Title	バーチャルリアリティ作業空間の展開：設計基準の策定と低ジッタータイピングシステムによる効率性の向上
Author(s)	XU, TIANSHU
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19927
Rights	
Description	supervisor: 長谷川 忍, 先端科学技術研究科, 博士

Doctoral Dissertation

Advancements in Virtual Reality Workspaces: Enhancing Efficiency Through Defined Design Standards and Low-Jitter Typing System

Tianshu Xu

Supervisor: Shinobu Hasegawa

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

March 2025

Abstract

Open-Plan Workspace (OPWS) is an office style that allows many employees to work simultaneously in a wall-less, partition-less environment. OPWS is characterized by a high sense of openness, low cost, promotion of cooperation, and enhancement of the collective wisdom of the team. Although OPWS has already proven its value, many shortcomings still exist. OPWS is not only full of auditory and visual interference, but the low level of privacy protection also causes psychological stress to employees and reduces work efficiency. Although many researchers have been working on solving these problems, they still cannot declare that the issues have been entirely resolved. One prominent limitation is that most proposals suggest creating additional workspaces, which incur extra costs. Virtual Reality Workspace (VRWS), a virtual personal space independent of OPWS, can alleviate the psychological pressure caused by the lack of privacy protection in public office environments. It can also reduce auditory and visual interference in the workspace. Thus, we hypothesize that VRWS technology has great potential to address the challenges in OPWS. However, no studies have shown how Virtual Reality (VR) environments can be designed to maintain or improve work efficiency. In traditional workspaces, some studies suggest that the office environment has a significant influence on work efficiency. Therefore, we believe that a properly designed VRWS can improve users' work efficiency.

Previous research has shown that a pleasant office environment should be a cozy space free from visual and auditory interference, with good lighting, controlled sound, and plenty of natural light. Although some studies have proposed solutions to improve the shortcomings of OPWS, it remains unclear whether these solutions can be applied to VRWS. This research aims to create a VRWS with the favorable characteristics of OPWS to maintain or improve work efficiency.

We adopted Semantic Differential (SD) analysis to compare the emotional responses of participants in both OPWS and VRWS, identifying differences between the two environments and exploring factors unique to VRWS that influence work efficiency.

Due to the difficulty of finding a typical noisy OPWS in the author’s region, we decided to use the Cave Automatic Virtual Environment (CAVE) system to simulate a typical noisy OPWS. The CAVE system is a projection-based virtual reality system, which consists of several projection screens surrounding the participants and can produce a completely immersive virtual environment. At the same time, mini speakers were arranged around the CAVE system to restore the simulated OPWS sound environment as much as possible. For the simulated OPWS content in the CAVE system, we selected

NASA's mission center, where one of the frequent activities involves information exchange among employees.

We hypothesized that a VRWS with excellent OPWS characteristics—namely, an environment free from visual and auditory interference, with good lighting, sufficient natural light, and privacy protection—would maintain or improve work efficiency. To meet these requirements, we implemented the following measures: A combination of Head Mounted Display (HMD) and noise-canceling headphones was used to eliminate visual and auditory interference. To create a pleasant lighting environment, we increased the brightness of the virtual model and used natural light sources instead of ordinary light sources, ensuring the entire virtual space was well-lit. To provide ample natural light, we designed large floor-to-ceiling windows to replace the walls on either side of the VRWS and positioned the virtual desk near the windows, allowing users to enjoy the scenic views outside. For privacy protection, we designed VRWS to be single use, ensuring users experience a personal office environment.

Despite participants' overall satisfaction with working in VR, some expressed concerns regarding efficiency. A significant factor was the perceived challenge of prolonged and consistent touch typing while wearing an HMD. Consequently, participants indicated the need to frequently remove and wear the HMD if needed to type something, which negatively impacted work efficiency and user enthusiasm.

An excellent VRWS should not only provide a well-designed virtual office environment but also feature robust typing assistance functionality. To address the challenges of typing in VR, we summarized the drawbacks of not supporting a physical keyboard and relying on additional auxiliary devices. To comprehensively solve these issues, we decided to use only the HMD-mounted camera to capture typing actions and reproduce them in real time in VR. This approach eliminates the need for auxiliary devices while allowing users to verify their typing hand positions in VR. However, typing in VR presents unique challenges. When using the HMD's camera to capture typing hand positions, the fingers are often obscured by the palm, making it difficult to obtain a complete hand contour. Therefore, a dataset of "obscured typing hands" was required to train the hand-tracking model.

We conducted a data collection experiment to create this dataset. Each participant engaged in a one-hour typing session, collecting a total of 21,900 images. Using OpenCV for data augmentation, we expanded the dataset to approximately 438,000 images, 20 times larger than the original dataset. Manual annotation was performed by human annotators with the assistance of MediaPipe to extract meaningful images of typing hands.

Additionally, Motion History Image (MHI), a computer vision technique for capturing temporal motion patterns, was applied to the data to extract motion-related features.

To construct the neural network model, we prioritized minimizing VR latency. We employed a 2-stream (2S) ResNet18 to process typing images and MHI data, followed by LSTM for further processing. The model outputs the coordinates of 42 key points, which are transmitted in real time to the VR controller, enabling accurate reconstruction of typing hand positions in VR. Kalman Filtering (KF) was applied to reduce jitter. An ablation study was conducted to evaluate the model's effectiveness.

To identify the optimal network framework for VR typing tasks, we compared several models, including some recently proposed ones, focusing on metrics such as latency, accuracy, and jitter. Our analysis confirmed that the 2S-LSTM model performed among the best. To further evaluate the typing assistance solution, we conducted comparative experiments. Participants performed typing tasks under normal typing conditions, using the 2S-LSTM model, and with two other VR typing solutions, followed by a questionnaire. Statistical analysis of the experimental and questionnaire results demonstrated that the 2S-LSTM solution effectively maintained typing efficiency.

Finally, we hypothesized that the better the typing support system, the smaller the impact on users (with normal typing having no impact). To test this hypothesis, we analyzed participants' typing behavior under different conditions. Results showed that the 2S-LSTM solution had the least impact on users.

This research addresses two major challenges in virtual office systems: designing an effective VRWS and overcoming text input difficulties in VR, laying a solid foundation for the future development of VRWS systems. By referencing the findings and conclusions of this study, future researchers and developers can build upon this work to create improved VR office assistance solutions for a broader range of users.

In terms of VRWS design, this research establishes a foundational comprehensive design standard by integrating principles from OPWS studies to create an environment with minimal visual and auditory interference, enhanced lighting, and strong privacy protection. These features contribute to a more comfortable and efficient VR office experience, supported by empirical evidence from user evaluations.

For typing in VR, a key novelty of this research lies in the integration of the Two-Stream (2S) architecture and the Kalman Filter (KF) to develop a low-jitter hand tracking system. This innovative combination enables the system to accurately track and replicate typing hand positions in real time, addressing a critical challenge in VRWS. The proposed 2S-LSTM-KF solution not only improves typing efficiency but also reduces latency and jitter, maintaining users' natural typing habits and enhancing overall productivity.

Through comprehensive evaluations comparing this model with other recent solutions, including state-of-the-art models, the 2S-LSTM-KF system demonstrated the best overall performance across latency, accuracy, jitter, and deployment feasibility.

Moreover, this study opens new possibilities for applying VR technology in fields such as remote work, education, and virtual collaboration. By establishing a design standard for VRWS and addressing usability challenges in VR text entry with a robust and scalable solution, this research paves the way for the widespread adoption of VRWS. The findings offer valuable insights for researchers and developers aiming to design more efficient and user-friendly virtual environments, contributing to the theoretical and practical advancement of VR technology in virtual office settings.

Keywords: virtual reality; workspace; work efficiency; low jitter; typing support; hand tracking.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Shinobu Hasegawa, for his unwavering support and guidance throughout my academic journey. From my undergraduate internship to my master's program and finally my doctoral studies, his mentorship has been invaluable. His insights and encouragement have shaped not only this dissertation but also my growth as a researcher and individual.

I am deeply thankful to my parents, whose endless support and belief in my potential have been my foundation throughout this long academic pursuit. Their love and sacrifices have provided me with the strength to persevere.

