

Verification of Operability in Object-based Interfaces

Atsuki Masuta and Yusuke Kajiwara*

Division of Production System Science, Komatsu University,
Shichomachi Nu1-3, Komatsu, Ishikawa 923-8511, Japan

(Received January 5, 2026; accepted May 28, 2026)

Keywords: virtual reality (VR), 3D user interface (3DUI), NASA Task Load Index (NASA-TLX)

We propose an object-based interface (OBI), in which functions are embedded within the objects themselves and operated by touching or moving them by hand, to address the issue that the separation of “functions” and “operating units” common in conventional IT devices makes intuitive operation difficult. Implementation in mixed reality is intended; however, owing to technical constraints, in this study, we implemented OBI in a virtual reality space and designed three forms: (1) surface touch, (2) object movement, and (3) object + hand gestures. 21 university students participated in comparative experiments with menu-based user interfaces (UIs) and other systems, conducted five times each across three tasks (calendar date selection, music identification and playback operation, and the presence/absence of mouse gestures) using a head-mount display (VIVE Focus Vision) and Manus Quantum Metagloves, and their completion time, number of errors, and NASA Task Load Index were evaluated. As a result, for Pattern 1 centered on touch, OBI was effective in reducing time and errors; moreover, for Pattern 2 involving object movement, the menu-based UI had a lower workload, and for Pattern 3, gestures alone were superior. From these results, it was demonstrated that while OBI tends to be effective for simple direct operations, it can increase the burden in tasks involving an increased amount of physical movement.

1. Introduction

In recent years, the spread of immersive technologies such as virtual reality (VR), augmented reality (AR), and mixed reality (MR) has increased opportunities to handle information, manipulate objects, and perform tasks within virtual spaces. Conversely, most current IT devices are based on planar operations—such as displays and mice or touch panels—where “functions” (what can be done) and “operating units” (where and how to touch) are often separated. As a result, users must mentally map which operations lead to specific outcomes, making intuitive operation difficult for beginners or nonexperts. Particularly in environments that presuppose physical movement, such as VR/MR, simply importing 2D-centric user interfaces (UIs) may detract from immersion and increase operational hesitation (search) and workload.

*Corresponding author: e-mail: yusuke.kajiwara@komatsu-u.ac.jp
<https://doi.org/10.18494/SAM6158>

To address this issue, we focus on AR and MR technologies. Since MR allows for the generation of virtual objects that can be treated as physical “things” existing in real space, it enables a design that “embeds functions within objects”. Specifically, users are expected to execute functions through physical actions—such as touching, grasping, and moving—using the target object itself as a cue, rather than selecting abstract menu items on a screen. While recent reports have analyzed 3D UI design patterns and research challenges in consumer VR games to provide suggestions for the research community,⁽¹⁾ game UIs are heavily affected by experience value and staging. To discuss effectiveness as a versatile UI method, it is necessary to systematically evaluate it under more general task conditions using both objective and subjective metrics.

Research on menu-type interfaces in VR is progressing by comparing representative methods such as raycasting, direct input, and marking menus.⁽²⁾ While these comparisons are useful for clarifying which menu method is advantageous under specific conditions, they are fundamentally based on the framework of “selecting functions from a list”, where functions and operation targets tend to be separated.

Therefore, we propose an object-based interface (OBI) as a concept to shift the function selection itself from menus to “objects”. OBI is an interface where functions are inherent to the object, allowing users to operate it by directly touching or moving it by hand.

While implementation in an MR environment is assumed in this study, owing to technical constraints, we implement OBI in a VR space to evaluate its operability. Specifically, we designed three operational forms to ensure intuitive use: (1) operation via touch (direct contact), (2) operation involving object movement, and (3) operation combining object manipulation with hand gestures. To verify OBI’s operability, we prepared tasks comparable to conventional menu interfaces and gesture operations, and evaluated them through experiments.

In the evaluation, workload factors such as “fatigue”, “frustration”, and “effort” are as crucial as operational speed. Thus, we use the NASA Task Load Index (NASA-TLX),⁽³⁾ a representative method of multifaceted workload measurement, to organize the characteristic differences between methods using both objective metrics (completion time and the number of errors) and subjective metrics (workload). The objective is to distinguish the conditions under which OBI is advantageous or disadvantageous and contribute knowledge to design guidelines for object-centric UIs in VR/MR.

In this study, OBI, conventional menu interfaces, and gesture operations were compared as UIs in VR space. In user experiments, multiple tasks (e.g., date selection, the activation of target objects, and gesture input) were set. We determined completion times and error counts, and evaluated subjective workload using NASA-TLX. Results confirmed that under operational conditions centered on touch, OBI contributes to reduced operation time and errors. On the other hand, under conditions involving movement volume, such as object relocation, OBI showed a tendency for increased time and higher subjective workload. These findings suggest that while OBI is effective for “simple direct operations after search”, caution is required in the design for tasks that inherently increase physical workload.

The contributions of this paper are as follows.

- An evaluation framework capable of multifaceted examination of differences between methods by combining objective metrics (time and errors) and subjective workload via NASA-TLX⁽³⁾ is presented.
- OBI was demonstrated to be advantageous under touch-centered conditions, suggesting environments where OBI application is promising (situations completed with simple direct operation).
- The application limits and design precautions for OBI (increase in search, movement volume, and burden) were clarified by showing that the load may increase in situations involving object movement.

This study has social significance in that a UI design direction that enhances intuitiveness and accessibility for users unfamiliar with VR by shifting VR operations from “menu selection” to “operation by directly touching the target object” was demonstrated. Furthermore, through evaluation using NASA-TLX alongside speed and accuracy, criteria for UI selection, including “low fatigue” as well as “speed”, are provided, contributing to the creation of design guidelines for VR/MR systems intended for practical use in education, training, remote work, and business support.

