Spatial Cognition Through Gestural Interfaces: Embodied Play & Learning with Minecraft

Jannah Issa[1[0009-0004-8429-5156]], Vishesh Kumar[1[0000-0001-5097-7678]], and Marcelo Worsley[1[0000-0002-2982-0040]]

Northwestern University, Evanston IL 60208, USA
jannahissa2026@u.northwestern.edu, {vishesh.kumar, marcelo.worsley}@northwestern.edu

Abstract. Increasingly accessible gesture-based interfaces offer powerful but underexplored potential for facilitating inclusive and engaging learning experiences. In this work, we explore creating and testing a hand and body pose-based gestural control interface for playing Minecraft. This work aims to be a starting point for making Minecraft (and other video games) based learning and play more inclusive and engaging by moving away from just keyboard-mouse controls. We study the viability and user-friendliness of our initial designs through surveys, interviews, and early co-design sessions, shaping gestures that resonate with diverse youth, enriching the learning and assessment of spatial cognition skills among them.

Keywords: Multimodal Interfaces · Gesture Based Control · Spatial Thinking · Game-based Learning

1 Introduction

Educational video games spanning homes, schools, and various out-of-school contexts have become indispensable tools for learning [16]. Their applications in diverse fields such as STEM, Social Sciences, and Special Needs education are well-documented [2, 8]. Minecraft, in particular, stands out due to its extensive usage across disciplines, as well as its potential for enhancing spatial cognition skills [16].

Video games in general and Minecraft in particular have been recognized for supporting development of a variety of spatial cognition skills [13, 15], a fundamental aspect of human intelligence. [4] highlight the potential design space for engaging spatial cognition using embodied interfaces. Gestural interfaces also offer increased access for children with physical disabilities limiting motor control corresponding to keyboard and mouse control, addressing the need for more inclusive game-based learning experiences [3].

There is significant and increasing evidence suggesting that incorporating gesturing into educational practices enhances learning outcomes [9]. In this work we identify hand and body based gestural moves captured through web cameras to interact with Minecraft play. This innovative approach enables a unique combination of leveraging the greater accessibility of physical actions for young audiences, while also engaging sensory-motor interactions facilitating an embodied engagement with spatial reasoning.

2 Methods

In this study, we use observation, survey, and interview data to explore the viability of such a design in terms of ease of learning and use. We highlight the potential for how such a system can provide deeper insights into how different gestures connect to learning and deepening spatial reasoning skills [12]. Qualitative and thematic analysis of survey data from earlier work helped us identify and categorize various gesture types: interactive, involving direct manipulation of the virtual environment, and control, focusing on user input and navigation. These corresponded to vista and environmental forms of embodiment [4]. We chose gestures to control movement (walking & jumping), game view (panning the camera sideways), and in-game actions (breaking/mining) in Minecraft, and created a system that can convert gestures into Minecraft-based actions in real-time – using MediaPipe's hand pose recognition system.

2.1 Technical Details

Details of the system include a gestural control interface designed for playing Minecraft, offering an alternative to conventional keyboard and mouse controls. The Python program code integrates OpenCV, MediaPipe, and Minecraft libraries to accomplish this. Hand landmarks are identified and categorized through a K-Nearest Neighbors (KNN) classifier. The MediaPipe library is utilized for hand landmark detection, capturing the landmarks for each hand and translating their physical positions into in-game actions. These landmarks undergo processing to extract features contributing to the classification of various gestures. In real-time, the webcam feed is analyzed to detect hand landmarks using the MediaPipe library. The identified landmarks are then input into the trained KNN classifier for predictions. The predicted actions subsequently trigger corresponding in-game actions, enhancing the Minecraft experience through the use of gestural controls. The system is seamlessly integrated with the Minecraft API, allowing the execution of actions within the game world based on recognized gestures. Various gestures, such as walking, breaking/attacking, placing/using, looking, and jumping in Minecraft, trigger distinct actions. This comprehensive system incorporates player movement, block manipulation, and camera control for a more immersive gameplay experience.