To my fiancée, Liu, thank you for your constant companionship and patience. Your presence has been my greatest source of comfort and motivation, especially during the challenging moments of this journey. When she read the first draft of my acknowledgments, she smiled and told me that she hoped I would write a little more about her. But what should I write? I thought about it for a while, but I still wasn't sure. Perhaps the simplest truth is this: we met in 2017, and eight years have passed since then. Through all these years, she has been by my side, ensuring that loneliness was something I never had to experience. For that, I am grateful beyond words.

I also wish to extend my heartfelt thanks to Dr. Sagorika Safinoor, whose guidance significantly improved my English communication skills. When she first came to my university as a graduate student, aspiring to enter the doctoral program, I became her "daily life assistant." At that time, my English was poor, and she spoke only English, so I had to speak with her while translating everything into Japanese. This experience unknowingly shaped my English skills, and if I can communicate fluently in English today, it is because of those moments. We became good friends, despite the age gap—sometimes she treated me like a younger brother, sometimes, perhaps, even like her own child. However, not long after she graduated, she was diagnosed with leukemia, and within a year, she was gone. I regret not being able to do more. Through this, I came to understand what it truly means to lose someone.

To everyone who has contributed to my academic and personal growth, directly or indirectly, thank you. This milestone would not have been possible without you.

Looking back, this journey has been incredibly long. If I start counting from when I

first entered university in 2012, it has been 14 years. But it stretches even further back—to elementary school, middle school, and high school. If I consider the years from when I first became aware of myself, my entire life up to this point has belonged to the era of being a student. And now, that chapter is coming to a close.

I have never been an exceptional student. In fact, I wouldn't even call myself someone who enjoys studying. Yet, somehow, I never stopped. From elementary school to middle school, from high school to university, from my master's to this doctoral program—I kept moving forward, without ever truly understanding why. And yet, here I am.

As I write these words, I am exactly where I once imagined I would be—sitting by the window, with the evening sun casting a warm glow, feeling the breeze drift into the room. A can of beer in my hand, the distant laughter of children playing at a nearby elementary school, and the occasional chirping of birds outside. This is how I choose to close the final chapter of my student life.

To whoever is reading these words, whether it be years from now or decades later—I don't know what I should say to you. Perhaps I don't need to say anything at all. Instead, I will simply do this: I raise my beer, on March 14, 2025, at 5:58 PM, to the sunset outside my window.

Let's have a toast—just you and me.

Cheers.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

In the wake of the COVID-19 pandemic, remote education and work have become integral aspects of our daily lives. This paradigm shift was driven by the need to maintain social distance and minimize physical interactions. Remote education and work offer significant advantages, including flexibility, accessibility, and the ability to maintain productivity in challenging circumstances [1, 2]. The technologies facilitating remote education and work include video conferencing platforms, collaborative document editing tools, and cloud-based storage solutions [3].

Amidst existing remote solutions, Virtual Reality (VR) emerges as a noteworthy innovation. By leveraging computer-generated simulations, VR provides users with an immersive experience that simulates real-world scenarios. This technology is characterized by the use of specialized hardware, such as head-mounted displays (HMDs) and motion-sensing controllers, enabling users to engage with computer-generated environments in a seemingly natural way [4]. The integration of VR into remote education and work has the potential to mitigate many of the challenges brought about by the pandemic, such as the lack of engagement and limited interactive capabilities in conventional remote platforms [5, 6].

VR technology is a groundbreaking field based on computer graphics, enabling the creation of virtual scenes and interactive elements. Users manipulate these elements through input devices, experiencing multisensory immersion seeing, hearing, touching, and even smelling within the virtual environment [7,8,9]. VR provides a level of immersion that transcends traditional digital interfaces, and users can feel almost as if they were in the real world [7]. Burdea and Coiffet introduced the concept of the 3I of virtual reality immersion, imagination, and interaction, which forms the foundation for a rich user experience [8].

In essence, VR technology represents the next generation of media for obtaining artificial information, offering a completely immersive and intuitive interactive experience. Burdea et al. defined the five classical components of a VR system—VR engine, software & databases, input / output devices, user, and tasks [8]. These components form the foundation for a VR system. Common input devices in VR systems

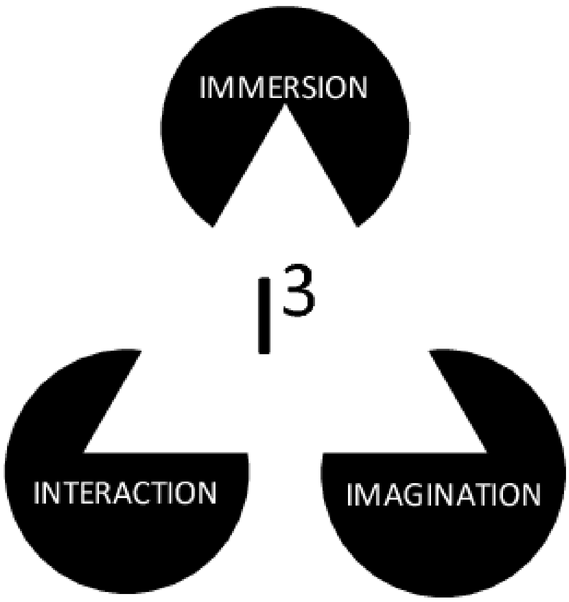


Figure 1.1: 3I of virtual reality [8].

are classified into two categories: manually operated and automatic capturing. Manual operation devices, such as keyboards and mice, are mature but may have a steep learning curve.

On the other hand, automatic tracking devices offer a higher degree of intuitiveness but may sacrifice accuracy due to the use of various sensors and algorithms [8,9,10]. VR systems require output devices to allow users to sense the virtual world. Visual sense is the most crucial, constituting 70% of the total sensing for a human being [7,11]. Common visual output devices include screens, projectors, HMDs, and holographic devices [12,13,14]. Among these, HMDs stand out as the most appropriate visual devices for VR training systems, offering full immersion at a reasonable cost.

We summarize the advantages of remote work in VR as follows:

- **Customizable Workspaces: [7,8,9,15]**

Users can personalize their virtual workspace, optimizing it for individual preferences and task requirements, potentially boosting focus and productivity.

- **Reduced Commuting Stress: [16,17]**

Eliminating the need for physical commuting contributes to reduced stress and fatigue, potentially positively impacting overall work efficiency.

- **Spatial Presence: [7,8,9,10,11,14]**

VR offers an unparalleled sense of spatial presence, allowing users to feel physically present in a meticulously crafted virtual office space. This spatial immersion fosters a deeper level of engagement and collaboration.

- **Global Collaboration: [18,19]**

Overcoming geographical constraints, VR enables seamless global collaboration. Colleagues from different corners of the world can converge in a shared virtual space in real-time, enriching teamwork with diverse perspectives and expertise.

- **Immersive Meetings: [20]**

VR transcends traditional video calls by offering immersive meeting experiences. Participants can engage with 3D models, interactive presentations, and collaborative tools in a shared virtual environment, enhancing the quality and depth of discussions.

- **Immersive Focus: [21,22]**

VR promotes a high level of concentration by minimizing external distractions. This immersive focus can lead to increased productivity, especially for tasks that require deep concentration and attention to detail.

- **Creative Imagination**: [23]

The imaginative aspect of VR allows users to visualize and conceptualize ideas in three-dimensional space. This creative environment fosters innovation, enabling users to explore and implement novel solutions to work challenges.

- **Enhanced Interaction: [7,16,17,18,19,20]**

  VR facilitates natural and intuitive interaction, whether it's manipulating virtual objects or collaborating with team members in a shared virtual space. This can lead to more effective communication and seamless collaboration.

Despite the above advantages, it still cannot be concluded that VR can effectively assist people in remote education / work. A notable concern is that the public commonly perceives VR as an entertainment tool, leading to skepticism about the potential negative impacts of virtual environments in remote education / work settings, highly customizable nature of VR further exacerbates these apprehensions. Regarding the perspective on efficiency and the design of virtual environments, it is crucial to delve deeper into these aspects, as they are among the issues addressed in this study.