2. Related Work

2.1 Menu-based interfaces in VR/MR

In VR/MR environments, UIs are essential to facilitate “control operations”, such as function calling, setting changes, and mode switching, in addition to the manipulation of virtual objects. Menu-based UIs have been widely adopted for this purpose, with the focus of design being the combination of elements such as presentation position (world-fixed vs body-referenced), input methods (e.g., raycasting, direct input, and gestures), layouts (e.g., linear and radial), and hierarchical structures (single layer vs multilayer).

To provide a foundation for overviewing the design space of menu-based UIs, Dachsel and Hübner provided classification axes and a taxonomy for 3D menus, systematizing the options for menu implementation in 3D applications.⁽⁴⁾ Comparative evaluations between menu methods are also progressing. Referring to the actual state of implementation in commercial VR applications, Wentzel *et al.* compared representative archetypes—raycasting, direct input, and marking menus—and demonstrated that speed and user experience vary depending on the menu hierarchy and the number of items.⁽²⁾ Furthermore, Santos *et al.* compared multiple VR menu methods and reported the characteristics of each from the perspectives of task completion time, error rate, and subjective evaluation.⁽⁵⁾

As described above, research on menu-based UIs has accumulated a wealth of knowledge regarding the optimization of input methods and layouts within the framework of “selecting from a list”. On the other hand, because menus present functions in an abstract manner, they have a structure in which the operation target (object) and the function (command) are easily separated. Therefore, a design that integrates functions into objects and uses the objects themselves as cues for operation (the OBI in this study) is based on a design philosophy different

from that of conventional menu-based UIs, and it is valuable to clarify its effective conditions and limitations through comparison.

2.2 Diegetic UI

In contexts where the quality of the VR experience (the immersion and consistency of the world-view) is emphasized, “diegetic UI” is widely used. This approach integrates the UI as objects, body representations, or environmental elements within the virtual world, rather than overlaying them on the screen. Steed *et al.* extracted 3D UI design patterns from consumer VR games and organized the challenges that should be addressed as research topics.⁽¹⁾ This perspective provides a direction for integrating the UI as a constituent element of the experienced world, rather than presenting it as a “separate control panel”.

The effects of diegetic UI are not uniform, and evaluations vary depending on the presented content and the task. Köhle *et al.* used health information presentation in a VR shooter as a subject to compare non-diegetic (e.g., fixed bars) and diegetic (e.g., wristwatch displays and bodily reactions) UIs, demonstrating that benefits in experience and challenges in readability and efficiency can coexist.⁽⁶⁾ Furthermore, research comparing UI modalities in VR (e.g., 2D, 3D, and voice) has revealed that evaluation axes such as efficiency, enjoyment, and ease of understanding can compete with one another.⁽⁷⁾

The findings of the above studies suggest that the evaluation of VR UIs should be examined in terms of both experience value and workload, rather than being determined solely by operation time. The OBI in this study exhibits an affinity with diegetic UI because it integrates the UI into the environment by assigning functions to objects. On the other hand, benefits in the experiential aspect do not always directly lead to improvements in efficiency or workload. Therefore, it is important to compare OBI with menu-based UIs under the same conditions and organize them from the perspectives of performance and task load.

2.3 Tangible UI

Design philosophies based on embodiment include tangible user interfaces (TUIs), direct manipulation, and reality-based interaction (RBI). Ishii and Ullmer’s “Tangible Bits” presented a vision of coupling digital information (bits) with physical objects (atoms), enabling people to handle information using their innate skills for “grasping and moving”.⁽⁸⁾ Fitzmaurice *et al.*’s “Bricks” introduced the concept of directly controlling virtual objects via graspable physical handles, demonstrating the potential of operations using physical proxies.⁽⁹⁾ Furthermore, through a survey of TUI research, Shaer and Hornecker organized concepts, applications, and evaluation frameworks, showing that TUI covers a broad design domain encompassing “perception, physical movement, and tool use”.⁽¹⁰⁾ Additionally, Hornecker and Buur categorized tangible interaction from the perspectives of physical space and social interaction, discussing the importance of experience, collaboration, and spatial configuration.⁽¹¹⁾

Direct manipulation is a design concept that creates the sensation of directly handling an object through “continuous operations on visible targets and immediate feedback”, a principle

whose effectiveness was emphasized by Shneiderman.⁽¹²⁾ On the other hand, Hutchins *et al.* discussed the advantages of direct manipulation from a cognitive perspective while also addressing its limitations, such as abstraction and representational constraints.⁽¹³⁾ RBI is a framework systematized by Jacob *et al.* that enhances ease of learning by incorporating knowledge and skills developed in the real world (e.g., physical laws, bodily movements, and spatial behaviors) into UI design.⁽¹⁴⁾

In UIs that rely on space, while the user's spatial memory may contribute to the operation, costs may arise in terms of searching and movement. Cockburn and McKenzie experimentally examined the effects of spatial memory in 2D and 3D spaces, and showed that spatial configuration can serve as an effective cue.⁽¹⁵⁾

What the results of these related studies indicate is a duality: designs leveraging embodiment, object-centricity, and spatial cues can enhance intuitiveness and ease of learning, yet the resulting increase in expression and physical movement volume may become a burden. The OBI in this study is positioned as an attempt to realize an “object-centric UI” within VR that aligns with the philosophies of TUI, direct manipulation, and RBI by assigning functions to objects without mediating through abstract “function lists” known as menus.

