



## Article

# Exploring Data Input Problems in Mixed Reality Environments: Proposal and Evaluation of Natural Interaction Techniques

Jingzhe Zhang <sup>1</sup>, Tiange Chen <sup>1</sup>, Wenjie Gong <sup>1</sup>, Jiayue Liu <sup>1</sup> and Jiangjie Chen <sup>2,\*</sup> <sup>1</sup> School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China<sup>2</sup> School of Design, Jiangnan University, Wuxi 214122, China

\* Correspondence: chenjiangjie@jiangnan.edu.cn

**Abstract:** Data input within mixed reality environments poses significant interaction challenges, notably in immersive visual analytics applications. This study assesses five numerical input techniques: three benchmark methods (Touch-Slider, Keyboard, Pinch-Slider) and two innovative multimodal techniques (Bimanual Scaling, Gesture and Voice). An experimental design was employed to compare these techniques' input efficiency, accuracy, and user experience across varying precision and distance conditions. The findings reveal that multimodal techniques surpass slider methods in input efficiency yet are comparable to keyboards; the voice method excels in reducing cognitive load but falls short in accuracy; and the scaling method marginally leads in user satisfaction but imposes a higher physical load. Furthermore, this study outlines these techniques' pros and cons and offers design guidelines and future research directions.

**Keywords:** human–computer interaction; mixed reality; natural interaction techniques; immersive analytics; multimodal interaction techniques



**Citation:** Zhang, J.; Chen, T.; Gong, W.; Liu, J.; Chen, J. Exploring Data Input Problems in Mixed Reality Environments: Proposal and Evaluation of Natural Interaction Techniques. *Future Internet* **2024**, *16*, 150. <https://doi.org/10.3390/fi16050150>

Academic Editors: Danilo Avola, Bruno Fanini and Marco Raoul Marini

Received: 12 March 2024

Revised: 24 April 2024

Accepted: 25 April 2024

Published: 27 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Recently, mixed reality (MR) interaction has emerged as a significant research focus [1]. The proliferation of consumer-grade MR devices has spurred the application of MR technology across diverse fields [2–4]. As MR's natural interaction techniques mature, tailoring design and evaluation methods to specific application scenarios and user needs is crucial for enhancing user immersion and satisfaction [5].

Immersive analytics merges MR with natural and embodied user interaction, offering three-dimensional information presentation, expanded display spaces, and the integration of physical references with corresponding data [6]. This method demonstrates utility in education, scientific visualization, immersive workspaces, and data embedding, among others, surpassing traditional graphic visualizations in supporting user exploration and comprehension of complex 3D data [7]. Substantial research has been dedicated to developing visualization methods for assorted data types within MR environments [6,8,9], including immersive analytics applications in collaborative work [10].

Thus, this study centers on immersive analytics as its primary application scenario. To improve users' abilities to perceive, recall, and comprehend complex data, visual analytics systems may need to allow users to achieve retrieval, deletion, and other controls over complex data information by way of inputting the data information. This includes adjusting, retrieving, and inputting complex information like colors and coordinates in data visualizations. To more concretely illustrate the significance and application context of our research, we envisioned a prevalent scenario within educational and scientific visualization contexts wherein students or researchers were tasked with exploring and analyzing three-dimensional datasets through mixed reality (MR) technology. In this scenario, adjusting the RGB color values of data visualizations became essential for users to effectively differentiate data layers or emphasize particular data points. Traditional numerical input

methods, such as keyboards or standard sliders, may prove unintuitive or cumbersome within an MR setting, thus diminishing user immersion and interaction efficiency. Present studies on data input in MR environments predominantly examine homogeneous methods like keyboarding [11,12] and clicking [13]. This study identified a gap in systematic and comprehensive research on data input tasks within immersive analytics, particularly concerning interaction efficiency and user experience [5], noting that numerical information plays a crucial role in data visualization, notably in scientific contexts [14], with this gap partially limiting visual analytics development [10]. Addressing this gap, this study concentrated on numerical input challenges, devising two multimodal interaction techniques and comparing them to traditional gesture or keyboard inputs in MR devices like HoloLens2. The objective was to explore and assess numerical input methods for immersive analytics in MR, aiming to bridge existing research gaps and advance MR technology's application in data analysis and visualization.

## 2. Related Works

Mixed reality (MR) merges the real and virtual worlds [1]. MR enables users to have a natural interaction with both virtual and real objects [15]. Consequently, MR systems have incorporated advanced techniques like eye tracking, facial expression analysis, and virtual control manipulation, diversifying information output channels including gaze, gesture, and speech, allowing for their use individually or in combination to enhance interaction modality and flexibility [5,16]. This section will review the literature on natural interaction methods in MR, input research, and data interaction within MR immersive analytics.

### 2.1. Natural Interaction Methods in Mixed Reality

Advancements in computer vision technology enable the implementation of gesture recognition in MR devices with minimal hardware costs [17]. Consequently, many MR devices, such as HoloLens and Meta Quest, predominantly utilize gesture-based interactions [13]. However, studies indicate that gestural interactions demand more complex movements than traditional devices like mice or keyboards, potentially causing user fatigue and discomfort [18]. The mismatch between motor and visual spaces necessitates reliance on visual feedback for gesture adjustment, rather than solely on proprioception, which can elevate cognitive load and the likelihood of misuse [19]. Furthermore, in target-intensive or precision-required environments, gestures may inadvertently select multiple targets simultaneously, necessitating a disambiguation mechanism for accurate target selection. However, disambiguation mechanisms may further increase cognitive load [20].

Research indicates that voice interaction ranks as the second most prevalent method in MR and augmented reality (AR) [21], highlighting its advantages in terms of high ease of use and low deployment cost [22]. Yet, challenges include language ambiguity and instability in acoustically challenging environments [23], indicating general limitations. Eye-tracking interaction is increasingly integrated into new MR devices. These devices often use the dwell method for selection, facing issues like misuse and inefficiency (Midas touch problem) [24].

This study posits that combining multiple interaction modalities into a multimodal approach could enhance interaction effects. Pfeuffer et al. [25] designed an interaction modality merging eye-tracking and gesture. Wagner et al. [13] experimentally showed that eye-tracking and gesture interaction outperform single-gesture interaction in solving clicking issues efficiently.

### 2.2. Input Methods in Mixed Reality

Derby, Rarick, and Chaparro [26] conducted experiments comparing the efficiency and user ratings of Clicker (remote control) versus gesture input for textual information on HoloLens, finding that participants' typing speeds and user ratings were higher with Clicker input than with gesture input; yet, accuracy did not significantly differ between the two methods. However, the newer HoloLens2 and the recently introduced Apple Vi-

sion Pro (Los Altos, CA, USA) have eliminated the Clicker or joystick, indicating a trend towards tool-free natural interaction in MR.