Another issue is mainly restrictions by HMDs. [23] One notable limitation is the obstruction of vision while wearing HMDs, which can impede users' ability to see their physical surroundings. Typing and interacting with a physical keyboard can be cumbersome, potentially affecting work efficiency. [24, 25] As such, addressing these challenges is crucial for optimizing the effectiveness of VR into remote education / work.

Additionally, one critical aspect that warrants attention in the realm of VR is the phenomenon known as jitter. Jitter refers to the subtle, unintended movements or shaking that occur in the virtual environment, often caused by complex technical factors. [23] Understanding jitter is pivotal, as it not only has implications for user comfort, leading to symptoms of motion sickness and dizziness, but also significantly affects the precision of user interactions within the VR space. [26] Jitter arises from a variety of sources, including tracking inaccuracies, latency in motion sensors, and limitations in the rendering capabilities of VR hardware. These factors collectively contribute to the subtle but impactful instability experienced by users, particularly when performing tasks that require fine motor skills, such as typing. The consequences of jitter on typing efficiency in VR are notable. Users may find it challenging to maintain a consistent typing speed and accuracy due to the unpredictable movements induced by jitter. This can lead to frustration and a decline in overall productivity, as the user struggles to synchronize their physical movements with the virtual representation of a keyboard. Moreover, the discomfort induced by jitter can exacerbate issues such as motion sickness, further hindering the user's ability to work efficiently in the VR environment. Addressing jitter in VR becomes imperative for ensuring a comfortable and efficient user experience, especially in applications where precise interactions, such as typing, are integral.

In this current study, we aim to develop a VR system that maintains user learning / work efficiency by addressing the previously mentioned issues. This will be explained in detail in the later section of this chapter.

1.2 Challenges in Text Input for VR Office Applications

In the post-pandemic era, remote work and online education have become integral aspects of daily life. Against this backdrop, VR technology has garnered significant attention as an innovative tool to enhance these activities. However, one major barrier prevents VR from reaching its full potential in professional settings—the lack of efficient text input methods.

A key challenge arises when typing while wearing HMD. As shown in Figure 1.2, for most users who cannot touch type, it is often necessary to remove the HMD to view the keyboard and their hands, type the required text, and then put the HMD back on to continue interacting with the VR environment. This repetitive process disrupts workflows, reduces productivity, and significantly lowers the appeal of VR for professional tasks.

Additionally, current VR typing solutions primarily rely on handheld controllers to select individual letters on a virtual keyboard, like Figure 1.3. While this approach facilitates text input to some extent, it is far too inefficient to meet the demands of office work, which requires fast and seamless typing. These shortcomings present a major obstacle to VR adoption in professional environments, where usability and efficiency are paramount.

Overcoming these challenges will require the development of innovative and intuitive text input methods designed specifically for VR environments. Without such advancements, VR's potential as a transformative tool for remote work and collaboration will remain severely constrained. Addressing these barriers would enable VR to truly revolutionize remote professional workflows and establish itself as an essential technology in modern work environments.

Figure 1.2: The obstruction of vision can reduce typing efficiency in VR. [27]

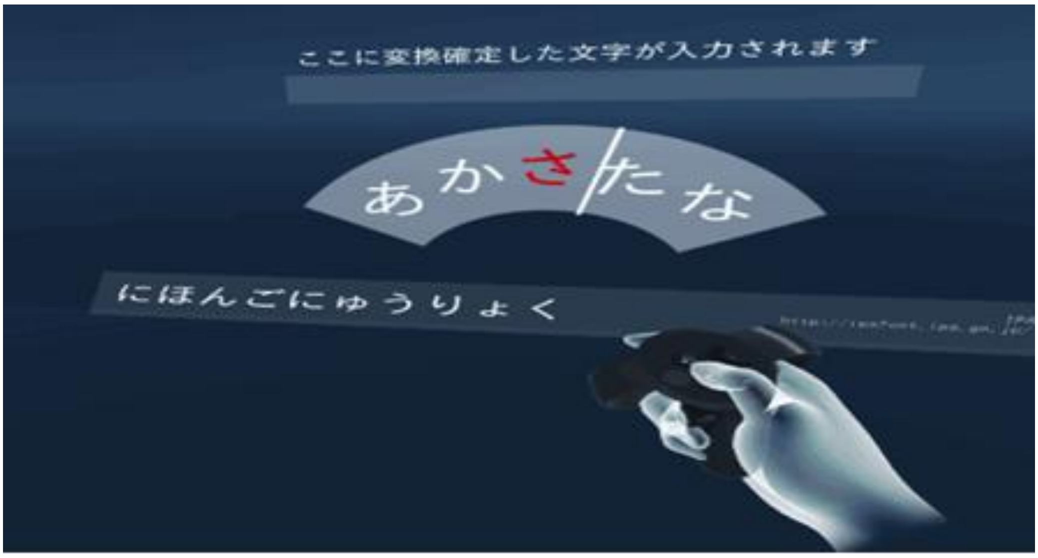


Figure 1.3: Existing VR typing solution is still not convenient. [28]

1.3 The Potential and Challenges of VR Workspaces

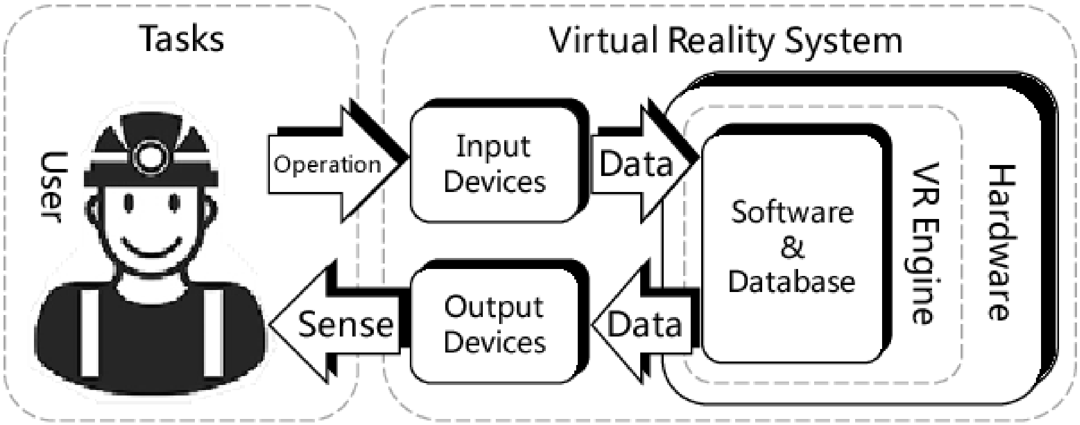


Figure 1.4: Components of a VR system. [8]

The integration of immersive realism in VR significantly enhances work efficiency. By creating virtual environments that closely mimic real-world scenarios, users can achieve a heightened sense of presence, surpassing traditional digital interfaces. This enhances presence fosters a focused and immersive work environment, potentially improving productivity and task completion. Microsoft has introduced the concept of VR workspace (VRWS) as a potential solution that supports typing. VRWS can seamlessly integrate into the existing open-plan workspace (OPWS) without incurring additional space costs, offering a virtual personal space independent of the open office environment [29]. This technology is posited to address the adverse effects of noise and privacy concerns by creating a distinct virtual workspace.

Some VRWS have already been employed to support office work, and notable examples include VRchat [30] and Oculus Virtual Desktop [31]. VRchat, a virtual reality-based social platform with over 2 million users, enables interactions between users as 3D character models. Oculus Virtual Desktop, with a substantial user base, provides excellent image quality and additional features to aid users in their work.

While VRWS holds promise in revolutionizing workspace dynamics, it's essential to recognize the inherent challenges. One primary challenge lies in the lack of standardized environmental design across VRWS applications. In a diverse array of VR environments, ensuring that the design doesn't inadvertently compromise user focus or work efficiency remains a significant concern. Addressing these challenges becomes pivotal for optimizing VRWS's potential and ensuring it aligns with user needs and expectations.