2.4 Positioning of this study

Previous studies have advanced the classification and comparison of VR menu-based UIs, and demonstrated that speed and user experience vary according to the input method, hierarchy, and the number of items. Additionally, with the spread of consumer VR games, “diegetic UI”—a design that integrates the UI within the world—has become common, highlighting the close relationship between experience value and UI design. Furthermore, research on TUI, direct manipulation, and RBI provides theoretical and empirical frameworks supporting the improvement of intuitiveness and ease of learning through object-centricity and the utilization of physical skills.

However, despite these findings, there is no sufficiently organized framework that enables the comparison of OBI”—which integrates functions into virtual objects—with conventional menu-based UIs under identical conditions to distinguish its advantages and limitations in terms of both objective metrics and subjective workload.

Therefore, in this study, we propose OBI and compares it with menu-based UIs to clarify the conditions under which OBI is advantageous (where cuing and intuitiveness are effective) and disadvantageous (where search, movement volume, and representational constraints become a burden), aiming to derive application guidelines for object-centric UIs in VR/MR. In the evaluation, in addition to objective metrics such as operation time and error counts, we measure subjective workload using the Japanese version of NASA-TLX,⁽¹⁶⁾ performing a comparison from the perspectives of both experience value and task burden.

3. Proposed Methodology

In this study, we propose OBI. The most significant feature of our OBI is that it centers on “objects that integrate functions and operation targets” as the primary subjects of interaction. By

touching objects—such as desk calendars, illustrations, and other items familiar in daily life—users can invoke or manipulate specific computer functions. By embedding functions within the objects themselves, it is expected that users can more intuitively understand “what to operate and how to operate it” on the basis of the object’s shape, position, and the way it is touched, providing a more natural operational feel than conventional planar GUIs.

OBI is structured around the following three operational forms.

(1) Touching the surface (side) of an object

This is the most basic form, where functions are invoked by touching or pressing the surface of an object. For example, by assigning a specific function to the side of a desk calendar, operations such as selecting a date or expanding a menu can be performed by touching that area. It is also possible to divide the touchable surface in accordance with the object’s shape and size to assign multiple functions. Furthermore, in addition to simple touch operations, combinations with more diverse gestures such as swiping and dragging can be considered. A conceptual diagram is shown in Fig. 1. This figure illustrates the concept of selecting a date on a desk calendar.

(2) Operations based on combinations of objects and positions

Next is a form in which functions are invoked by moving an object from one location to another. Inspired by everyday actions—such as rearranging paper documents or returning a book to a shelf—actions are triggered by moving an object to a specific position. Using spatial information—specifically, where an object has been placed—as a cue, users can more naturally recall subsequent operations. Since this leverages memory based on bodily sensations (kinesthetic memory), an improvement in operability is expected. A conceptual diagram is shown in Fig. 2. The blue square represents the object and the white squares represent the designated positions for invoking functions. By moving the object to a designated position, a predetermined function can be used.



Fig. 1. (Color online) Concept of OBI operation by directly touching the surface of an object.

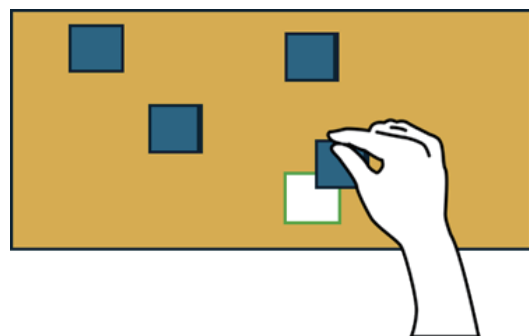


Fig. 2. (Color online) Concept of OBI operation by moving an object to a designated position.

(3) Combined operation using objects and hand gestures

Finally, we propose a form that invokes complex functions by performing hand gestures while holding an object. For example, by performing a “click-like” gesture with the fingertips while holding a physical object, complex operations—such as switching modes or moving and selecting a cursor—can be realized. It is believed that by merging the sensation of actually holding an object with the movements of the hands and fingers, the sense of operability for the entire UI will be further enhanced.

4. Experiment

4.1 Purpose

The purpose of this study is to compare the operability of OBI and menu-based interfaces and to present the strengths and challenges of OBI. The results of this study are expected to provide beneficial suggestions for identifying the optimal interfaces in MR and AR.

In this experiment, the following three patterns of tasks were conducted to compare each of the three proposed items (1) to (3): a task to select a date on a calendar (Pattern 1), a task to listen to music and perform operations to play that music (Pattern 2), and an operation using mouse gestures (Pattern 3). The specific experimental procedures are shown in Table 1. To minimize the effects of habituation, participants practiced sufficiently for each case and then performed five measured trials per case. For all problems, the task was continued until it was successfully achieved. Additionally, a 5 s interval was provided between trials. After completing two patterns, a 5 min break was taken. The order of the patterns was randomized for each participant using pseudo-random numbers, which were generated as uniform random numbers. The

Table 1
Experimental procedures.

No.	Item
1	Experiment description
2	Pre-experiment questionnaire
3	Installation of manus quantum metagloves
4	Calibration of the manus quantum metaglove
5	Vive focus vision adjustment (IPD adjustment)
6	Experimental practice for cases $i-j$ ($i = 1, 2, 3, j = 1, 2$) (2 patterns)
7	Case ij : live experiment (2 patterns)
8	Post-experiment questionnaire i
9	Break
10	Vive focus vision adjustment (IPD adjustment)
11	Experimental practice for cases $i-j$ ($i = 1, 2, 3, j = 1, 2$) (1 patterns)
12	Case ij : live experiment (1 pattern)
13	Post-experiment questionnaire i
14	Installation of manus quantum metagloves
15	Calibration of the manus quantum metaglove

measured data included start time, end time, and the number of errors for each trial; furthermore, the Japanese version of NASA-TLX was administered as a post-experiment questionnaire using a 7-point scale.