Yu et al. [27] assessed three text input methods: DwellType, TapType, and GestureType, finding that users tended to use GestureType and TapType and were less satisfied with DwellType. Dudley et al. [28] compared a physically attached keyboard to a hovering keyboard in virtual reality (VR) for input efficiency, revealing that physical alignment improved speed and accuracy, albeit limiting usage scenarios compared to the more flexible hovering keyboard [29]. Adhikary and Vertanen [11] found voice input superior to keyboard input in speed and accuracy. Ahn and Lee [12] developed an interaction method combining eye-tracking and touch, utilizing the speed advantage of eye-tracking to quickly select keyboard sections, followed by simple touch gestures to input characters. However, there has been no systematic study focused on the methods and effectiveness of numerical input within Mixed Reality environments.

### 2.3. Data Interaction in Immersive Analytics

Immersive analytics aims to ascertain whether new interfaces and display technologies can be used to foster deeper data analysis and exploration [30]. Previous studies have demonstrated that immersive environments enhance engagement and can intensify emotions. These factors contribute to making information more accessible [9].

Mota et al. [31] introduced a focus and context visualization technique for multi-geometry data, enabling user interaction with real-time data through VR controllers. Sicat et al. [6] developed DXR, an MR-based tool for data visualization and analysis, supporting selection, filtering, and more via bare-handed interaction. Cordeil et al. [10] created the IATK tool based on DXR, which offers enhanced data interaction, such as filtering attributes through sliders or input fields or encoding data attributes by adjusting visual channels like color, shape, and size. Büschel, Lehmann, and Dachselt [32] developed MIRIA, a visual analysis tool using AR headsets that allows analysts to visually explore spatial interaction data in context-rich environments, preserving context and enhancing data comprehensibility and comparability. The system supports multi-person collaboration, integrating gestures, eye-tracking, speech, and entity interactions with data.

While immersive analytics has seen notable achievements, numerous studies have identified the input of complex or precise data as a significant gap [9,32,33], hindering the visual analysis of crucial data in MR environments. A systematic assessment of interaction methods' efficiency across different data analysis tasks [14] or the development of data input methods for complex tasks could expand the range of information suited for immersive analytics on MR platforms [9].

MR advancements in natural interaction methods like gestures, speech, and eye-tracking have enhanced interaction variety and user experience flexibility; yet, research in the critical area of numerical input remains under-explored. Especially in immersive analytics, the challenge of inputting complex or precise data is notably significant, limiting MR's application potential and hindering user comprehension and analysis of complex data. Currently, an evident research gap still exists in effectively performing numerical input in MR, particularly in enhancing accuracy and efficiency through natural interaction.

## 3. Interaction Technique Design

This study introduces two multimodal data input techniques: Bimanual Scaling and Gesture and Voice, and compares them with three benchmark methods. These methods are used to input floating-point data with varying degrees of precision in a 3D environment, where the data panel is at different distances from the user.

### 3.1. Bimanual Scaling

The "Bimanual Scaling" interaction model aims to address accuracy issues stemming from gesture pointers or proximity interactions, employing multimodal interactions to enhance the interaction's overall fluidity. Building upon traditional gesture interactions, "Bi-

manual Scaling” incorporates eye-tracking selection in the “selection” phase, uses hand distance for data magnification during the “adjustment” phase, and the termination of a pinching gesture with both hands as the signal for “confirmation”. The user interaction process with “Bimanual Scaling” is illustrated in Figure 1. Interaction steps include the following: (1) The user looks at the numerical panel that needs to be operated. (2) The user pinches and moves forearms to alter the inter-hand distance. To reduce the value of the data, users decrease the distance between their hands, and to enlarge, they do the opposite. This gesture is similar to the HoloLens2 default zoom action, facilitating user understanding. (3) By maintaining the pinch gesture, users can achieve the continuous adjustment of values. Upon completion of adjustments, users can release their hands to signify the confirmation of the input result. Implementation-wise, reflecting user tendencies for broad initial adjustments followed by fine-tuning to approach the target value, Bimanual Scaling’s hand-distance magnification function is non-linear, accommodating coarse-to-fine input refinement. The function is as follows:

$$y = \begin{cases} (1 + C \cdot \text{SmoothStep}(0, 0.5, x')) \cdot x' & \text{if } x' \leq 0.5 \\ (1 + C \cdot \text{SmoothStep}(1, 0.5, x')) \cdot x' & \text{if } 0.5 < x' \leq 1 \\ x'^{0.5} & \text{if } x' > 1 \end{cases} \quad (1)$$

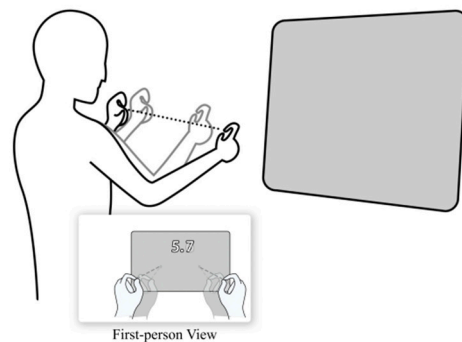


Figure 1. An illustration of the Bimanual Scaling technique.

In the above function,  $x'$  is defined as  $x' = \sqrt{\frac{x}{x_0 + 0.0001}}$ ;  $x$  is the current distance between two hands and  $x_0$  is the original distance between two hands. The constant value  $C$  was designed to adjust the sensitivity of this operation and, on our current device, the value  $C$  was set at 0.45. The values 0.5 and 1 were chosen as thresholds to create distinct scaling modes or sensitivity ranges within the interaction. For  $x' \leq 0.5$ , the system entered a “fine-tuning” mode where smaller hand movements resulted in more precise adjustments. For distances  $x' > 1$ , the system shifted to a “coarse adjustment” mode where hand movements lead to larger, more significant changes. The function part  $0.5 < x' \leq 1$  allowed for a seamless transition between precise and broad adjustments based on the scale of their hand movements, enhancing the interaction’s intuitiveness and efficiency.

The SmoothStep(a,b,t) function is defined as follows:

$$\text{SmoothStep}(a, b, t) = \begin{cases} 0 & \text{if } t \leq a \\ \frac{(t-a)^2 \cdot (3b-2t-a)}{(b-a)^3} & \text{if } a < t < b \\ 1 & \text{if } t \geq b \end{cases} \quad (2)$$

### 3.2. Gesture and Voice

Different interaction methods present distinct advantages and limitations. Eye tracking offers precision; yet, executing actions like confirmations solely through eye tracking poses challenges. The prolonged use of gesture-based interactions may induce fatigue. While voice input facilitates the input of complex content, discerning and isolating explicit

commands from lengthy sentences remains challenging. Hence, this study seeks to utilize the strengths while mitigating the weaknesses of each method by proposing a multimodal interaction technique named “Gesture and Voice”. This method employs eye tracking for preliminary input selection, pinch gestures for confirmations, and voice commands for entering floating-point data. It was anticipated that merging these interaction strengths would enhance the overall user experience.