Figure 1.5: The scene of VRchat office and Oculus Virtual Desktop. [30, 31]

1.4 The issue of Jitter

Jitter in VR is the phenomenon of fluctuations in the signal that is generated by the controller or the headset and sent to the VR software. [32,33,34] Jitter can affect both the position and the orientation of the tracked objects, causing them to appear unstable or inaccurate in the virtual environment. Jitter can have a negative impact on the user's performance, experience and comfort in VR. [33,34]

One of the factors that causes jitter in VR is the technology used for tracking. [35] Different sensors have different levels of precision, latency and noise, which can affect the quality of the signal. For example, optical tracking systems can suffer from occlusion, interference or reflection issues, while magnetic tracking systems can be affected by metallic objects or other magnetic fields.

Jitter can also be caused by software issues, such as rendering problems, performance heuristics or reprojection settings. Such as infrared or visible light cameras [36]. This type of jitter is observable across a spectrum of VR tracking devices, encompassing cameras embedded in headsets, hand-tracking systems like Leap Motion [37], and technologies like the Kinect [38]. The primary objective of these sensors is to ascertain the absolute pose of the input device, whether it be a VR controller associated with commercial VR HMD or the user’s hands in the case of Leap Motion. The sensors rely on detecting visually salient entities, such as beacons, shapes, or markers, enabling the tracking algorithm to determine the device’s pose.

However, challenges arise when these beacons, shapes, or markers are not consistently visible to the sensors due to occlusion. In such instances, the tracking algorithm may experience difficulties, leading to abrupt changes in the virtual VR controller’s pose as perceived by the user. Even when all markers or beacons are fully visible, the tracking algorithm's output may still exhibit noise in the pose, attributable to simplifying assumptions within the algorithm or limitations inherent to the sensors.

Jitter can be measured by its magnitude and frequency. The magnitude of jitter is the amount of deviation from the expected position or orientation of the tracked object, while the frequency of jitter is how often the deviation occurs. Jitter can be classified into two types: random jitter and deterministic jitter. Random jitter is unpredictable and varies in magnitude and frequency, while deterministic jitter is predictable and has a fixed pattern or source.

Jitter can have various effects on the user's interaction with VR. For example, jitter can reduce the accuracy and precision of pointing tasks, especially for distal targets or small objects. Jitter can also increase the difficulty and frustration of selection tasks, especially when using a button press to confirm the selection. Jitter can also affect the user's perception of depth, distance and size of virtual objects, as well as their sense of presence and immersion in VR. Jitter can also cause motion sickness or cybersickness, which are symptoms of discomfort or nausea induced by VR.

Figure 1.6: Jitter results in inaccuracies in the representation of the virtual model's position, leading to difficulties in interaction. The position of the virtual objects in the image appears blurred and unclear due to jitter, preventing real-time and accurate representation of their locations.

Figure 1.7: "The timing of the pulse". One of the performance indicators of stability is jitter characteristics. Jitter causes VR hand position to differ slightly over time. Middle image show the fingers are bent, but right image show fingers are straight. In typing task, this kind of difference will reduce typing accuracy.

1.5 Motivation / Purpose of this research

In the post-pandemic era, Virtual Reality Workspaces (VRWS) emerged as promising solutions to overcome the limitations of traditional Online Productivity Workspaces (OPWS), especially in remote work and education settings. However, the effective integration of VRWS into professional environments faces two critical barriers: the lack of standardized design principles and the inefficiency of text input methods. This research aims to address these challenges through the following two objectives:

- Objective 1: Establishing Design Standards for VRWS

Despite the insights provided by existing OPWS research, their direct applicability to VRWS remains largely unexplored. To ensure VRWS can support user attention and productivity, there is a pressing need to establish universally applicable design principles tailored to the unique characteristics of virtual environments. This research investigates how factors within VR environments influence work efficiency and proposes design standards that align with best practices from OPWS while leveraging the immersive potential of VR.

Key research questions:

RQ1: How can VRWS uphold or improve work efficiency?

RQ2: What factors in virtual reality environments affect work efficiency?

- Objective 2: Overcoming Text Input Challenges in VR

Efficient text input remains one of the most significant technical challenges in VR, especially in immersive office settings where productivity is essential. Existing solutions, such as wearable devices or specialized controllers, introduce high costs and usability issues [29,39,40,41]. Additionally, jitter caused by rendering inconsistencies undermines the accuracy of virtual hand movements, making text input unreliable. This research proposes an innovative approach utilizing the built-in cameras of head-mounted displays (HMDs) to capture hand movements and typing actions. By analyzing back-of-hand images, motion history images (MHI), and employing a two-stream LSTM network with Kalman filtering (KF), this method aims to reduce jitter, improve text input accuracy, and maintain natural typing habits.

Key research questions:

RQ3: How can the built-in cameras of HMDs accurately detect typing actions, even when the line of sight is obstructed?

RQ4: How can jitter be reduced to enhance the accuracy of virtual hand movements in VR, thereby improving the user experience?

RQ5: How does the effectiveness of a VR typing solution influence the degree of change in users' typing habits compared to regular typing under different performance?

By addressing these two critical challenges: establishing design standards for VRWS and improving text input solutions, to advance VR technology as a practical tool for professional productivity, ensuring its effective integration into modern workspaces.

1.6 Dissertation Organization

This dissertation is organized as follows:

- Chapter 1 Introduction
  - 1.1 Overview
  - 1.2 Challenges in Text Input for VR Office Applications
  - 1.3 The Potential and Challenges of VR Workspaces
  - 1.4 The issue of Jitter
  - 1.5 Motivation / Purpose of this research
  - 1.6 Dissertation Organization
- Chapter 2 Literature Review
  - 2.1 Methodology of Literature Review

- 4.1 Data Collection
- 4.2 Motion History Image
- 4.3 Network Architecture
  - 4.3.1 Two Stream ResNet18
  - 4.3.2 LSTM
  - 4.3.3 Kalman Filter
  - 4.3.4 Key Point
- Chapter 5 Performance
  - 5.1 Ablation Study
  - 5.2 Performance Comparison
    - 5.2.1 Participants and Equipment
    - 5.2.2 Comparison Conditions
    - 5.2.3 Metrics and Data Collection
  - 5.3 Result
    - 5.3.1 Result of Ablation Study
    - 5.3.2 Result of Performance Comparison
  - 5.4 Discussion
    - 5.4.1 Discussion on Ablation Study
    - 5.4.2 Discussion on Performance Comparison
- Chapter 6 Typing Experiment

    - 6.1 Experiment Design
        - 6.1.1 Participants
        - 6.1.2 Equipment
        - 6.1.3 Experimental Conditions
        - 6.1.4 Experiment Procedure
        - 6.1.5 Data Collection
    - 6.2 Questionnaire
    - 6.3 Result
    - 6.4 Discussion
- Chapter 7 Typing Behavior
    - 7.1 Preliminary Experiment
    - 7.2 Typing Behavior Experiment
        - 7.2.1 Participants
        - 7.2.2 Equipment
        - 7.2.3 Experimental Conditions and Procedure
        - 7.2.4 Data Collection
        - 7.2.5 Use Typing Habit Data to Cluster
        - 7.2.6 Use Typing Habit Data to Re-clustering
        - 7.2.7 Use Typing Habit Difference Data to Cluster
        - 7.2.8 Statistical Test
        - 7.2.9 Result Summary
        - 7.2.10 Discussion

Chapter 2

Literature Review

This chapter embarks on a state-of-the-art review focusing on empirical studies addressing working in VR. By delving into existing literature, we aim to illuminate the current landscape, identify gaps, and pave the way for our research endeavors.

2.1 Methodology of Literature Review

The search strategy utilized the following keywords and operands: keyword group a: (“virtual reality” or “VR”) combine with keyword group b: (“workspace”, “office”, “office work”, “open plan workspace”), keyword group c: (“typing”, “text input”, “keyboard”), keyword group d: (“hand tracking”), keyword group e: (“jitter”) and keyword group f: (“design criteria,” “design standards,” “efficiency,” “attention”). The inclusion criteria for the search encompassed conference proceedings and journal articles published between 2001 and 2023. The initial search yielded a total of 78 abstracts.

After excluding abstracts that contained relevant keyword groups but lacked significant relevance (n=19), the remaining 59 publications underwent a thorough full-text review. Subsequently, 53 publications were deemed relevant for inclusion in this review, with 6 non-empirical research excluded. Notably, while studies on text input challenges and hand tracking in VR are relatively abundant, research explicitly addressing design standards for VRWS remains scarce. This gap highlights the need for further investigation into how VRWS design principles can enhance work efficiency and attention management. The literature review process is visually represented in Figure 2.1.