4.2 Environment

Figure 3 shows (1) VIVE Focus Vision, (2) Manus Quantum Metagloves, (3) VIVE Tracker, (4) desktop PC, (5) desk, (6) chair, (7) base stations, (8) monitor, (9) participant, and (10) experimenter. Participants wore the VIVE Focus Vision and the Manus Quantum Metagloves only on their dominant hand. However, in the mouse-click gesture experiment, all participants performed the task with their right hand, as even left-handed individuals generally use a mouse with their right hand. The VIVE Focus Vision was operated via PC streaming using VIVE Business Streaming 2.2.4 beta. The height of the physical desk was adjusted to match that of the virtual desk. Regarding the distance between the chair and the desk for Patterns 1 and 2, participants were allowed to move forward or backward during the practice period to find a comfortable position that minimized any decrease in operability related to arm length, and a fixed position was determined within the sensor's tracking range. The starting hand position was specified as being on the participant's lap. Pattern 3 was conducted with the participant's hands resting on the physical desk. The base stations were placed diagonally and adjusted in height to ensure there were no blind spots.

4.3 Methods

4.3.1 Experimental method for Pattern 1

Figures 4 and 5 show the application screens for OBI and the menu-based interface for Pattern 1, respectively. For Pattern 1, a question is presented in text format, and the task requires

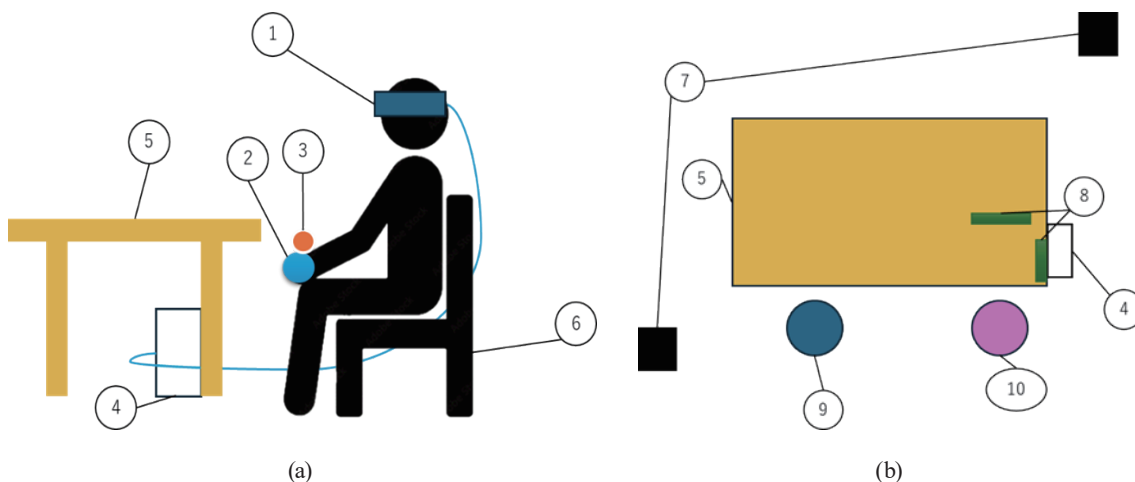


Fig. 3. (Color online) Experimental environment: (a) horizontal and (b) plan views.



Fig. 4. (Color online) OBI scene for Pattern 1: calendar date selection task.

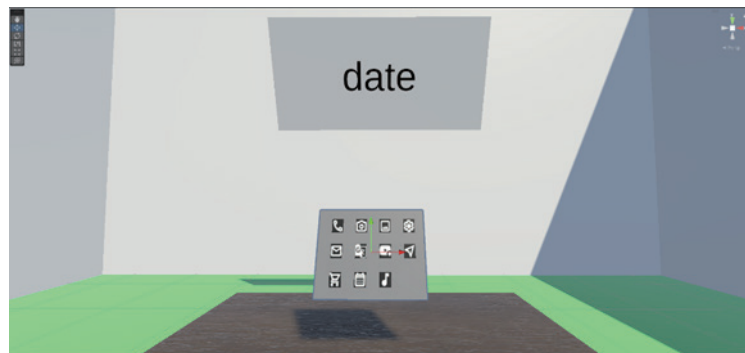


Fig. 5. (Color online) Menu-based interface scene for Pattern 1: calendar date selection task.

the participant to operate the respective interface to select the same date as specified in the question until the correct answer is achieved. In OBI, a desk calendar and other objects, totaling 11 items, are placed on a desk object. The participant must locate the desk calendar from among the multiple objects and touch the date corresponding to the question using their index finger. Regarding functionality, only touching the calendar dates is enabled; no functions are assigned to the other objects. Furthermore, no function for picking up objects is provided. In the menu-based interface, the operation follows the sequence “press the menu button, press the calendar icon on the menu screen, and press the calendar date”. The number of icons on the menu screen, including the calendar icon, was set to 11 to match OBI. In both interfaces, a confirmation screen is displayed upon a correct or incorrect answer. If the answer is incorrect, an “X” button appears in the upper right of the confirmation screen; the participant must press it and redo the previous operation until the correct answer is reached. Additionally, the positions of the objects in OBI and the icons on the menu screen in the menu-based interface are randomized for each trial.

4.3.2 Experimental method for Pattern 2

Figures 6 and 7 show the application screens for OBI and the menu-based interface for Pattern 2, respectively. The OBI condition for Pattern 2 intentionally required three actions—



Fig. 6. (Color online) OBI scene for Pattern 2: music selection task.