2.2 User Study

We build on prior work not elucidated in this paper due to IRB limitations, wherein we surveyed 15 participants spanning 6-17 years of age from a summer camp. We collected initial expectations and preferences concerning gaming gestures through a pre-survey. We then presented a preliminary gesture based minecraft play system, explaining to them the specifics of gesture-action relationships to help provide initial insights on the engagement fostered through this system. Post playtesting, participants engaged in a post-survey where we asked them about their preferences for using gestures versus traditional keyboard/mouse controls, the perceived effectiveness of each gesture in gameplay scenarios, and further ideation on gestures they felt would be relevant.

After improvements, we analyze in this study a second round of testing with youth aged 11-14 years old, randomly selected from a Minecraft club organized in a public middle school in the Midwestern United States. Students in this club play after school hours at their school, in an opt-in program through personal interest in Minecraft. The participants in our study included 4 boys and 1 girl. in the context of a common spatial reasoning task recreated in Minecraft [7] – where participants navigate an environment, examine and manipulate complex 3D objects, and are then assessed around their memory of the objects' details and abilities .

2.3 Spatial Task

We wanted to evaluate students' spatial cognitive abilities in the spatial challenge and find out how they use gestures to recall information. In order to make this investigation easier, we built a virtual structure in Minecraft that is intended to resemble the commonly used mental rotations test [11]. First, the gestural interface was shown to the students and questions about their preferences in Minecraft were asked. They moved over the standard Minecraft terrain to become accustomed to the virtual world and learn which gestures corresponded to which activities.

Fig. 1).

Moving around the virtual world involved spatial awareness, understanding the gestures that triggered actions, and building an understanding of the correspondence between gestures and in-game controls. Following this, students were informed about participating in a memory game, where they would be tested after they finished playing. Placed just outside the meticulously created virtual structure, participants were tasked with identifying items within (a bed, bales of hay, a pumpkin, leaves, and chests), thereby fostering spatial memory. Their challenge was to mentally map the locations of various items strategically placed within the structure. Participants were given explicit instructions to traverse the virtual structure while utilizing the gestural interface to identify and locate objects. This hands-on approach aimed to reinforce the mental representation of spatial information through the physical enactment of movements, providing a comprehensive understanding of the intricate relationship between gestures and spatial memory.

Spatial mapping played a pivotal role in our study, aiming to unravel how students utilized gestures to perceive, comprehend, and mentally represent spatial relationships between objects and environments. Gestural interfaces, identified as crucial tools in this investigation, facilitated kinesthetic learning. This approach

Fig. 1. In-Game structure students interacted with, as well as the items to locate

directly linked physical movements to cognitive processes, thereby enriching spatial understanding by incorporating bodily experiences into the learning process [5]. Fig. 2).

After an initial gameplay phase, we transitioned participants to the testing segment. During this phase, we presented them with a visual representation of the structure in its regular orientation, prompting them to recognize its shape. Once the students acknowledged the image as the structure, we asked them a series of questions about item locations within the structure, with notes taken to track the nuances of their gestures. This detailed observation aimed to discern if students naturally incorporated gestures during the recall of spatial information. Adding an additional layer of complexity, an outline of the structure 2 was rotated 90 degrees either to the right or left. We then queried them again about their ability to recall the locations of items within the altered spatial configuration.

Throughout this process, we underscored the importance of adaptive and flexible interactions with spatial content. Learners demonstrated their ability to dynamically adjust their gestures based on the evolving spatial context, promoting adaptability in spatial thinking. An illustrative example included the use of gestures to rotate through virtual space, demonstrating adaptability to varying spatial scenarios and fostering a flexible spatial mindset.