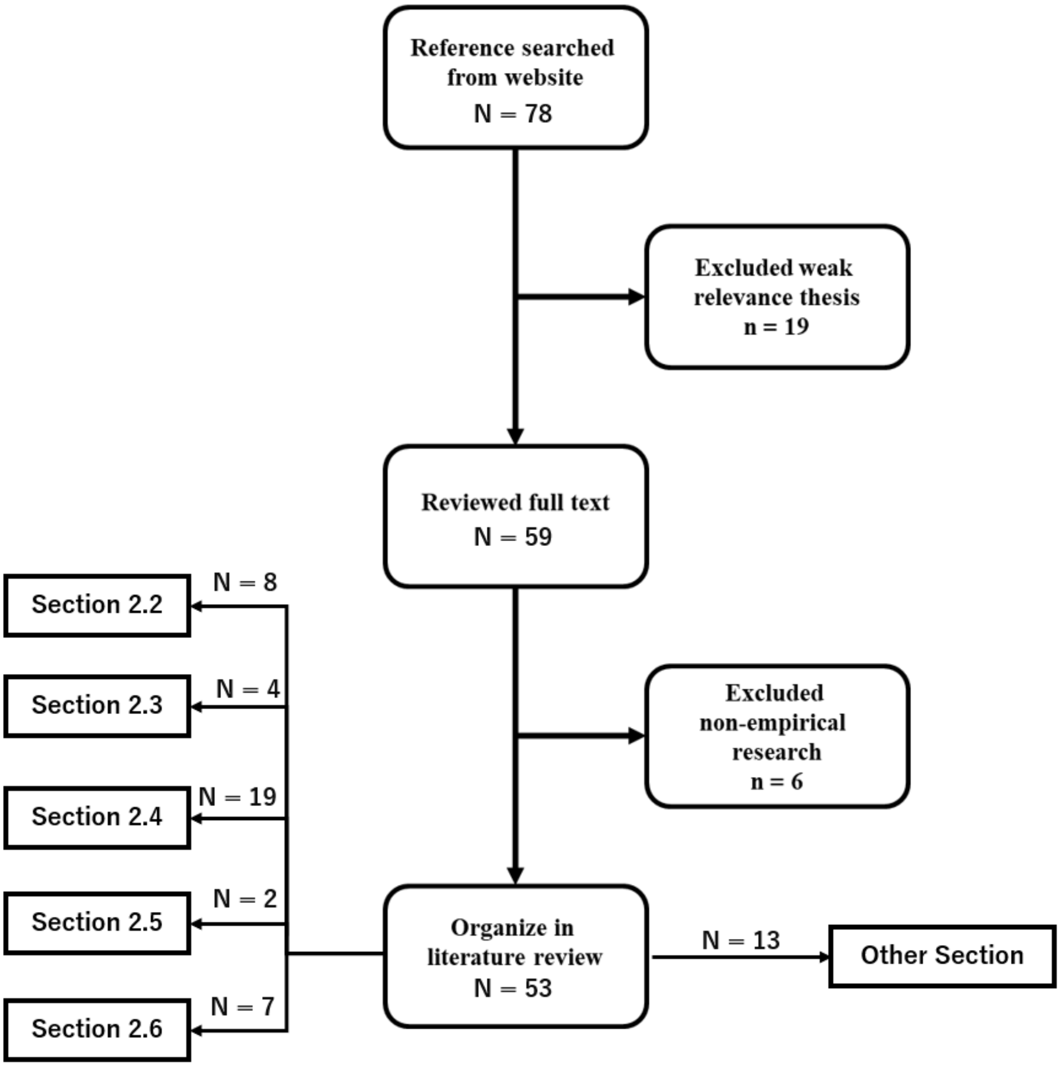


Figure 2.1: The process of literature search

2.2 Work in OPWS

Roelofsen's research found that noise from the OPWS reduces work efficiency [42]. Oommen claimed the OPWS often produces adverse effects such as noise, stress, conflict, high blood pressure, and high turnover rate, etc. among them [43]. Treasure thinks the noise has the most apparent impact on work efficiency. Compared to quiet rooms, noise interference in OPWS reduces work efficiency by one third [44]. Research from Xymax

Real Estate Institute showed that it would be tight in some countries with demanding space utilization requirements. For example, in Japan, some companies are often unable to find enough space [45]. Evans and Johnson found that noise from the OPWS creates extra stress on workers [46]. Humphries believes that a pleasant office environment should be a cozy space that has no visual and auditory interference [47]. Veitch et al. and Karasek & Theorell found good lighting, a controlled sound environment, and plenty of natural light in the OPWS is suitable for work efficiency [48, 49].

2.3 Work in VRWS

In recent years, the application of VR for office work has garnered significant attention, particularly with the rise of remote work and the increasing need for virtual collaboration. The term "office work" encompasses a spectrum of knowledge and administrative tasks traditionally conducted in a physical office environment or in proximity to other work environments, such as laboratories, healthcare facilities, or manufacturing sites. The evolving landscape of office work, shaped by digitalization, technological advancements, and remote work policies, has blurred the boundaries of where and how work is conducted [50].

Several studies have delved into novel prototypes utilizing VR for office tasks. Biener et al. introduced a unique approach, featuring side-by-side and in-depth screens with micro-movements for input, offering applications in confined spaces [51]. Kim and Shin explored side-by-side screens akin to traditional desktop usage, unveiling untapped potential for interaction patterns on large VR screens [52]. Both studies emphasized the potential for mobile office work in on-the-go situations and limited spaces, proving beneficial for professionals working in confined or crowded environments [51,52]. Furthermore, the immersive nature of VR environments has been recognized as an advantage in terms of information security, as sensitive information remains enclosed within the virtual space [52].

However, challenges associated with wearing HMD include physical discomfort and difficulties accommodating eyeglasses within the headsets. [52] Visual discomfort and simulator sickness are also acknowledged concerns. Shen et al. identified fatigue as a potential challenge, emphasizing the need to address user exhaustion during extended VR sessions. [53]

Despite the advancements, the impact of VR environment design on work efficiency remains understudied. Current VR environments heavily rely on personal customization, leading to skepticism about their positive impact on work efficiency. Our investigation reveals a gap in research regarding VRWS design standards for maintaining or improving work efficiency. While some studies propose solutions for OPWS, it is unclear if these can be directly applied to VRWS.

To summarize, the lack of research into standardized VRWS design principles to improve work efficiency represents a significant gap. This gap forms the foundation for Chapter 3, where we propose specific methodologies to explore VRWS design standards and solutions.

2.4 Type in VR

HMD coupled with a keyboard is a basis for a full VRWS in which users can enjoy a motion-independent robust and immersive virtual office environment [39]. However, one barrier is no robust text entry. Entering long text will become difficult because users wearing HMD cannot see their hands and keyboard. Figure 1.2 illustrates the difficulty, showcasing how simulated keyboards may lack haptic feedback for experienced typists, and not seeing one's hands while typing can be challenging for inexperienced typists [52,52]. In order to see the hands and keyboard while typing in VR, hands or keyboards should be recognized and shown in virtual reality. At present, there are mainly two approaches to solve typing inefficiency problems: the traditional solutions and the machine learning solutions.

Traditional solutions:

Traditional approaches have explored diverse methods to enhance VR typing. For instance, the Leap Motion device was employed to track users' hand movements and gestures, presenting a circular virtual keyboard with 26 keys arranged in two concentric rings [54]. Users could interact by tapping keys with their fingers or employing gestures like swiping or pinching.

Another insightful study compared three distinct conditions for text input in VR: using a standard physical keyboard with mechanical switches, a touch-sensitive physical keyboard with capacitive switches, and a virtual keyboard with mid-air gestures [55]. This comparative analysis sheds light on the nuanced differences in performance and user preference across these varied input methods.

Handwriting input, a popular choice in VR, was investigated in a study where an optical motion capture system tracked users' hand movements. A haptic glove provided tactile feedback, and a speech recognition engine converted handwritten text into speech [56]. This multifaceted approach not only captures the nuances of handwriting but also integrates speech recognition for a comprehensive text input experience.