Fig. 7. (Color online) Menu-based interface scene for Pattern 2: music selection task.

selecting the correct object, grasping it, and placing it at a designated location—because this pattern corresponds to the second proposed OBI form: operation based on the combination of an object and a position. The purpose was not to make the number of operation steps identical to that of the menu-based interface, but to examine whether the spatial placement of an object can function as an intuitive operation for invoking a function. Therefore, Pattern 2 evaluates the trade-off between the potential intuitiveness of object-placement operation and the additional physical task cost caused by grasping and moving the object.

For Pattern 2, a question is presented via audio (music), and the task requires participants to operate the objects or icons corresponding to the music in each interface until the correct answer is achieved. There are a total of eight background music (BGM) sound sources used for the questions. The sound sources used in the questions are shown in Table 2. The illustrations used for the objects and icons were related to music. Participants were asked to memorize the correspondence between the sound sources and the objects or icons before the practice session. In OBI, objects with illustrations are placed on a desk object in a number equal to the number of sound sources. The operation involves selecting the object that matches the music from among multiple objects, grasping it, and placing it at a specific location (indicated by a white frame). If the answer is correct, the music stops; however, if it is incorrect, the music continues to play, requiring the participant to move the object from the specific location to another spot and repeat

Table 2
BGM sound sources used.

1	Happy Birthday to You
2	Jingle Bells
3	Orpheus in the Underworld
4	Grandfather's Clock
5	Twinkle, Twinkle, Little Star
6	Buzz, Buzz, Buzz
7	The Other Day, I Met a Bear
8	Froggy's Song

the process until the correct answer is reached. For the menu-based interface, the number of icons on the menu screen was set equal to the number of sound sources. The operation follows the sequence “press the menu button, press the music icon”. In both interfaces, a confirmation screen is displayed upon a correct or incorrect answer. If the answer is incorrect, an “X” button appears in the upper right of the confirmation screen; the participant must press it and redo the previous operation until the correct answer is reached. Additionally, in the OBI scene, a reset button is placed on the right side to return objects to their initial positions in case they are dropped off the desk. Participants were told during the experiment that this was for emergency use only. In this study, it was never used.

4.3.3 Experimental method for Pattern 3

Figure 8 shows the application screen for Pattern 3. For Pattern 3, questions are presented in text format, and the task requires participants to select the corresponding icon using a gesture similar to a mouse click, performed under two conditions, namely, hand gesture with physical objects and hand gesture without physical objects (Figs. 9 and 10, respectively), until the correct answer is achieved. There are 16 icons in total, divided into a first and a second page. The icon names used were intuitive and familiar from daily life. Participants were asked to memorize the icons and their names before the practice session. In both environments, participants were instructed to start with their hands on the desk. The only operations possible with this gesture are cursor movement by moving the hand forward, backward, left, or right, and a click operation performed by moving the index finger. Under the “hand gesture with physical objects” condition, no electronic functions were assigned to the mouse itself. In both cases, a confirmation screen is displayed upon a correct or incorrect answer. If the answer is incorrect, an “X” button appears in the upper right of the confirmation screen; the participant must press it and redo the previous operation until the correct answer is reached.

4.4 Subjects

A total of 21 university students (18 males and 3 females) participated in the study. According to the pre-experiment questionnaire, the participants' attributes were as follows: age [mean (M)



Fig. 8. (Color online) BI scene for Pattern 3.



Fig. 9. (Color online) Pattern 3 condition: hand gesture with a physical object.



Fig. 10. (Color online) Pattern 3 condition: hand gesture without a physical object.

= 21.8 years, $SD = 1.05$]; VR experience (eight with no experience, nine with experience of about once a year, one with experience of about once a month, two with experience of about once a week, and one almost every day); hand circumference ($M = 199.9$ mm, $SD = 11.8$); hand length ($M = 186.8$ mm, $SD = 9.6$); and height ($M = 1689.0$ mm, $SD = 66.7$). 20 participants were right-handed and one was left-handed. Written informed consent was obtained from all participants.

4.5 Post-experiment questionnaire

For the post-experiment questionnaire, the Japanese version of NASA-TLX was used, as in Ref. 3. Prior to the experiment, participants were asked to read Table 1 of Ref. 3 to understand the nuances of each evaluation item. Subsequently, the experimenter asked simplified versions of

each question orally, and participants were instructed to provide their ratings on a 7-point scale, using a score of 4— corresponding to their everyday computer operation—as the baseline.

4.6 Analysis

4.6.1 Pre-experiment questionnaire

The mean and standard deviation for age, hand circumference, hand length, and height obtained from the pre-experiment questionnaire were calculated.

4.6.2 Completion time

During the experiment, data for the start and end times were collected for each of the five trials conducted per case. From these data, the completion time was determined, and the mean and standard deviation were calculated for each case. Furthermore, using the average completion time across the five trials for each participant per case as the data point, an F-test was performed for each pattern. The null hypothesis was that “there is no difference between the population variances of the two groups”, with a significance level of 5%. Subsequently, for cases where the null hypothesis was not rejected by the F-test, Student’s t-test assuming equal variance was performed. For cases where it was rejected, Welch’s t-test was performed. The null hypothesis for these tests was that “there is no difference between the mean values of the two groups”, with a significance level of 5%.

4.6.3 Number of errors

When an incorrect selection that did not follow the shortest path to the correct answer was made during the experiment, this was recorded as error count data. From this data, the mean error count and standard deviation were calculated for each case.