The Gesture and Voice method is illustrated in Figure 2. Interaction steps include the following: (1) The user looks at the target panel to pre-select an item. (2) Confirmation is initiated by the user through a pinch gesture, triggering the voice input process. While maintaining the pinch gesture, the system records the user’s voice. Simultaneously, a visual prompt on the data panel indicates ongoing recording. (3) Upon releasing the pinch gesture, the recording stops, and the system converts the voice input into a floating-point number for display.

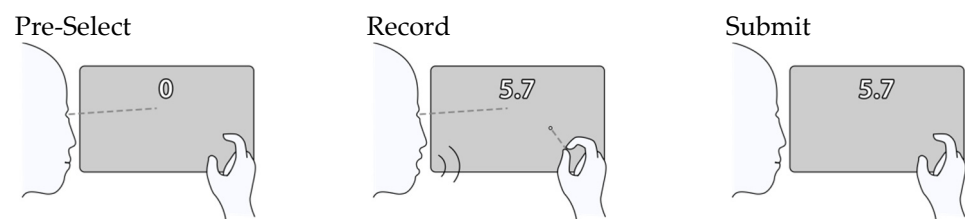


Figure 2. An illustration of the Gesture and Voice technique.

### 3.3. Benchmark Methods

The benchmark interaction methods, depicted in Figure 3, formed the basis for comparative analysis in this experiment. This experiment employed three standard numerical input methods from Mixed Reality Toolkit (MRTK) as benchmark methods for comparison with the two multimodal interaction methods investigated in this study. The three interaction methods are described as follows:

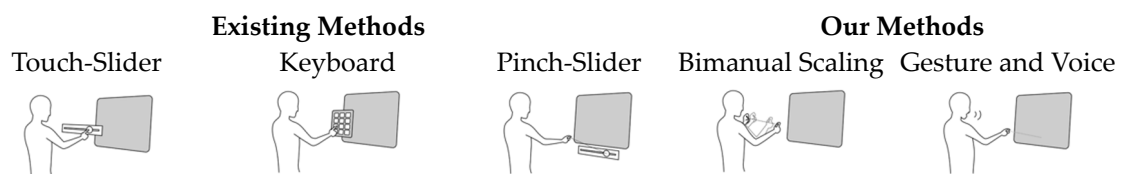


Figure 3. All techniques mentioned in this paper.

**Pinch-Slider:** This interaction technique capitalizes on the natural gesture of pinching. Users simulate a pinching gesture, which, in turn, activates a virtual gesture pointer. This pointer is then used to interact with a slider mechanism presented within the mixed reality space.

**Touch-Slider:** In contrast to the Pinch-Slider, the Touch-Slider method emphasizes direct tactile interaction. It allows users to manipulate numerical values by physically touching and dragging a virtual slider using their fingers. This method simulates the conventional slider control found in many graphical user interfaces but translates it into the MR environment.

**Keyboard:** This method involves using a virtual numeric keyboard, where values are inputted by directly touching the keys with a finger.

### 3.4. Comparison

Table 1 compares the properties of five interaction techniques: Touch-Slider, Keyboard, Pinch-Slider, Bimanual Scaling, and Gesture and Voice; Figure 3 shows the interaction process of all five interaction methods. A key advantage of the proposed interaction methods is their use of eye–hand coordination, enabling users to complete input tasks with-

out diverting their gaze from the data source. Contrary to traditional slider methods, these interactions offer flexible input ranges without fixed boundaries, allowing users to adjust both range and precision within a single task, potentially improving input efficiency.

**Table 1.** Summary of similarities and differences between the techniques we studied.

	<b>Touch-Slider</b>	<b>Keyboard</b>	<b>Pinch-Slider</b>	<b>Bimanual Scaling</b>	<b>Gesture and Voice</b>
Confirm trigger	Touch	Touch	Pinch	Pinch	Pinch
Data input method	Slide	Touch	Slide	Scale	Voice
Modalities	Hand	Hand	Hand	Gaze and hand	Gaze, hand, and voice
Interaction metaphor	Slider	Keyboard	Slider	Scaling	Voice input
Gesture type	Motion (slide) + symbolic (touch)	Symbolic (touch)	Motion (slide) + symbolic (pinch)	Motion (scale) + symbolic (pinch)	Symbolic (pinch)

In summary, the primary objectives of this research were as follows:

1. Introduce and evaluate the effectiveness of two novel multimodal data input techniques, “Bimanual Scaling” and “Gesture and Voice”, within an MR environment.
2. Compare these novel techniques against three benchmark methods (Pinch-Slider, Touch-Slider, and Keyboard) to assess their user efficiency in inputting floating-point data at varying distances.

By establishing these objectives, this study aims to contribute to the development of more intuitive and efficient interaction techniques for mixed reality environments, ultimately enhancing the user experience.

#### 4. Evaluation

##### 4.1. Participants

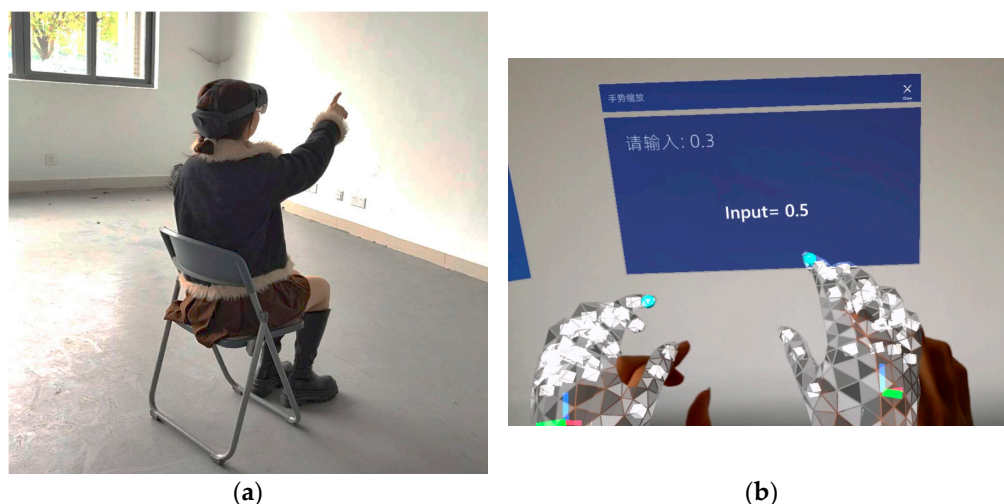
The research experiment involved 27 participants (14 females and 13 males) aged between 18 and 30. The majority (74.07%) reported that their daily work and study frequently entailed data reading, analysis, processing, or visualization. Participants rated their MR knowledge across three levels: “unfamiliar”, “somewhat familiar”, or “having a clear concept”, with over half of the participants (59.26%) reporting they were “somewhat familiar” with MR technology, a portion (37.04%) claiming a clear concept of MR, and one participant (3.7%) expressing unfamiliarity with MR. Approximately half (51.85%) had at least once used a dedicated VR, AR, or MR device, excluding cell phones and tablets. Among those, 40.74% had experienced devices featuring natural interaction modalities, including gesture and eye-tracking interactions.