After the initial series of questions, we further questioned participants about item locations, recall strategies post-rotation, and whether the gestural interface significantly contributed to their memory recall. We asked additional questions to explore general preferences and experiences with the gestural interface, providing a comprehensive understanding of the intricate interplay between spatial

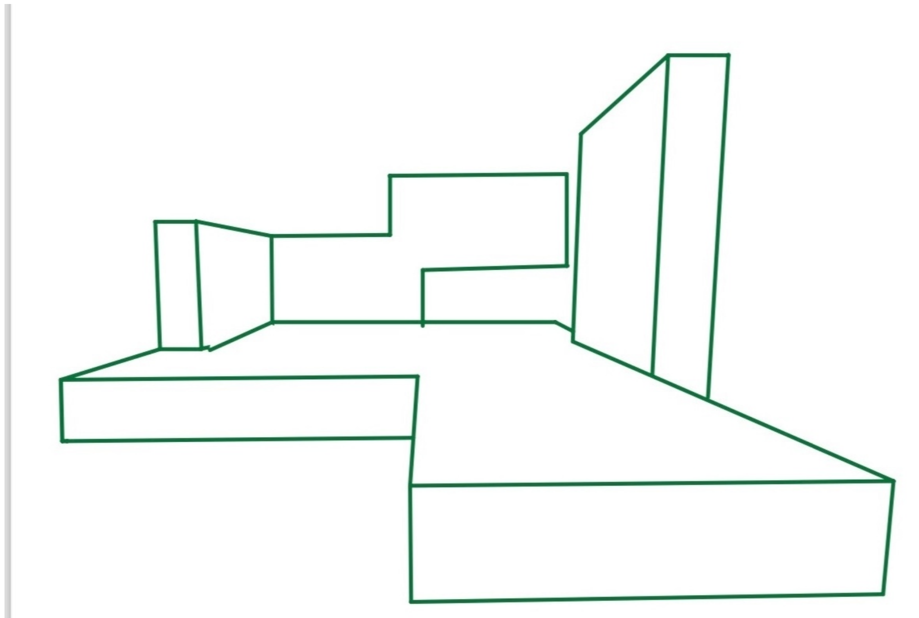

Fig. 2. Visual Representation of structure for Recall test

cognition, gestures, and memory recall within the virtual environment we meticulously designed.

3 Findings

In the post-play survey, we asked participants several questions to provide us with a detailed description of their interaction with the virtual environment. This exploration helped us to better understand how their gestures and descriptive language assisted in their spatial cognition and memory recall abilities within the virtual environment and the spatial task. The detailed exploration of spatial cognition and memory recall abilities within the virtual environment was not limited to individual experiences; rather, it extended across several user studies involving multiple students. As each participant engaged in the post-play survey, we present here a diverse array of strategies employing gestures and descriptive language to navigate and recall elements within the simulated space.

For instance, while prompting a student with questions regarding the location of specific objects, the student used very descriptive direction language, while also turning their body to correspond with the directions they were using to explain the location of the object. The student also utilized other objects as reference points, repeatedly referencing the overall layout of all objects within the structure to further explain instructions for finding each item. Notably, the student used descriptive directional language to articulate depth within the structure, using hands to simulate layers and visually express the spatial relationships. When the structure was rotated 90 degrees, the student was able to recall locations, guided to be descriptive in explaining how objects related to each other. After the recall portion of the survey, the student expressed a preference for playing with gestures over a keyboard, indicating enjoyment in using hands for game interaction. When asked about the impact of using hands on memory, the student confirmed that gestures enhanced visualization and improved memory recall. Fig. 3).

Fig. 3. Photo of student moving hand down

Similarly, another student was asked to identify the location of the bed without directly pointing at the picture. The student described the bed's location in relation to the length of platforms and shapes on the structure.As the orientation of the structure was changed, the student used hands more while describing the bed's location, noting the shapes and layout of the structure as reference points. Even after the structure was rotated, the student used gestures to provide directions for his description of the bed's new location relative to the new orientation.When asked about how they remembered the bed's location, the student explained breaking down the platform into shapes, faces, and corners, pointing down while describing the bed's position. The student mapped out several

Fig. 4. Photos of student moving hand front and back

pieces of hay and used gestures to convey depth, demonstrating a comprehensive reliance on hand movements for spatial recall.

Other students extensively used hands to outline the overall shape of the structure, employing flat hands to represent faces during recall. After the structure was rotated, the students mimicked the rotation with hands to assist in recall. This particular participant's reliance on gestures suggested a strong connection between hand movements and spatial memory.