4.6.4 Post-experiment questionnaire

Using the data obtained from the Japanese version of NASA-TLX, the mean and standard deviation of the overall scores were calculated. Additionally, the mean and standard deviation for each subscale of the Japanese version of NASA-TLX were calculated. For both cases, the Wilcoxon signed-rank test was performed for each pattern. The null hypothesis was that “the median of the differences between the two groups is zero”, with a significance level of 5%.

5. Results

5.1 Completion time

Table 3 shows the mean and standard deviation of the completion time, as well as the results of the F- and t-tests. The F-test showed a significant difference in completion time for Pattern 1.

Table 3
Experimental results for completion time.

		Mean (s)	SD (s)	<i>f</i> -value	<i>p</i> -value	<i>t</i> -value	<i>p</i> -value
Pattern 1	OBI	3.285	1.322	8.988	0.000	−9.108	0.000
	Menu	6.139	2.103				
Pattern 2	OBI	5.646	1.999	1.472	0.395	3.649	0.001
	Menu	4.450	1.633				
Pattern 3	Hand gesture with physical objects	5.913	2.455	1.711	0.238	5.384	0.000
	Hand gesture without physical objects	3.844	2.19				

Additionally, the *t*-test revealed significant differences across all patterns. The results indicated that for Pattern 1, OBI exhibited a shorter mean completion time; for Pattern 2, OBI showed a longer mean completion time; and for Pattern 3, the “with object” condition resulted in a longer mean completion time.

5.2 Number of errors

Table 4 shows the results of the mean and standard deviation for the number of errors. Regarding the mean number of errors, a substantial difference was observed For Pattern 1, with OBI resulting in fewer errors. For Pattern 2, OBI also had fewer errors, and for Pattern 3, the “with-object” condition resulted in more errors; however, in both of these cases, almost no notable difference was observed.

5.3 Post-experiment questionnaire

Table 5 shows the results of the mean and standard deviation for the overall workload scores in the Japanese version of NASA-TLX. The mean overall workload scores were higher for the OBIs of Patterns 1 and 2, and for the “with object” condition of Pattern 3. At a 5% significance level using the Wilcoxon signed-rank test, no significant difference was observed for Pattern 1. For Patterns 2 and 3, the null hypothesis was rejected at the 5% significance level.

Table 6 shows the results of the mean, standard deviation, test statistics from the Wilcoxon signed-rank test, and *p*-values for each item in the Japanese version of NASA-TLX. At a 5% significance level using the Wilcoxon signed-rank test, no significant difference was observed for Pattern 1. For Pattern 2, the null hypothesis was rejected for mental demand, physical demand, temporal demand, effort, and overall burden. For Pattern 3, it was rejected for physical demand, performance, effort, and overall burden.

6. Discussion

6.1 Pattern 1

We discuss the evaluation results for Pattern 1 of this study. The results of the evaluation of completion time showed that it was, on average, shorter for OBI. This is likely because, in OBI,

Table 4
Experimental results for number of errors.

		<i>Mean</i> (time)	<i>SD</i> (time)
Pattern 1	OBI	0.333	0.483
	Menu	1.286	1.678
Pattern 2	OBI	0	0
	Menu	0.381	0.805
Pattern 3	Hand gesture with physical objects	0.381	0.740
	Hand gesture without physical objects	0.190	0.512

Table 5
Experimental results for overall workload scores.

		<i>Mean</i>	<i>SD</i>	Statistics	<i>p</i> -value
Pattern 1	OBI	4.476	0.605	75.5	0.179
	Menu	4.246	0.612		
Pattern 2	OBI	4.587	0.646	36	0.004
	Menu	4.095	0.581		
Pattern 3	Hand gesture with physical objects	4.421	0.656	50.5	0.024
	Hand gesture without physical objects	3.921	0.666		

the operation only requires searching for the target object and simply tapping the calendar date. In contrast, conventional menu-based interfaces require hierarchical operations—selecting the menu button, selecting the application, and then tapping the calendar date—which is considered to consume extra time. Furthermore, as suggested in Ref. 6, the faster completion may be affected not only by the complexity of the hierarchical structure but also by the possibility that spatial memory is more effective than menu screens.

The mean number of errors was also lower for OBI. It is inferred that in the menu-based interface, factors such as the complexity of the hierarchical structure and the size of each operation target may have made accidental taps more likely to occur. Moreover, the subjective workload evaluation using NASA-TLX showed no significant differences between any of the items.

These results suggest that for Pattern 1 involving touch operations, OBI—which arranges objects in 3D space—is more useful than conventional menu-based interfaces in terms of both task completion time and the number of errors. For Pattern 1, the OBI condition was effective because the object cue and the required operation were directly associated. Participants could interpret the calendar object as the target of date selection and complete the task by directly touching the required date. In this case, the object reduced the need to search through a hierarchical menu structure. Therefore, the task cost was mainly limited to visual search and direct touch, and the cognitive cost of mapping a command to an operation target was relatively low. This suggests that OBI is appropriate when the function of an object is semantically clear and the required operation can be completed by direct contact.

Table 6
Experimental results for each item in the Japanese version of NASA-TLX.