##### 4.2. Hardware Devices and Experimental Setup

This study’s experimental system was developed with the Mixed Reality Toolkit (MRTK) in Unity for Microsoft HoloLens2 (field of view 43° × 29°), which supports hand and eye tracking (viewing angle accuracy of 1.5°) [13]. Example views of the experimental system in our study are shown in Figure 4.

Participants were seated in a spacious and quiet room (with a noise level smaller than 60 dB) for the duration of the experiment. The eye tracker was calibrated for each study participant at the start.

Since the experimental materials were initially in Chinese, we employed ChatGPT to correct and optimize the English translation during the preparation of this work [34]. This usage of ChatGPT ensured that the translation retained the technical accuracy and contextual relevance of the original materials [35].

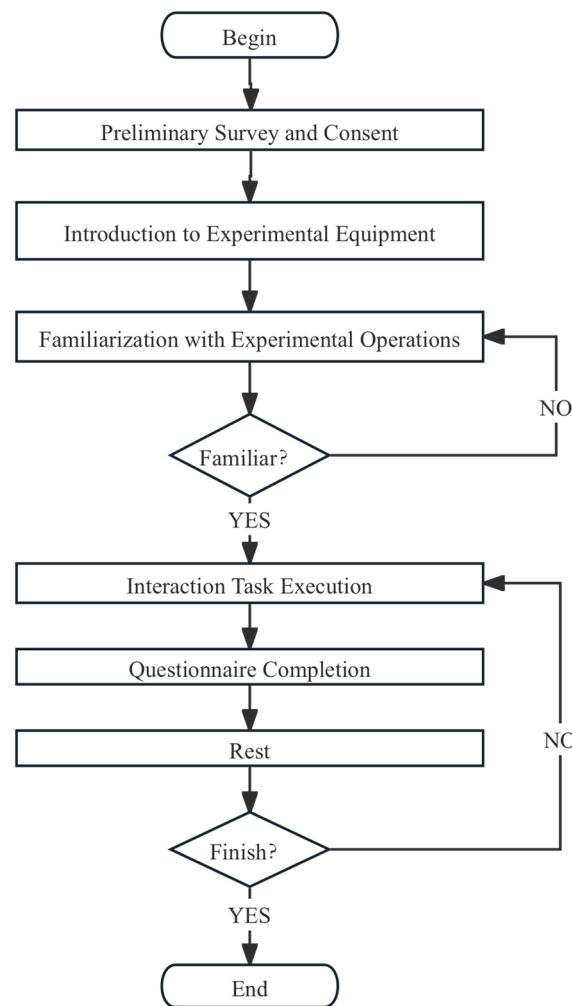


**Figure 4.** Examples of the experimental system in the user study from outside (a) and user (b) perspectives. The “请输入” in (b) means “Please enter”.

#### 4.3. Task and Procedure

The overall procedure of the experiment is shown in Figure 5. The experiment was divided into five sessions. Initially, participants signed a consent form, which confirmed their voluntary participation and permitted the use of their data for research purposes. They also completed a questionnaire that collected demographic details, including age, education, and field of study or work, and assessed their baseline knowledge of MR by querying their familiarity and previous experiences with using mixed reality technologies and devices featuring natural interaction capabilities. Following this, they watched a demonstration video to gain an initial understanding of the experimental process and operations. Subsequently, participants donned the HoloLens2 device and launched the instructional program to acquaint themselves with its basic usage and several fundamental interaction methods required for the experiment. Participants were allowed to run the instructional program multiple times until they found themselves to be familiar with these interaction techniques.

The experiment then officially commenced. The formal experiment featured five sections, each utilizing a distinct interaction method. During the experiment, participants were presented with a randomly generated number on a panel directly in front of them. Participants were tasked with accurately entering the data using the specified data input method. Random numbers comprised three types: a floating-point number between 0 and 1, a floating-point number between 1 and 10, and an integer between 1 and 10, addressing both precision levels (ten and one hundred) and data types (integer and floating-point). Interaction distances between the panel and the participant were set at 0.5 m, 1 m, and 2 m, varying to assess the impact of distance on interaction performance. Random number types and interaction panel distances were combined into  $3 \times 3 = 9$  scenarios, with each data input task repeated three times per scenario in a section, resulting in  $9 \times 3 = 27$  experiments per section. In total, the input task was conducted  $9 \times 3 \times 5 = 135$  times across the experiment. Upon completing all input tasks in a section, participants filled out a NASA-TLX questionnaire for that interaction. Participants had approximately 15 min rest periods between sections. Completing the entire experimental process took each participant approximately 40–60 min. The experiment yielded a total of  $27 \times 5 = 135$  NASA-TLX questionnaires and  $27 \times 135 = 3645$  input data points.



**Figure 5.** An illustration of the experiment process.

#### 4.4. Evaluation Metrics

In our study, we used two main metrics to evaluate the performance and subjective task load. Here are the detailed descriptions:

**Task Completion Time (TCT):** In order to accurately measure how long it took each participant to complete a given task, the experiment recorded the moment a target first appeared within the task scenario, as well as the instance when the participant successfully executed the required input action. This method ensured a comprehensive assessment of the participants' task completion times.

**NASA-TLX Questionnaire:** The experiment utilized the NASA-TLX (Task Load Index) questionnaire [13] to evaluate the subjective task load of participants following each interaction technique. It encompasses six questions on mental, physical, and temporal demands, as well as perceived success, effort, and frustration, rated on a 7-point Likert scale from very low to very high.

In summary, this section meticulously outlined the experimental design, participant demographics, task procedures, and evaluation metrics, aiming to rigorously investigate the effectiveness of various interaction methods within MR environments. This study sought to assess the impact of these technologies on user experience and task performance efficiency, ultimately contributing to the optimization and innovation of data input and interaction techniques in MR environments.