Utilizing a Bluetooth keyboard paired with a HMD presents another avenue for VR typing. This method allows users to type on a physical keyboard while observing a virtual representation in VR. However, it is noteworthy that this approach is contingent on having a specific keyboard model (such as the Logitech K830) and may not exhibit optimal functionality in diverse environments or postures.

Motion sensors and cameras come into play as an alternative approach, exemplified by Tap ID—a wristband with motion sensors that analyze bone vibrations during finger taps. [57] This facilitates typing into a virtual keyboard or interacting with virtual objects on various surfaces. Nevertheless, challenges related to accuracy, reliability, and calibration for different users or surfaces may impede the seamless adoption of this method.

Gesture recognition emerges as a distinct avenue, with approaches like Thumb Air, and STAR. [58,59] These methods leverage hand gestures to emulate key presses on a virtual keyboard or simulate smartphone typing. While aiming to capitalize on users' familiarity with existing input devices, they also acknowledge potential challenges such as low accuracy, recognition errors, and limitations in supporting diverse text input, including punctuation or symbols.

Diverging from these studies, our research maintains a steadfast commitment to the use of a physical keyboard. While this choice introduces certain technical challenges, opting for a physical keyboard with 3D tactile feedback preserves users' typing habits, ensuring a familiar and comfortable experience. This approach aligns with the principle of being “easy to adapt for users with keyboard input experience”.

Machine learning solutions:

Among all other solutions, the machine learning solution is considered as the most potential support solution.

Hwang et al. proposed a method to estimate 3D human pose from a monocular fisheye camera mounted on a VR headset [60]. Erwin et al. introduce a system to recognize 3D hand poses from a wrist-worn camera via a deep neural network [61]. Jang et al. presents a metaphoric gesture interface for manipulating virtual objects with an egocentric viewpoint [62].

Hand tracking is a technology that enables the detection and tracking of the position, depth, speed, and orientation of a user’s hands using various methods such as headset cameras [63], LiDAR arrays [64], or external sensor stations [65]. This tracking data is analyzed and processed to create a virtual, real-time representation of the user’s hands and their movements within the virtual world. This representation is subsequently transmitted to the respective application or video game being used, allowing users to interact naturally with the virtual environment using their hands.

Unfortunately, LiDAR arrays, or external sensor stations, these kinds of wearable hand tracking solutions often hinder typing efficiency due to the requirement of wearing extra devices. Deep learning solutions offer cost advantages as they only require the cameras embedded in the HMD, eliminating the need for additional hardware [60,61,62]. This also means that the Deep learning solution has less impact on typing efficiency because it does not need to wear extra devices.

However, typing as a task presents unique challenges. When users wear HMD and type in a VR environment, their fingers are often obstructed by the back of their hand. It makes HMD’s cameras difficult to capture the complete view of the typing hands. As a result, the accurate tracking of typing hands positions becomes challenging. A study has been conducted to estimate finger positions during typing by utilizing subtle variations on the back of the hand, using a wrist-mounted camera [60]. Inspired by their work, our approach also focuses on visual features on the back of the hand, extending it to support richer, full typing hands position estimation. Our approach builds upon the insights from their research, focusing on the visual features on the back of the hand, and extending it into a robust and practical VR typing support system.

The mentioned studies each have their own shortcomings. In summary, we believe that solutions that do not support the use of a physical keyboard are often less user-friendly and may deviate from familiar typing methods, potentially affecting typing efficiency. A considerable portion of the research requires additional auxiliary devices, and some studies even necessitate users to wear special equipment to support VR typing. This not only increases office costs but also causes inconvenience associated with wearing such devices.

In terms of target users, our research primarily focuses on the majority of users who cannot touch type and require visual guidance to type effectively while wearing an HMD. According to a survey, approximately 70% of Americans are unable to touch type, meaning they rely on visual feedback to locate keys and maintain typing accuracy [66]. For these users, it is essential to provide a solution that allows them to see their hands while typing in VR. In addition, for experienced typists capable of blind typing, such solutions may seem unnecessary. However, even for this group, challenges such as physical keyboard positioning and misalignment in VR environments can still arise, underscoring the need for robust and adaptable hand-tracking technologies.

In the virtual reality (VR) environment, while enabling a "window" to display the real-world keyboard and hands can serve as a practical workaround (a solution similar to those employed in augmented reality, AR), this approach compromises the immersive experience that is central to VR. This immersion is a key advantage of VR, and such a method is often regarded as a compromise that fails to satisfy users who prioritize complete virtual immersion. Moreover, compared to VR-based workspaces, AR provides a lower level of immersion, making it less effective at preventing environmental distractions (such as real-world visual or auditory interference) and weaker in terms of privacy protection. In contrast, the enclosed nature of VR workspaces (VRWS) allows them to shield users from external disturbances, making them more suitable for tasks requiring a high degree of focus and thus maintaining their unique advantages.

To further balance the trade-off between input efficiency and immersion, considerable research has focused on developing advanced hand-tracking technologies tailored for VR. As discussed in Section 2.4, technologies like Leap Motion, Thumb Air, and gesture recognition aim to enable natural and seamless hand representation without relying on external "windows" to display the real world. Although these solutions show significant potential, they still face limitations in terms of accuracy, usability, and compatibility with users' existing typing habits.

In our pursuit of enhancing office work in VR, we focus on addressing these challenges by developing solutions that integrate physical keyboards with advanced hand-tracking technologies. These efforts form the foundation for Chapter 4, where we detail our methodology and propose a novel approach to overcoming VR typing inefficiencies.

2.5 Typing Behaviors

Despite numerous studies exploring user behavior in VR, our limited investigation reveals a gap in direct research on typing behaviors in VR, particularly regarding variations in typing speed, accuracy, and user preferences. However, several studies, although their goal is not explicitly addressing VR typing behaviors but still provide valuable insights that inspire further exploration.

One valuable avenue of inquiry involves the analysis of user's VR touch typing models. [67] Unlike other classic touch-typing studies, physical keyboards are replaced by tablets, users wearing HMD engaging in touch typing tasks. Figure 2.2 illustrates the evolving positions of users' typing finger touchpoints over time, showcasing unique aspects of VR typing behavior.

It is crucial to note that in this VR touch typing task, the deviation in finger touchpoints dynamically adjusts on the virtual keyboard corresponding to the tablet, preventing frequent typing errors and experiment interruptions. This distinct feature sets this study apart from others, where virtual keyboards do not dynamically adapt to user's evolving typing patterns. Therefore, the observed changes in finger touchpoints in this study can be argued to represent, to a certain extent, the unique variations in typing behavior specific to VR.

In addition, it is imperative to acknowledge certain limitations within this study. The choice of substituting a physical keyboard with a tablet inherently impacts user's typing experiences. Users accustomed to the three-dimensional tactile feedback of a physical keyboard may find the tablet solution unaccustomedness, leading to alterations in typing behaviors such as changes in typing speed and adjustments in the use of preferred fingers.

Furthermore, it is essential to recognize that the primary objective of this related work was not specifically to explore VR typing behaviors. Consequently, the study, while offering valuable insights, should be considered as a preliminary exploration, providing a foundational understanding but not exhaustive insights into the realm of VR typing behaviors.

Another research found differences in hand movement area were observed across various typing conditions [68]. Figure 2.3 displays the movement patterns of the right and left hands based on tracking data from the left hand and right-hand markers in a two-dimensional plane.

To comprehensively understand the shifts in VR typing behaviors, further investigations specifically tailored to unravel the intricacies of VR-native typing behaviors are warranted. Addressing the influence of the choice of input devices and aligning the study objectives more closely with the exploration of VR typing behaviors would contribute to a more nuanced and holistic understanding of the subject matter.