4 Discussion

The findings from our user studies underscore the impact of integrating gestural interfaces into Minecraft gameplay for enhancing spatial cognition skills among youth. This section summarizes the key findings derived from the user studies, offering a comprehensive exploration of the impact of gestural interactions on spatial description, embodied memory and recall, visualization, orientation, shaping, and depth.

1. Spatial Description through Gestures

The adept utilization of body and hand gestures by students showcased a remarkable proficiency in spatial description. For instance, during spatial tasks, students employed sweeping hand motions to vividly illustrate the layers of virtual objects, providing a tangible representation of depth. A notable example emerged where a student effectively used descriptive directional language to explain the precise location of an object, simulating layers and enhancing spatial understanding. Students frequently employed specific hand gestures, such as flattening their hands, to represent faces or surfaces of the virtual structure. This nuanced use of gestures served as a symbolic representation, aiding in the

Fig. 5. Student using hands flat to represent faces of structure

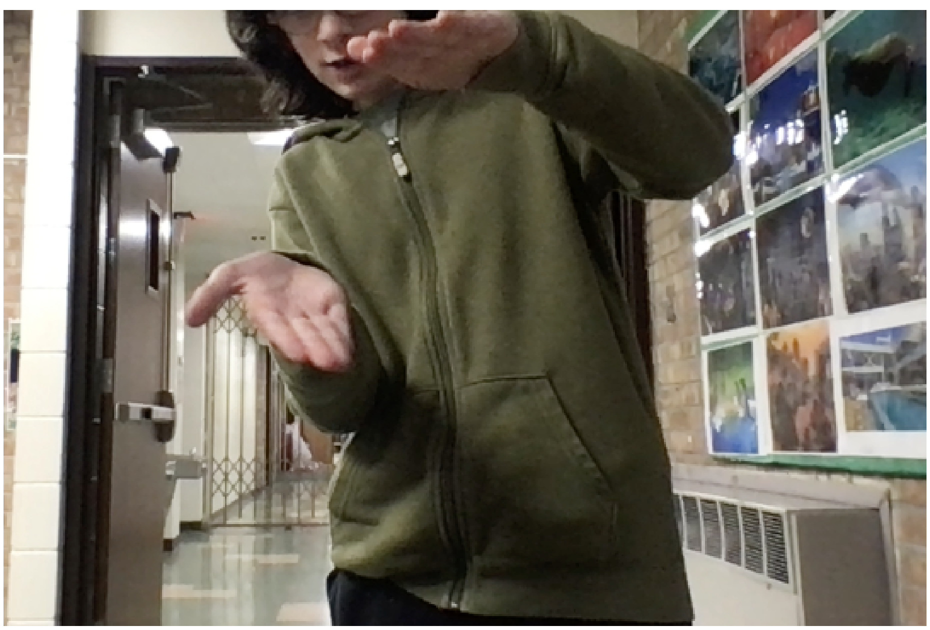

Fig. 6. Student turning hand faces to represent structures rotation during recall

description of the layout and orientation of objects within the Minecraft environment. For instance, students flattened their hands to symbolize the faces of the structure during recall, providing a visual and kinesthetic reference for the arrangement of virtual elements.

2. Embodied Memory and Recall Gestures became a powerful tool in facilitating memory recall, allowing students to embody their mental representations of the virtual environment. Despite the challenges posed by the rotation of the virtual structure, students seamlessly integrated gestures into their recall process. This adaptability underscored the effectiveness of gestural interactions in enhancing spatial memory. For instance, in the observation of one student, the inquiry about the location of an object demonstrated the powerful role of gestures in memory recall. When prompted to remember the position of an object, the student actively turned his body, using physical movement to enhance his mental recall process. The student not only used descriptive directional language like "behind the bed and under the pumpkins" but also employed gestures with his hands to simulate layers, providing a tangible representation of the virtual environment's depth. Additionally, when the structure was rotated, the student's ability to recall locations, guided by descriptive gestures, highlighted the adaptability and effectiveness of gestural interactions in spatial memory. The student's preference for playing the game with gestures further emphasized the positive impact on memory recall, as he confirmed that using his hands helped him visualize the virtual environment more effectively. The gestures served as an aid, enabling the student to anchor their memory in tangible spatial references.