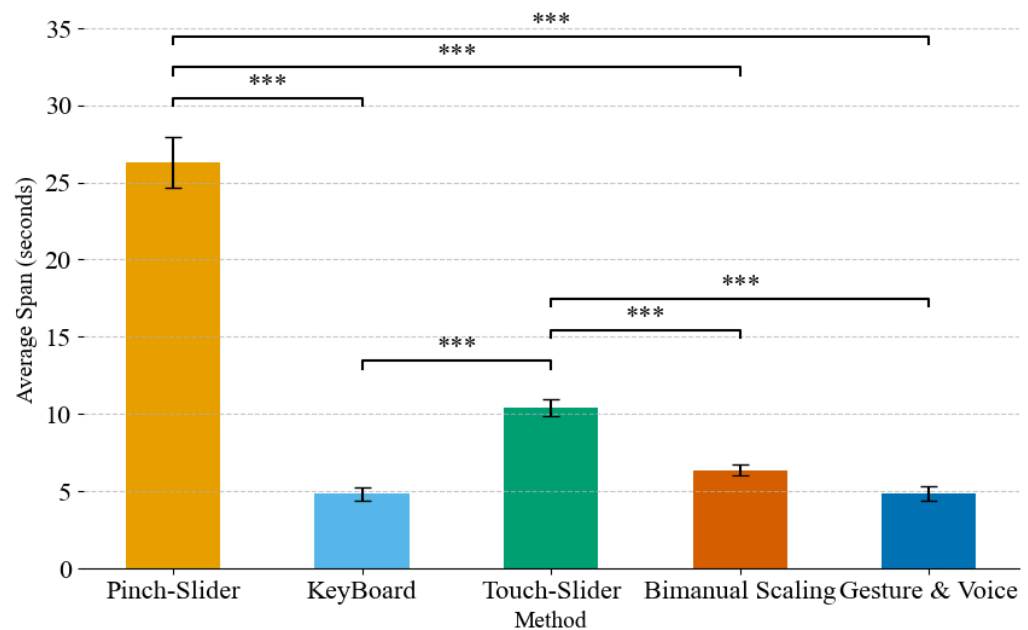


### 5. Result and Analysis

This study aimed to analyze the experimental data through quantitative and qualitative methods in order to assess the performance of different interaction techniques in terms of task completion time, cognitive load, and user satisfaction. A series of statistical methods were employed to analyze the performance of different interaction techniques in the above aspects. Initially, an ANOVA test was conducted on all subjective and objective quantitative outcomes. In cases of heteroscedasticity, the Welch test was utilized. Post-hoc pairwise comparisons were performed using the Bonferroni correction method to control the overall error rate of multiple comparisons. For data failing to meet the assumption of homogeneity of variances, the Games–Howell test was deployed to investigate specific differences among interaction methods.

#### 5.1. Task Completion Time

Figure 6 shows the average interaction time for each interaction method. In terms of task completion time, the data indicated significant differences between interaction methods ( $F_{585.26}^4 = 54.807, p < 0.001, \eta^2 = 0.276$ ). No significant differences were found in interaction time between the Touch-Slider and Pinch-Slider. However, both the Touch-Slider and Pinch-Slider methods showed significant differences in task completion time compared to other interaction methods ( $p < 0.001$ ), with both slider-based methods exhibiting significantly longer interaction times compared to the keyboard and the study’s proposed multimodal method, suggesting issues with the slider input method’s efficiency in data input and highlighting the superior efficiency of this study’s proposed methods.



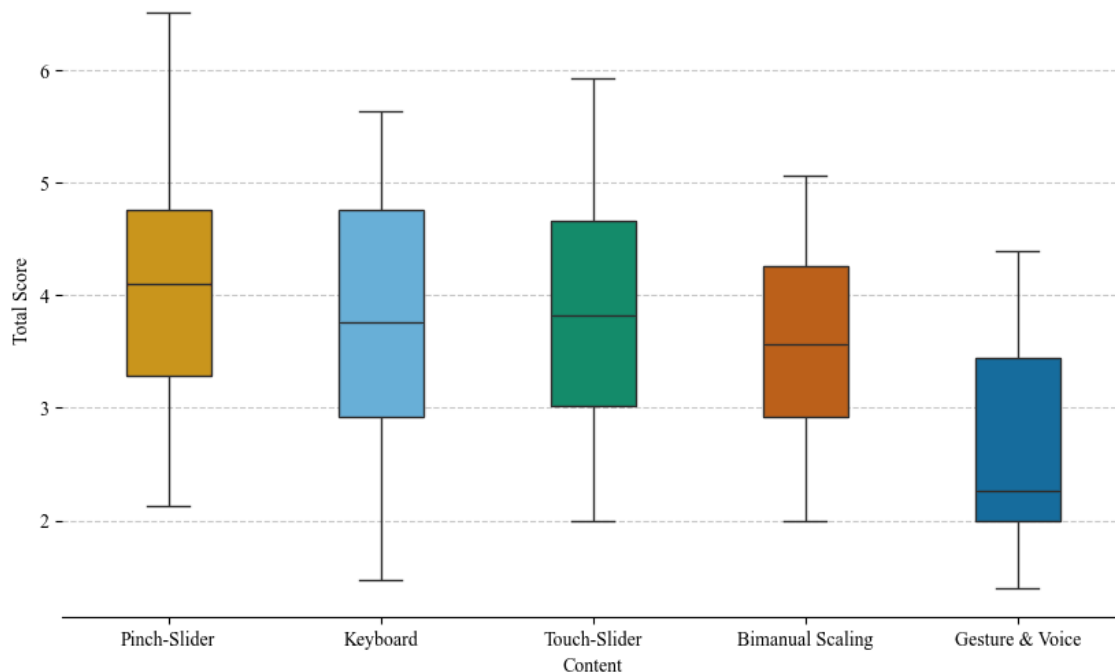
**Figure 6.** Average interaction time for each interaction method. Statistical significance is shown as \*\*\* for  $p < 0.001$ .

No significant differences in efficiency were observed among Bimanual Scaling, Gesture and Voice, and Keyboard; yet, significant differences existed compared to the remaining methods ( $p < 0.001$ ). This suggests that while this study’s proposed methods offer advantages over the slider method in terms of efficiency, they do not show a significant difference in efficiency compared to the numeric keyboard.