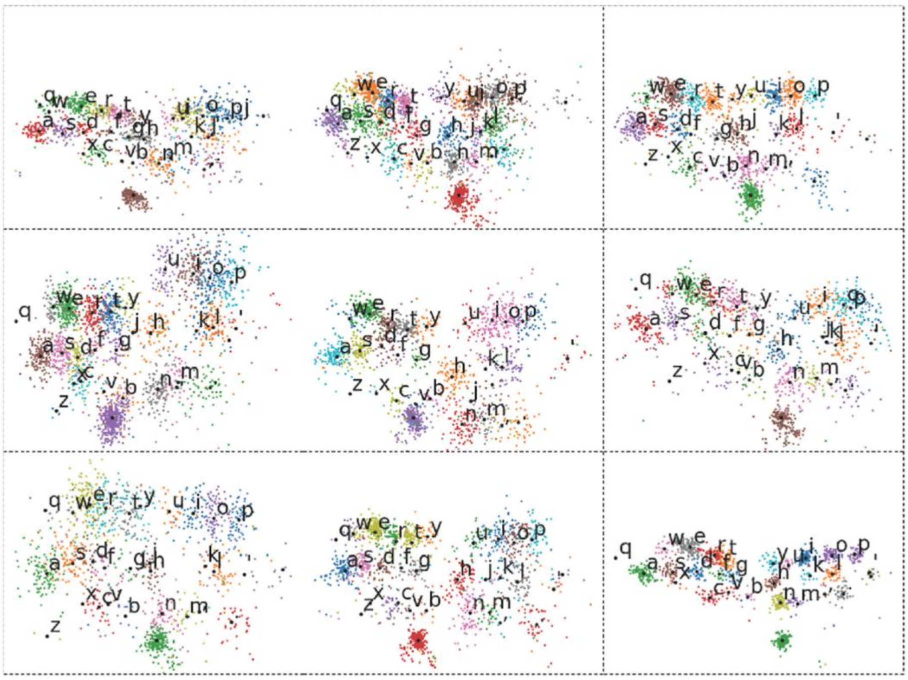


Figure 2.2: The positions of users' typing finger touchpoints [68]. The positions of the

letters represent the user’s intuitive recognition of touch-typing locations.

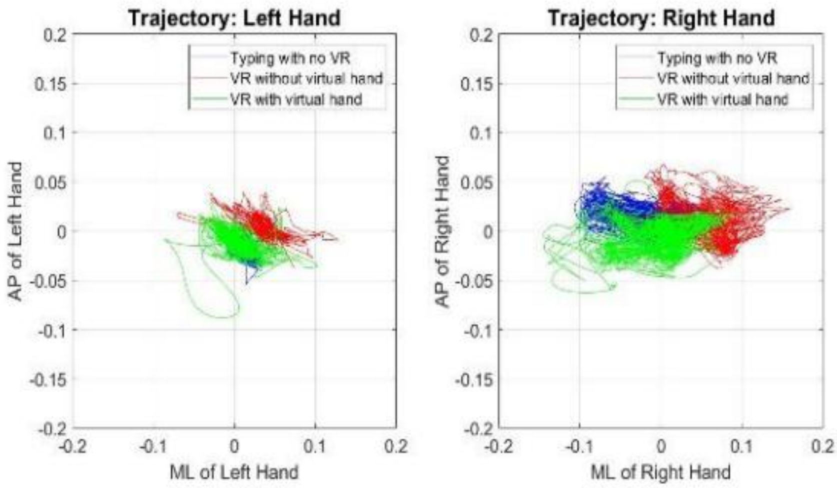


Figure 2.3: The positions of users’ typing finger touchpoints [68]. The positions of the letters represent the user’s intuitive recognition of touch-typing locations.

2.6 Jitter in VR

In VR systems, jitter refers to small fluctuations in the signal and is a significant factor that can adversely affect motor performance and user experience. Despite continuous technological advancements, effectively reducing or eliminating jitter remains a challenge, especially in tracking systems that are integrated into various HMDs. The impact of jitter on VR systems has been extensively studied by various researchers. Teather et al. [69] conducted an analysis and found that even a small amount of spatial jitter (0.3 mm) in the input device could noticeably decrease user performance. Moreover, it has been observed that larger jitter levels have a more pronounced negative effect on user performance, especially when dealing with smaller targets [70]. Batmaz et al. [71] also support this finding, noting that user performance declines significantly in terms of time, error rate, and throughput as the jitter level increases. Additionally, Moaaz et al. [72] conducted experiments where they artificially introduced 0.5°, 1°, and 1.5° jitter to the VR system, leading to a substantial increase in the user’s error rate with each the increment in jitter level.

In conclusion, considering the above research highlighting the adverse effects of jitter on user performance in VR systems, we firmly believe that an efficient VR typing support system must exhibit low jitter characteristics.

Reducing jitter in VR systems is a challenging task that requires careful consideration of various factors. Because the causes of jitter are complex and varied, there have been various studies proposing different methods to reduce jitter:

- Improving the hardware and software of the tracking system. This can include using more accurate sensors, faster processors, higher bandwidth, and better algorithms to minimize the noise and latency in the signal [73,74,75].
- Calibrating the tracking system regularly and properly. This can ensure that the tracking system is aligned with the physical and virtual spaces, and that the errors are minimized or corrected [73,74].
- Applying filtering techniques to smooth out the signal. This can include using low-pass filters, Kalman filters, or predictive filters to reduce the high-frequency fluctuations and estimate the future position and orientation of the tracked object [73,74].
- Adjusting the parameters of the VR system according to the task and user preferences. This can include changing the field of view, the level of detail, the rendering quality, and the motion scale to optimize the performance and comfort of the VR system [73,75].

Applying filtering techniques to smooth out the signal is a common approach in mitigating high-frequency fluctuations in tracking data. Prior studies have employed methods such as low-pass filters, Kalman filters, and predictive filters to reduce jitter and estimate the future position and orientation of the tracked object [73,74]. However, the unique characteristics of typing movements make some of these techniques less effective for this specific task.

Typing actions are characterized by their continuous, highly repetitive, and subtle nature. These micro-movements, while crucial for accurate text input, can easily be misinterpreted by low-pass filters as high-frequency noise. As a result, low-pass filters may excessively "smooth" these fine movements, inadvertently compromising the accuracy of hand tracking. Similarly, predictive filters, which rely on modeling future positions based on previous movement patterns, struggle to perform effectively in this context. The high similarity and repetitive nature of typing actions can lead to prediction errors, as the filter may fail to distinguish between slight variations in consecutive movements.

In contrast, the Kalman filter proves to be the most suitable choice for addressing the challenges of typing tasks in VR. Its ability to dynamically adapt to subtle changes in movement allows it to accurately track the small, precise actions involved in typing without over-smoothing or introducing significant prediction errors. The Kalman filter's capability to balance noise reduction and motion fidelity makes it particularly well-suited for maintaining the accuracy and responsiveness required for text input tasks.

Based on these comparisons, our method employs a Kalman filter to ensure precise tracking of typing movements, minimizing jitter while preserving the subtle nuances of hand actions. This approach not only enhances the typing experience in VR but also aligns with the specific demands of text input tasks, as demonstrated in subsequent sections of this study.

Chapter 3

Exploration of VRWS Design

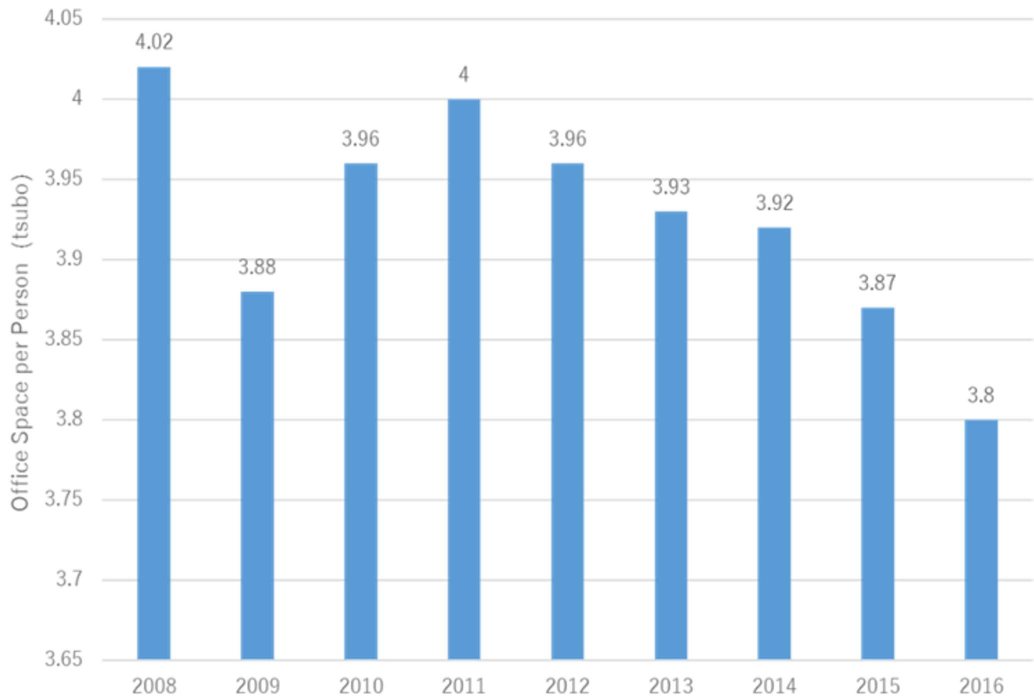
3.1 Overview

OPWS is an office style that allows many employees to work simultaneously in the wall-less, partition-less environment [76,77], as shown in Figure 3.1. OPWS is characterized by a high sense of openness, low cost, encouraging cooperation, and improving the collective wisdom of the team. More and more companies have chosen this kind of office since its birth in the last century.