3. Visualization and Orientation Beyond spatial description, students used gestures to deconstruct the virtual structure into shapes, corners, and extruding pieces, significantly contributing to memory recall. After the rotation of the virtual structure, students mimicked the rotation with their hands to assist in recall. This gesture served as a simulation technique, allowing students to mentally visualize and recall the reoriented structure. For example, one student illustrated the integration of gestures into the visualization and orientation of the virtual environment. The student, when asked to identify the bed without directly pointing, employed gestures to describe the bed's location in relation to the length of platforms and shapes on the structure. Even after the structure was turned, the student used hand movements to convey directions and elaborate on the bed's new position. The student's breakdown of the platform into shapes and corners, coupled with pointing and mapping with hand movements, demonstrated how gestures contributed significantly to the visualization and orientation of the virtual space.The use of hands to replicate the rotation underscored the interactive and embodied nature of gestural interactions in Minecraft gameplay.

4. Shaping and Depth Gestures also assisted in conveying depth, as students employed nuanced hand movements to represent spatial relationships within the virtual environment. In a notable example, a student effectively used gestures to map out the location of objects, providing both visual and kinesthetic cues. The observations of various students provided insights into how gestures played a pivotal role in conveying depth within the virtual environment. One student, in outlining the overall shape of the structure, used hand movements extensively. Additionally, the use of flat hands to represent the faces of the structure during recall and mimicking the rotation with hands showcased how gestures were employed to provide both visual and kinesthetic cues. The detailed gestures, such as waving hands to map out hay or signaling depth by moving hands forward and back, highlighted the dynamic and multi-dimensional aspects of gestural interactions in shaping spatial understanding. In several instances, students utilized flat hands to symbolize faces or surfaces of the virtual structure during recall. This mapping technique was observed as a means to break down the complex structure into simpler components, aiding in memory recall. Our documented students used flat hands to represent the orientation of platforms, shapes, and objects within the virtual space. This dynamic gesture provided a tangible and interactive way to convey spatial relationships.

The observations of students in Minecraft gameplay, particularly the use of gestures in spatial tasks, corroborate Atit et al.'s findings [1] where students utilized body and hand gestures to articulate spatial information, highlighting the intrinsic connection between gestures and spatial cognition. We built on their work to ask students to describe the layout and depth of a virtual environment, and students effectively used a variety of gestures in doing the same. The use of sweeping hand motions to illustrate layers of virtual objects in Minecraft parallels the students in the research study who used gestures to convey the intricate layers of rock. This shared emphasis on gestures as a means of providing a tangible representation of depth suggests a consistent pattern in the use of gestural interactions for spatial understanding. We also notice a similarity in the gestures used for directions, including pointing upwards for elevation and circular motions for object orientation. Both studies converge on the notion that gestures serve as a dynamic and expressive tool for spatial description and understanding.

We also notice students use gestures while engaged in verbal memory recall and spatial understanding, which also influences their communication within the virtual environment of Minecraft. This connection between gestures and talk mirrors the work by Morsell & Krauss [5] which explicate how gestures, especially in the role of engaging spatial memory, influence speech production.

5 Conclusion

Early findings highlight a mix of affordances and limitations. While movement and navigation tasks are sometimes more laborious using hand poses, corroborating ongoing findings around interaction fatigue [10], players report easier viewfinder and object manipulation using gestures. Some participants perceived movement tasks as more laborious when using hand poses. There were instances where the program was not adapted well to the hands of children, and the use of extra fingers posed challenges. These limitations underscore the need for ongoing refinement and adaptation to ensure optimal usability and inclusivity. We find a noticeable increase in usage of embodied gestures in follow-up spatial cognition tasks by players who utilized the gestural interfaces more during gameplay. Youth provided a variety of complex gestures and interactions that connect complex concepts which are often laborious tasks in gameplay to easy to demonstrate gestures, including symbols for loops, multi-step constructions, and multi-axis viewfinder which we are currently working on creating.