		Mental Demand				Physical Demand			
		Mean	SD	Statistics	p-value	Mean	SD	Statistics	p-value
Pattern 1	OBI	4.714	1.030	19.5	0.404	5.571	0.849	23	0.053
	Menu	4.476	0.906			5.000	1.024		
Pattern 2	OBI	4.952	0.999	13.5	0.039	5.905	0.811	6.5	0.001
	Menu	4.333	0.891			5.143	0.639		
Pattern 3	Hand gesture with physical objects	4.190	0.732	0.663	0.860	5.095	0.868	5	0.003
	Hand gesture without physical objects	4.190	0.663			4.333	0.943		

Temporal Demand				Performance			
Mean	SD	Statistics	p-value	Mean	SD	Statistics	p-value
4.048	0.898	37	0.872	3.000	1.380	36	0.475
4.048	1.362			2.810	1.401		
4.429	0.904	4.5	0.021	2.524	1.531	38.5	0.606
4.000	0.926			2.333	1.127		
4.048	0.950	23	0.179	3.095	1.509	32.5	0.030
3.857	1.037			2.333	1.039		

Effort				Frustration			
Mean	SD	Statistics	p-value	Mean	SD	Statistics	p-value
5.048	0.950	47.5	0.462	4.476	1.180	28.5	0.378
4.810	1.052			4.333	1.247		
5.238	0.868	21.5	0.025	4.476	1.258	12.5	0.221
4.524	1.096			4.238	1.306		
5.238	0.971	17	0.013	4.857	1.082	49	0.180
4.381	1.174			4.429	1.218		

Overall burden			
Mean	SD	Statistics	p-value
4.762	1.019	36	0.140
4.476	0.79		
5.095	1.151	30	0.033
4.571	1.094		
4.952	0.898	30	0.010
4.286	1.075		

6.2 Pattern 2

For Pattern 2, the results showed that the mean completion time was longer for OBI. This is likely because operations in OBI involve the action of the user personally carrying and moving objects; compared with a menu-based interface that requires only simple touch operations, the movement distance and physical workload are increased.

The mean number of errors for OBI was 0, whereas the menu-based interface had a mean of approximately 0.3, indicating a slight occurrence of errors. However, as this difference is small, it can be considered that there is no substantial difference in the frequency of operational errors. When using the menu screen, mis-taps during touch operations may have been the cause.

Furthermore, in the NASA-TLX evaluation, the overall workload score for OBI was higher than that for the menu-based interface. When the Wilcoxon signed-rank test was conducted for each subscale (mental demand, physical demand, temporal demand, effort, or overall burden), a significant difference in median was observed. Specifically, the following points can be noted.

- **Mental Demand:** It is considered that cognitive load increased in OBI because the manipulation of moving target objects is more complex than simple touch operations.
- **Physical Demand:** In OBI, as the operation targets move across the screen, it is necessary to move the arms and fingers more extensively; it is inferred that the physical workload increased compared with conventional touch operations.
- **Temporal Demand:** Since objects are scattered across the screen in OBI, the process of grasping the distances and spatial relationships between targets may have heightened the sense of time pressure.
- **Effort:** It is considered that the effort required for OBI operations was rated higher owing to the combined effects of the increased demands mentioned above and differences in habituation.

Taking these points into account, the results suggest that for Pattern 2, a menu-based interface using only simple touch operations is more suitable in terms of user workload and completion time than OBI that combines object manipulation with spatial relationships. Regarding task cost, OBI imposes a higher workload because it involves the actions of picking up and placing objects, whereas the menu-based interface requires only simple touch operations. Concerning cognitive cost, OBI is considered to be more demanding because users must consciously perform the unfamiliar action of grasping virtual objects in VR without the tactile feedback of physical items. While there is a possibility that cognitive cost could be reduced in an MR environment—where users can perform the familiar act of holding real objects—further experiments are necessary to demonstrate any superiority over menu-based interfaces.

Furthermore, although the objects in this experiment were placed randomly, OBI might prove to be more effective than a menu-based interface if the objects are organized in a regular and orderly manner. The OBI condition for Pattern 2 involved more physical operation steps than the menu-based interface because it was designed to evaluate an object-and-position-based operation, which is one of the proposed forms of OBI. In this operation, the user invokes a function by selecting an object and placing it at a designated position. Therefore, Pattern 2 should be interpreted as an evaluation of the trade-off between the intuitive correspondence of object placement and the additional task cost caused by grasping and moving the object. The results indicate that, under the current implementation, the additional actions of grasping and placing the object were not compensated by the benefit of object-based cues. Therefore, the OBI

design used for Pattern 2 was less suitable than the menu-based interface for this particular music selection task. This finding clarifies an important limitation of OBI: when a function can be invoked by a simple menu selection, adding object movement may increase task cost rather than improve operability.

6.3 Pattern 3

For Pattern 3, the mean completion time was longer under the hand gesture with physical objects condition, and the number of errors also increased under the same condition. Furthermore, post-experiment questionnaire results confirmed that the overall subjective workload was higher under the hand gesture with physical objects condition.

The primary factor behind these results is likely the significant gap between the experimental setup and the mouse operations that users are accustomed to in daily life. Conventional mice are ergonomically designed and optimized to eliminate the need for large finger movements during click operations. In contrast, while the mouse used in this experiment followed the conventional design, its operational feel and the load experienced when physically holding it differed from normal mouse usage, potentially imposing an extra burden on the users.

Although data on “missed clicks” (instances where a click was performed but no response was triggered) were not collected in this experiment, such phenomena were observed during the trials. We believe that it is necessary to properly collect and analyze such data in detail in the future. This result suggests that simply adding a physical object does not necessarily improve operability.

In this task, the physical object did not provide an additional functional cue that might have helped the user understand the operation. Instead, holding the object may have constrained finger movement and introduced a mismatch between the familiar experience of using a mouse and the gesture-based input used in the experiment. Therefore, for gesture-based interaction, OBI should be designed so that the physical object clearly supports the intended action rather than merely being held by the user.