#### 5.2. NASA-TLX

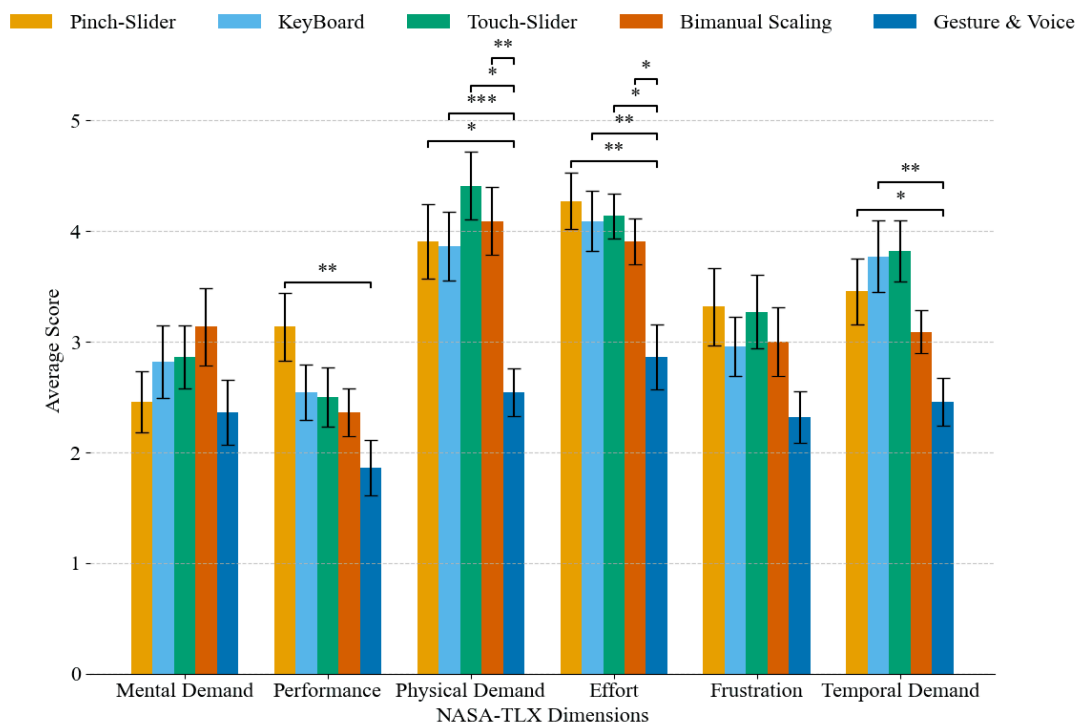
Figure 7 shows the mean workload of the five sessions. From the perspective of overall user evaluations, data analysis showed significant differences between each interaction

method ( $F_{52.254}^4 = 0.002$ ,  $p = 0.002$ ,  $\eta^2 = 0.183$ ). Post-hoc analysis showed that the cognitive load of Gesture and Voice was significantly lower than that of several other interaction methods. There were no significant differences in cognitive load between the remaining several interaction methods. This means that the Gesture and Voice interaction designed in this study was significantly better than the other interaction methods in terms of cognitive load, which is because the Gesture and Voice method simplifies the complexity of the input operation and avoids excessive hand movements. Although the Bimanual Scaling interaction method proposed in this study had some advantages in terms of interaction efficiency, it did not have significant advantages in terms of cognitive load compared with the current methods.



**Figure 7.** Mean workload of the five sessions.

Figure 8 shows the mean score of each NASA-TLX dimension. From the perspective of the six dimensions of NASA-TLX, no significant differences were observed in scores between groups for mental demand ( $F_{52.402}^4 = 0.998$ ,  $p = 0.417$ ,  $\eta^2 = 0.041$ ) and frustration ( $F_{52.173}^4 = 2.185$ ,  $p = 0.083$ ,  $\eta^2 = 0.067$ ). However, significant differences were observed in physical demand ( $F_{52.051}^4 = 8.391$ ,  $p < 0.001$ ,  $\eta^2 = 0.219$ ), temporal demand ( $F_{52.006}^4 = 5.079$ ,  $p = 0.002$ ,  $\eta^2 = 0.171$ ), performance ( $F_{52.325}^4 = 2.641$ ,  $p = 0.044$ ,  $\eta^2 = 0.117$ ), and effort ( $F_{52.197}^4 = 3.889$ ,  $p = 0.008$ ,  $\eta^2 = 0.199$ ) across different interaction methods. This indicates that neither mental demand nor frustration significantly contributes to variances in users' subjective perceptions. Firstly, this suggests that the tasks in this experimental design posed a moderate level of difficulty without imposing an unreasonable cognitive load on participants. Secondly, it demonstrates that regardless of the interaction method used, the outcomes of task completion align closely with users' psychological expectations.



**Figure 8.** Mean score of each NASA-TLX dimension. Statistical significance is shown as \* for  $p < 0.05$ , \*\* for  $p < 0.01$ , and \*\*\* for  $p < 0.001$ . Error bars indicate standard deviation.

In terms of physical demand, the Gesture and Voice group scored significantly better than the other groups ( $p < 0.01$ ), with no significant differences observed between the rest. This further demonstrates that interaction methods involving complex gestures can induce fatigue, lowering user ratings during data input tasks.

Regarding temporal demand, Gesture and Voice was significantly faster than the Pinch-Slider ( $p = 0.013$ ) and Keyboard input methods ( $p = 0.003$ ), with no significant differences among the other groups. This suggests that Gesture and Voice is quicker both subjectively and objectively for users. The Bimanual Scaling method proposed in this study was slightly better in terms of users' subjective perception of time spent compared to existing solutions (compared to Pinch-Slider  $p = 0.847$ , Keyboard  $p = 0.385$ , and Touch-Slider  $p = 0.223$ ), though not significantly. This may have been due to the fluidity of the continuous scaling operations causing users to underestimate the actual time consumed. Conversely, keyboard input, despite being objectively faster than both sliding input methods, tended to be perceived as more time-consuming due to its discontinuity and the additional cognitive load of searching for characters on the keyboard.

For performance, a significant difference was observed between the Gesture and Voice and Pinch-Slider methods proposed in this study, while differences among other groups were not significant. The Bimanual Scaling method, although slightly better than existing solutions (compared to Pinch-Slider  $p = 0.249$ , Keyboard  $p = 0.981$ , and Touch-Slider  $p = 0.995$ ), was still not significantly superior.

Regarding effort, voice interaction was significantly easier than all other interaction methods except for scaling interactions. No significant differences were observed among the other interaction methods. This indicates that users found voice interaction notably less strenuous. The Bimanual Scaling interaction method proposed in this study showed some superiority over current common interaction methods, though the effect was not markedly significant.

### 5.3. Discussion

This study evaluated the performance of different interaction techniques in terms of task completion time, cognitive load, and user satisfaction, yielding several key findings.

First, the experimental results indicate that in terms of interaction efficiency, the Touch-Slider and Pinch-Slider methods, based on slider components, are significantly inferior to Bimanual Scaling, which utilizes a two-handed scaling action. This finding suggests that Touch-Slider and Pinch-Slider, traditional slider input methods, necessitate fine touch or pinch movements to adjust the slider, thereby increasing the operation's physical complexity and cognitive load. Conversely, Bimanual Scaling, through the user's two hands' natural collaborative movements for zooming in and out of the data range, significantly simplifies the operation and reduces hand-eye coordination stress, thereby enhancing interaction efficiency and lessening the cognitive burden.

Furthermore, the intuitive interaction strategy of Bimanual Scaling diminishes the necessity for precise control, thereby enabling users to concentrate more on the task itself rather than on manipulating the interface. This enhances the interaction's naturalness and user satisfaction. Consequently, as an optimized method of interaction, Bimanual Scaling not only augments operational speed and efficiency but also improves the user experience.

Second, in terms of cognitive load, the experimental findings demonstrate that the Gesture and Voice interaction method significantly reduces cognitive load, underscoring its capability to streamline operations and lessen the burden, particularly for scenarios that require users to multitask.