By incorporating gesture principles into instructional design, educators can craft more engaging and effective learning environments for students, particularly in spatial thinking and STEM fields, where success is closely linked to spatial abilities. The increased usage of embodied gestures observed in participants who extensively used gestural interfaces during gameplay suggests that integrating such interfaces in educational tasks can enhance spatial cognition. Moreover, when applied to teamwork tasks, gestural interfaces not only bolster spatial understanding but also foster social interaction, collaboration, and innovative learning approaches with richer social manners. The hands-on and interactive nature of gestural interfaces can potentially make learning more enjoyable, leading to heightened participation and sustained interest. Additionally, the gestural interface system serves as a novel tool for educators to assess students' spatial reasoning skills, offering insights into their problem-solving approaches.

By integrating gestural interfaces and spatial tasks, educators are equipped with a potent arsenal that not only nurtures spatial cognition but also fosters social interaction, collaboration, and new ways of learning not only with their whole bodies but also in richer social manners [14]. By considering inclusive and embodied ways of enriching spatial cognition, this research charts a course toward a more equitable, inclusive, and empowering educational landscape.

References

1. Atit, K., Gagnier, K., & Shipley, T.: Student Gestures Aid Penetrative Thinking. In: Journal of Geoscience Education, 63, pp. 66-72. https://doi.org/10.5408/14-008.1 (2015)
2. Cai, Y., Goei, S. L., & Trooster, W. (Eds.): Simulation and serious games for education. Springer (2016)
3. Cezarotto, M. A., & Chamberlin, B.: Towards accessibility in educational games: a framework for the design team. In: InfoDesign-Revista Brasileira de Design da Informação, 18(2) (2021)
4. Clifton, P. G., Chang, J. S. K., Yeboah, G., Doucette, A., Chandrasekharan, S., Nitsche, M., ... & Mazalek, A.: Design of embodied interfaces for engaging spatial cognition. Cognitive research: principles and implications, 1, pp. 1-15 (2016)
5. Morsella, E., & Krauss, R. M.: The Role of Gestures in Spatial Working Memory and Speech. In: The American Journal of Psychology, 117(3), pp. 411–24. https://doi.org/10.2307/4149008 (2004)
6. Freina, L., & Ott, M.: A literature review on immersive virtual reality in education: state of the art and perspectives. In: The international scientific conference elearning and software for education. Vol. 1, No. 133, pp. 10-1007 (April 2015)
7. Guay, R. B., & McDaniel, E. D.: The Purdue spatial visualization tests: A brief history. In: Educational and Psychological Measurement, 37(2), pp. 391-397 (1977)

8. Kangas, M.: Creative and playful learning: Learning through game co-creation and games in a playful learning environment. In: Thinking skills and Creativity, 5(1), pp. 1-15 (2010)
9. Koumoutsakis, Theodora, et al.: Gesture in Instruction: Evidence from Live & Video Lessons (2016)
10. Ma, F., Song, F., Liu, Y., & Niu, J.: Quantitative Analysis on the Interaction Fatigue of Natural Gestures. In: IEEE Access, 8, 190797-190811 (2020)
11. Shepard, R. N., & Metzler, J.: Mental rotation of three-dimensional objects. Science, 171(3972), pp. 701-703. (1971)
12. Skulmowski, A., & Rey, G.D.: Embodied learning: introducing a taxonomy based on bodily engagement and task integration. In: Cogn. Research, 3, 6. https://doi.org/10.1186/s41235-018-0092-9 (2018)
13. Subrahmanyam, K., & Greenfield, P. M.: Effect of video game practice on spatial skills in girls and boys. In: Journal of applied developmental psychology, 15(1), 13-32 (1994)
14. Wang, L. H., Chen, B., Hwang, G. J., Guan, J. Q., & Wang, Y. Q.: Effects of digital game-based STEM education on students' learning achievement: a meta-analysis. In: International Journal of STEM Education, 9(1), pp. 1-13. (2022)
15. Worsley, M., & Bar-El, D.: Spatial reasoning in Minecraft: An exploratory study of in-game spatial practices. In: Computer Supported collaborative learning, 2. (2020)
16. Zhonggen, Y.: A meta-analysis of use of serious games in education over a decade. In: International Journal of Computer Games Technology. (2019)