6.4 OBI design implications based on task cost and cognitive cost

The findings suggest three design guidelines for OBI in VR/MR environments. First, OBI is suitable for tasks in which the object itself clearly indicates the function and direct touch can complete the operation. Second, OBI should be carefully applied to tasks that require object movement, because the increase in physical movement may outweigh the benefit of intuitive object cues. Third, when OBI is combined with hand gestures, the physical object should be designed to support the gesture mechanically, perceptually, or semantically; otherwise, it may increase operational burden. In this sense, OBI should be regarded not as a universally superior alternative to menus, but as a design approach whose effectiveness depends on the balance between intuitive object cues and additional task costs.

In addition, OBI may be particularly effective for tasks in which location information itself is the target of input. For example, arranging objects, specifying positions, placing markers, or

assigning information to locations may benefit from OBI because the user's spatial action directly corresponds to the information being entered. The present Pattern 2 task used object placement as a method for selecting music, but the task goal itself was not location input. Therefore, future experiments should evaluate tasks in which spatial placement directly represents the intended information. Such experiments would allow a fairer assessment of the advantages of OBI over menu-based interfaces for location-related operations.

7. Conclusions

In this study, we proposed OBI and evaluated its operability through comparative experiments with conventional menu-based interfaces and gesture-based operations in a VR environment. The first contribution of this study is that it presented an evaluation framework for OBI that combines objective performance measures, such as completion time and the number of errors, with subjective workload measured using NASA-TLX. This framework made it possible to examine not only whether an interface was fast or accurate, but also whether it imposed mental, physical, or temporal burden on users.

The second contribution is that the experiment demonstrated the effectiveness of OBI under touch-centered conditions. For Pattern 1, OBI reduced both completion time and the number of errors compared with the menu-based interface. This result suggests that OBI is effective when the object itself provides a clear cue for the operation and when the task can be completed by direct touch.

The third contribution is that the experiment clarified the limitations and design precautions of OBI. For Pattern 2, which required participants to select, grasp, and place an object, the menu-based interface showed a lower workload and shorter completion time than OBI. For Pattern 3, hand gesture without a physical object was superior to hand gesture with a physical object. These results indicate that OBI is not universally superior to menu-based or gesture-based interfaces. Rather, its effectiveness depends on the balance between intuitive object cues and additional task costs, such as physical movement, spatial manipulation, and constraints on hand or finger motion.

Future work will include increasing the number and diversity of participants, improving the gender balance, adding objective behavioral metrics such as hand travel distance, touch position, and missed-click frequency, and conducting experiments in AR and MR environments. In addition, future studies should include evaluations of tasks in which OBI can be expected to be more suitable than menu-based interfaces, such as entering location information, placing objects at specific positions, arranging multiple objects in space, and assigning functions to spatial regions.

References

- 1 A. Steed, T. M. Takala, D. Archer, W. Lages, and R. W. Lindeman: *IEEE Trans. Vis. Comput. Graph.* **27** (2021) 4171. <https://doi.org/10.1109/TVCG.2021.3106431>
- 2 J. Wentzel, M. Lakier, J. Hartmann, F. Shazib, G. Casiez, and D. Vogel: *IEEE Trans. Vis. Comput. Graph.* **31** (2025) 4868. <https://doi.org/10.1109/TVCG.2024.3420236>
- 3 S. Haga and N. Mizukami: *Jpn. J. Ergon.* **32** (1996) 71. <https://doi.org/10.5100/jje.32.71> (in Japanese).
- 4 R. Dachselt and A. Hübner: *Comput. Graph.* **31** (2007) 53. <https://doi.org/10.1016/j.cag.2006.09.006>

- 5 A. Santos, T. Zarranandia, P. Díaz, and I. Aedo: Proc. 2017 ACM Int. Conf. Interactive Surfaces and Spaces (ACM, 2017) 294. <https://doi.org/10.1145/3132272.3132277>
- 6 K. Köhle, M. Hoppe, A. Schmidt, and V. Mäkelä: Proc. IFIP Conf. Human-Computer Interaction, Lecture Notes in Computer Science **12934** (Springer, Cham, 2021) 3. https://doi.org/10.1007/978-3-030-85613-7_1
- 7 D. Hepperle, Y. Weiß, A. Siess, and M. Wölfel: Comput. Graph. **82** (2019) 321. <https://doi.org/10.1016/j.cag.2019.06.003>
- 8 H. Ishii and B. Ullmer: Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 1997) 234. <https://doi.org/10.1145/258549.258715>
- 9 G. W. Fitzmaurice, H. Ishii, and W. A. Buxton: Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 1995) 442. <https://doi.org/10.1145/223904.223964>
- 10 O. Shaer and E. Hornecker: Found. Trends Hum.-Comput. Interact. 3 (2010) 1. <https://doi.org/10.1561/11000000026>
- 11 E. Hornecker and J. Buur: Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 2006) 437. <https://doi.org/10.1145/1124772.1124838>
- 12 B. Shneiderman: Computer **16** (1983) 57. <https://doi.org/10.1109/MC.1983.1654471>
- 13 E. L. Hutchins, J. D. Hollan, and D. A. Norman: User Centered System Design: New Perspectives on Human-Computer Interaction, D. A. Norman and S. W. Draper, Eds. (Lawrence Erlbaum Associates, Hillsdale, 1986) Chap. 5.
- 14 R. J. K. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum: Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 2008) 201. <https://doi.org/10.1145/1357054.1357089>
- 15 A. Cockburn and B. McKenzie: Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (ACM, New York, 2002) 203. <https://doi.org/10.1145/503376.503413>
- 16 S. G. Hart and L. E. Staveland: Human Mental Workload, P. A. Hancock and N. Meshkati, Eds. (North-Holland, Amsterdam, 1988) Chap. 7. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)