From the perspective of interaction time, it was observed that both Bimanual Scaling and Keyboard interactions exhibited superior performance in input time. However, considering the subjective cognitive load, these methods did not demonstrate a significant advantage over the Touch-Slider and Pinch-Slider interactions. This may be attributed to the increased fatigue and cognitive stress from interacting with virtual objects and continuous adjustments of gestures.

The Gesture and Voice interaction method enhances interaction efficiency through its intuitive and natural approach, reflected in both reduced objective interaction time and subjective cognitive stress dimensions. Bimanual Scaling offers a cohesive gesture input method, comparable to Keyboard interaction regarding both objective interaction time and subjective cognitive stress, yet demonstrates greater input efficiency. Conversely, Touch-Slider and Pinch-Slider interactions fall short in both efficiency and range, revealing notable deficiencies in data input performance.

From the perspective of visual analytics applications, this study reveals the complexity of numerical input tasks in MR, emphasizing that interaction choices for data input in mixed reality interaction systems must be adjusted based on the usage environment. Among the five interaction methods discussed in this study, Gesture and Voice emerges as the method with the lowest cognitive stress and highest efficiency. However, in contexts like noisy environments or those requiring confidentiality, Gesture and Voice may not be optimal, underscoring the need for designers to account for environmental constraints and offer suitable alternatives [22]. Bimanual Scaling and Virtual Keyboard offer simple, rapid methods for brief, small-scale data input or when the Gesture and Voice method is impractical. These two methods demand less from the environment, especially Bimanual Scaling, which necessitates minimal hand-eye coordination and no extra space entities, offering a robust interaction mode for analyzing complex conditions like motion or confined spaces. However, designers need to pay particular attention to fatigue issues caused by continuous hand movements with these two interaction methods and avoid their use in long-duration, large-volume data entry tasks to prevent fatigue. Pinch-Slider and Touch-Slider, based on the slider component, offer tangible, continuous interaction methods; feedback from some participants suggests that these methods are more intuitive. The slider may be more intuitive for situations requiring the continuous input of indeterminate values. Nevertheless, this study's results indicate the disadvantages of the slider component in interaction

efficiency and cognitive load for precise inputs, suggesting the need to avoid these two methods in visual analytics applications requiring precision.

## 6. Conclusions and Limitations

This study focused on the innovation of interaction methods, user experiments, and quantitative analysis for numerical data input within immersive visual analytics based on MR. Initially, it introduced two multimodal interaction methods: Gesture and Voice and Bimanual Scaling. Subsequently, it conducted an experiment comparing these proposed interaction methods with three established MR input methods. The experiment revealed that the Gesture and Voice method significantly outperforms current industry-standard input methods in both objective and subjective metrics, whereas Bimanual Scaling shows advantages in objective metrics and certain subjective metrics, also demonstrating resilience to environmental noise and greater generalizability. Furthermore, it identified that Pinch-Slider and Touch-Slider interactions necessitate increased hand–eye coordination and finer manipulation, leading to significantly greater cognitive stress in delicate numerical input tasks. The experiment’s findings contribute uniquely to overcoming the obstacles associated with inputting precise or complex data in immersive environments, offering guidance for the design of more complex future immersive visual analytics systems. It is anticipated that these interaction techniques will find broader applications, offering users a more natural, efficient, and enjoyable interaction experience.

Nevertheless, this study has several limitations. Firstly, the focus was primarily on numerical data input rather than a comprehensive examination of all potential interaction methods in MR environments. For other data types (such as text and graphics), the effectiveness and applicability of the proposed interaction techniques might differ. Hence, the experiment’s findings may not be universally applicable to all data input tasks. Secondly, technology implementation limitations could have influenced the experimental outcomes. For instance, speech input accuracy may be contingent on speech recognition algorithms, while gesture and eye-tracking interaction accuracy could depend on hardware performance. Such technological constraints may impinge on user experience and interaction efficiency. Moreover, interactions were tailored to specific hardware (e.g., HoloLens2) and software (e.g., Unity), potentially restricting their applicability across various platforms and devices. Factors such as lighting, noise, and varied hardware and software environments can significantly impact the performance and user experience of the interaction methods. Additionally, while user experience was evaluated through the NASA-TLX questionnaire, this study did not include an evaluation of long-term usage. User adaptations and preferences for interaction styles are likely to evolve over time, especially after prolonged use. Thus, longitudinal user studies could yield more profound insights. Furthermore, this study’s limitations include a small sample size of 27 and the narrow age range of participants under 30. Future research should aim to increase the sample size and include a broader age demographic to enhance the findings’ generalizability. Lastly, the experiment’s controlled conditions may not have fully reflected the real-world applications of MR technologies.

**Author Contributions:** Conceptualization, J.C. and J.Z.; methodology, J.Z. and J.C.; software, J.Z. and T.C.; validation, J.C. and W.G.; investigation, J.L.; resources, J.Z. and J.C.; data curation, T.C. and W.G.; writing—original draft preparation, J.Z. and J.L.; writing—review and editing, J.C. and J.Z.; visualization, J.Z. and T.C.; supervision, J.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National College Student Innovation and Entrepreneurship Training Program no. 202310295002Z.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to sensitive nature.

