aged and older adults. Applied ergonomics **65**, 437–448 (2017)

[32] Sauer, G., Lazar, J., Hochheiser, H., Feng, J.: Towards a universally usable human interaction proof: evaluation of task completion strategies. ACM Transactions on Accessible Computing (TACCESS) **2**(4), 1–32 (2010)

[33] Chellapilla, K., Larson, K., Simard, P., Czerwinski, M.: Designing human friendly human interaction proofs (hips). In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 711–720 (2005)

[34] Hitaj, D., Hitaj, B., Jajodia, S., Mancini, L.V.: Capture the bot: Using adversarial examples to improve captcha robustness to bot attacks. IEEE Intelligent Systems **36**(5), 104–112 (2020)

[35] Roshanbin, N., Miller, J.: A survey and analysis of current captcha approaches. Journal of Web Engineering, 001–040 (2013)

[36] Hernandez-Castro, C.J., Ribagorda, A.: Pitfalls in captcha design and implementation: The math captcha, a case study. computers & security **29**(1), 141–157 (2010)

[37] Rao, M.K., Maniraj, M., Ganga, B.S.: Improved video captcha. Journal of Emerging Technologies in Web Intelligence **6**(4), 416–416 (2014)

[38] Fu, S., Tamir, N., Sundaram, S., Chai, L., Zhang, R., Dekel, T., Isola, P.: Dreamsim: Learning new dimensions of human visual similarity using synthetic data. arXiv preprint arXiv:2306.09344 (2023)

[39] Wah, C., Maji, S., Belongie, S.: Learning localized perceptual similarity metrics for interactive categorization. In: 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 502–509 (2015). IEEE

[40] Harris, D., Wilson, M., Vine, S.: Development and validation of a simulation workload measure: the simulation task load index (sim-tlx). Virtual Reality **24**(4), 557–566 (2020)

[41] Brooke, J.: Sus: A quick and dirty usability scale. Usability Evaluation in Industry

(1996)

[42] Feld, N., Bimberg, P., Weyers, B., Zielasko, D.: Simple and efficient? evaluation of transitions for task-driven cross-reality experiences. IEEE Transactions on Visualization and Computer Graphics (2024)

[43] Husung, M., Langbehn, E.: Of portals and orbs: An evaluation of scene transition techniques for virtual reality. In: Proceedings of Mensch Und Computer 2019, pp. 245–254 (2019)

[44] Pointecker, F., Friedl, J., Schwajda, D., Jetter, H.-C., Anthes, C.: Bridging the gap across realities: Visual transitions between virtual and augmented reality. In: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 827–836 (2022). IEEE

[45] Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology **3**(3), 203–220 (1993)

[46] Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G.: Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. Behavior research methods **41**(4), 1149–1160 (2009)

[47] Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 945–953 (2015)

[48] Alzahrani, M., Usman, M., Jarraya, S.K., Anwar, S., Helmy, T.: Deep models for multi-view 3d object recognition: a review. Artificial Intelligence Review **57**(12), 1–71 (2024)

[49] Nguyen-Phuoc, T.H., Li, C., Balaban, S., Yang, Y.: Rendernet: A deep convolutional network for differentiable rendering from 3d shapes. Advances in neural information processing systems **31** (2018)

[50] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[51] Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., Williams, I.: Inclusive augmented and virtual reality: A research agenda. International Journal of Human–Computer Interaction **40**(20), 6200–6219 (2024)

[52] Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., Williams, I.: Inclusive ar/vr: accessibility barriers for immersive technologies. Universal Access in the Information Society **23**(1), 59–73 (2024)

[53] Dudley, J., Yin, L., Garaj, V., Kristensson, P.O.: Inclusive immersion: a review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. Virtual Reality **27**(4), 2989–3020 (2023)

[54] Wu, Z., Wang, D., Zhang, S., Huang, Y., Wang, Z., Fan, M.: Toward making virtual reality (vr) more inclusive for older adults: Investigating aging effect on target selection and manipulation tasks in vr. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, pp. 1–17 (2024)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- reduecedCAPTCHADEMO.mp4