**Acknowledgments:** During the preparation of this work, the authors used ChatGPT in order to check grammar and enhance the English language used. After using this tool, the authors reviewed and edited the content as needed and we take full responsibility for the content of the publication. We also thank the anonymous reviewers who provided valuable comments on the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Rokhsaritalemi, S.; Sadeghi-Niaraki, A.; Choi, S.-M. A Review on Mixed Reality: Current Trends, Challenges and Prospects. *Appl. Sci.* **2020**, *10*, 636. [[CrossRef](#)]
- Flavián, C.; Ibáñez-Sánchez, S.; Orús, C. The Impact of Virtual, Augmented and Mixed Reality Technologies on the Customer Experience. *J. Bus. Res.* **2019**, *100*, 547–560. [[CrossRef](#)]
- Jiang, H. Mobile Fire Evacuation System for Large Public Buildings Based on Artificial Intelligence and IoT. *IEEE Access* **2019**, *7*, 64101–64109. [[CrossRef](#)]
- Walters, S.M.; Hirsch, S.E.; McKown, G.; Carlson, A.; Allen, A.A. Mixed-Reality Simulation with Preservice Teacher Candidates: A Conceptual Replication. *Teach. Educ. Spec. Educ.* **2021**, *44*, 340–355. [[CrossRef](#)]
- Papadopoulos, T.; Evangelidis, K.; Kaskalis, T.H.; Evangelidis, G.; Sylaiou, S. Interactions in Augmented and Mixed Reality: An Overview. *Appl. Sci.* **2021**, *11*, 8752. [[CrossRef](#)]
- Sicat, R.; Li, J.; Choi, J.; Cordeil, M.; Jeong, W.-K.; Bach, B.; Pfister, H. DXR: A Toolkit for Building Immersive Data Visualizations. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 715–725. [[CrossRef](#)] [[PubMed](#)]
- Zhao, Y.; Jiang, J.; Chen, Y.; Liu, R.; Yang, Y.; Xue, X.; Chen, S. Metaverse: Perspectives from Graphics, Interactions and Visualization. *Vis. Inform.* **2022**, *6*, 56–67. [[CrossRef](#)]
- Filho, J.A.W.; Stuerzlinger, W.; Nedel, L. Evaluating an Immersive Space-Time Cube Geovisualization for Intuitive Trajectory Data Exploration. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 514–524. [[CrossRef](#)] [[PubMed](#)]
- Kraus, M.; Fuchs, J.; Sommer, B.; Klein, K.; Engelke, U.; Keim, D.; Schreiber, F. Immersive Analytics with Abstract 3D Visualizations: A Survey. *Comput. Graph. Forum* **2022**, *41*, 201–229. [[CrossRef](#)]
- Cordeil, M.; Cunningham, A.; Bach, B.; Hurter, C.; Thomas, B.H.; Marriott, K.; Dwyer, T. IATK: An Immersive Analytics Toolkit. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 200–209.
- Adhikary, J.; Vertanen, K. Text Entry in Virtual Environments Using Speech and a Midair Keyboard. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 2648–2658. [[CrossRef](#)] [[PubMed](#)]
- Ahn, S.; Lee, G. Gaze-Assisted Typing for Smart Glasses. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, New Orleans, LA, USA, 20–23 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 857–869.
- Wagner, U.; Lystbæk, M.N.; Manakhov, P.; Grønbaek, J.E.S.; Pfeuffer, K.; Gellersen, H. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; pp. 1–15.
- Ens, B.; Bach, B.; Cordeil, M.; Engelke, U.; Serrano, M.; Willett, W.; Prouzeau, A.; Anthes, C.; Büschel, W.; Dunne, C.; et al. Grand Challenges in Immersive Analytics. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–17.
- Speicher, M.; Hall, B.D.; Nebeling, M. What Is Mixed Reality? In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, UK, 4–9 May 2019; pp. 1–15.
- Kang, H.J.; Shin, J.; Ponto, K. A Comparative Analysis of 3D User Interaction: How to Move Virtual Objects in Mixed Reality. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22–26 March 2020; pp. 275–284.
- Kerdvibulvech, C. A Review of Augmented Reality-Based Human-Computer Interaction Applications of Gesture-Based Interaction. In *HCI International 2019—Late Breaking Papers*; Stephanidis, C., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 233–242.
- Newbury, R.; Satriadi, K.A.; Bolton, J.; Liu, J.; Cordeil, M.; Prouzeau, A.; Jenny, B. Embodied Gesture Interaction for Immersive Maps. *Cartogr. Geogr. Inf. Sci.* **2021**, *48*, 417–431. [[CrossRef](#)]
- Sidenmark, L.; Clarke, C.; Zhang, X.; Phu, J.; Gellersen, H. Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–13.
- Shi, R.; Zhang, J.; Yue, Y.; Yu, L.; Liang, H.-N. Exploration of Bare-Hand Mid-Air Pointing Selection Techniques for Dense Virtual Reality Environments. In Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; pp. 1–7.
- Nizam, S.M.; Abidin, R.Z.; Hashim, N.C.; Lam, M.C.; Arshad, H.; Majid, N. A Review of Multimodal Interaction Technique in Augmented Reality Environment. *Int. J. Adv. Sci. Eng. Inf. Technol* **2018**, *8*, 1460. [[CrossRef](#)]

22. Hanifa, R.M.; Isa, K.; Mohamad, S. A Review on Speaker Recognition: Technology and Challenges. *Comput. Electr. Eng.* **2021**, *90*, 107005. [[CrossRef](#)]
23. Li, J. Recent Advances in End-to-End Automatic Speech Recognition. *APSIPA Trans. Signal Inf. Process.* **2022**, *11*, e8. [[CrossRef](#)]
24. Plopski, A.; Hirzle, T.; Norouzi, N.; Qian, L.; Bruder, G.; Langlotz, T. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality. *ACM Comput. Surv.* **2022**, *55*, 1–39. [[CrossRef](#)]
25. Pfeuffer, K.; Mayer, B.; Mardanbegi, D.; Gellersen, H. Gaze + Pinch Interaction in Virtual Reality. In Proceedings of the 5th Symposium on Spatial User Interaction, Brighton, UK, 16–17 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 99–108.
26. Derby, J.L.; Rarick, C.T.; Chaparro, B.S. Text Input Performance with a Mixed Reality Head-Mounted Display (HMD). In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Seattle, WA, USA, 28 October–1 November 2019; Volume 63, pp. 1476–1480. [[CrossRef](#)]
27. Yu, C.; Gu, Y.; Yang, Z.; Yi, X.; Luo, H.; Shi, Y. Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for Hmds. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 4479–4488.
28. Dudley, J.; Benko, H.; Wigdor, D.; Kristensson, P.O. Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. In Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, 14–18 October 2019; pp. 289–300.
29. Biener, V.; Ofek, E.; Pahud, M.; Kristensson, P.O.; Grubert, J. Extended Reality for Knowledge Work in Everyday Environments. In *Everyday Virtual and Augmented Reality*; Simeone, A., Weyers, B., Bialkova, S., Lindeman, R.W., Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 21–56. ISBN 978-3-031-05804-2.
30. Zhang, Y.; Wang, Z.; Zhang, J.; Shan, G.; Tian, D. A Survey of Immersive Visualization: Focus on Perception and Interaction. *Vis. Inform.* **2023**, *7*, 22–35. [[CrossRef](#)]
31. Mota, R.C.R.; Rocha, A.; Silva, J.D.; Alim, U.; Sharlin, E. 3De Interactive Lenses for Visualization in Virtual Environments. In Proceedings of the 2018 IEEE Scientific Visualization Conference (SciVis), Berlin, Germany, 21–26 October 2018; pp. 21–25.
32. Büschel, W.; Lehmann, A.; Dachselt, R. MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; Association for Computing Machinery: New York, NY, USA, 2021.
33. Reski, N.; Alissandrakis, A. Open Data Exploration in Virtual Reality: A Comparative Study of Input Technology. *Virtual Real.* **2020**, *24*, 1–22. [[CrossRef](#)]
34. Lee, T.K. Artificial Intelligence and Posthumanist Translation: ChatGPT versus the Translator. Available online: <https://www.degruyter.com/document/doi/10.1515/applirev-2023-0122/html> (accessed on 11 March 2024).
35. Rice, S.; Crouse, S.R.; Winter, S.R.; Rice, C. The Advantages and Limitations of Using ChatGPT to Enhance Technological Research. *Technol. Soc.* **2024**, *76*, 102426. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.