(Source: https://bunshun.jp/articles/-/2293)

Figure 3.1: JAL procurement office.

Although OPWS has already proven its value, but still there exist many shortcomings. The environment of OPWS not only directly affects people's health and enthusiasm for work but also affects work efficiency [76,77,78]. Not only the OPWS full of auditory and visual interference, but also the low level of privacy protection causes psychological stress to employees and reduces work efficiency. Although many researchers have been working on it to solve these problems, they still cannot declare that these problems are entirely solved. The obvious point is that most of the proposals suggest creating an additional workspace that needs extra cost. For example, the proposal of providing employees with various additional spaces to alleviate the problem [79], which will be very difficult in some countries with demanding space utilization requirements, such as Japan, and some companies are often unable to find enough space [80]. Even the workspace for per person is decreasing year by year, as shown in Figure 3.2.



(Source: Xymax Real Estate Institute,
https://www.xymax.co.jp/english/research/images/pdf/20160921.pdf)

Figure 3.2: Office space per person in Tokyo

On the other hand, VRWS, which is a virtual personal space independent of OPWS, has the potential to solve the psychological pressure caused by the lack of privacy protection of employees in the OPWS environment because it can reduce the auditory and visual interference in the workspace. However, no studies have shown how VR (Virtual Reality) environments can be designed to maintain or improve work efficiency. On the other hand, there are some opinions that virtual reality technology cannot benefit the work itself [81].

According to our investigation, there is currently no research to confirm what VRWS design standards can maintain or improve work efficiency. Although there have been a couple of research proposing solutions to improve the shortcomings of OPWS, it is not sure whether the solutions for OPWS can apply to VRWS.

3.2 Design of OPWS and VRWS

OPWS is very popular all over the world, and different types of work content will also produce OPWS with different characteristics. For example, the call center is a typical noisy OPWS, because answering a call is an essential task in the call center. In this environment, work noise is unavoidable. There are also diverse different types of OPWS. For example, librarians rarely worry about noise.

It is difficult to find a typical noisy OPWS in the area where the author lives. In order to control experimental settings, we decided to use the CAVE (Cave Automatic Virtual Environment) system to simulate a typical noisy OPWS.

The CAVE system is a projection-based virtual reality system, which consists of several projection screens surrounding the participants and can produce a completely immersive virtual environment. At the same time, mini speakers were arranged around the CAVE system to restore the simulated OPWS sound environment as much as possible. Therefore, the CAVE system used in this experiment can make the participants feel the real appearance of a noisy OPWS very well. The experimental arrangement of this study based on the CAVE system is shown in Figure 3.3.

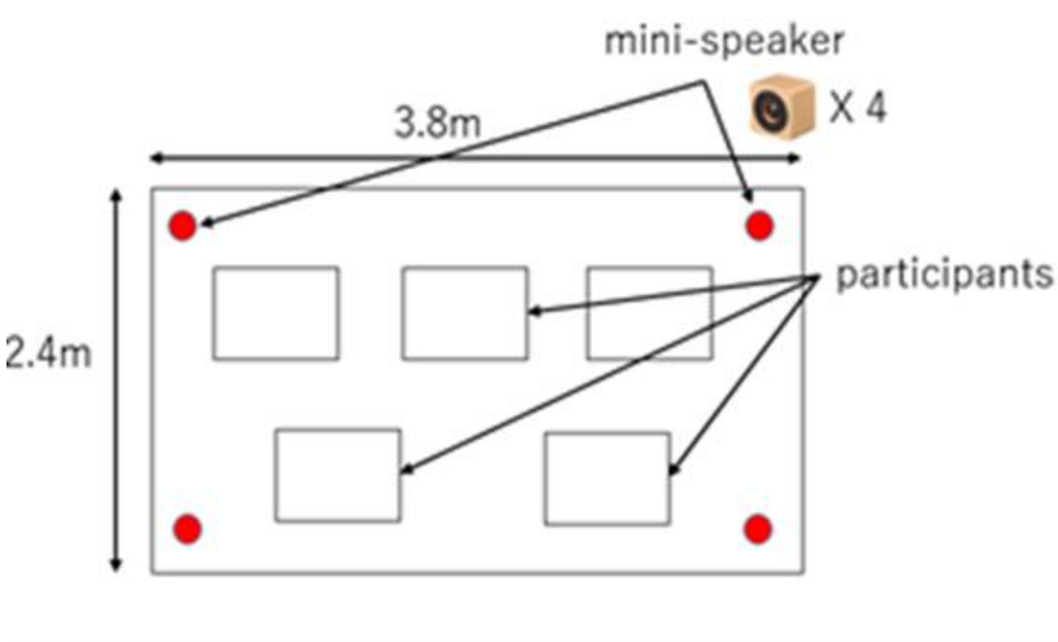


Figure 3.3: Experimental arrangements in the CAVE system

Five participants in the group performed experiments together in the CAVE system. The system included five seats, each equipped with a laptop and a mouse to allow the participants to perform their “work.” The purpose of the CAVE system was to simulate a surrounding environment that, while not directly related to the participants' tasks, could introduce external factors that impact their performance.

For the content played in the CAVE system, the simulated OPWS chosen for this experiment was the mission center of NASA (National Aeronautics and Space Administration) [82]. This environment was specifically selected to replicate the noise and dynamic characteristics of an OPWS, such as background noise, colleagues moving around, and frequent interruptions. One of the frequent activities simulated in this content was the exchange of information among employees, further contributing to the sense of a bustling and potentially distracting workspace. Figure 3.4 shows the mission center of NASA, and Figure 3.5 illustrates the scene in the OPWS condition.

By immersing participants in this environment, the experiment aimed to replicate the stressors commonly associated with OPWS conditions, allowing for a better understanding of how such external factors influence participants' performance in tasks requiring focus and attention.

(Source: https://yaruzou.net/gstv-space-view-iss)

Figure 3.4: Mission center of NASA

Figure 3.5: The scene of the experiment in the CAVE system

3.3 VRWS

We assumed a VRWS with excellent OPWS characteristics, which was an environment without visual and auditory interference, with good lighting, sufficient natural light, and privacy protection, would expect to maintain or improve work efficiency. In order to make VRWS meet the above requirements, we did the following steps.

Avoiding visual and auditory interferences:

To avoid visual and auditory interference from the environment, we decided to use a combination of HMD and noise-canceling earphones. The HMD could completely isolate the visual interference in the environment, and the muffler headphones could eliminate most of the auditory interference. Figure 3.6 shows a combination of HMD and noise-canceling earphones.

Figure 3.6: HMD and noise-canceling earphones

The HMD used in this experiment is Acer Windows Mixed Reality headset AH101. Table 1 shows HMD attribute.

Table 3.1: HMD attribute.

Model number	AH101
Field of view	95° (Fresnel lens)
Display size	2.89 inch ×2
Screen resolution	2880 x 1440 (simple eye：1440 x 1440)
Refresh rate	60 Hz (HDMI 1.4) / 90 Hz (HDMI 2.0)
Size	195.8 (W)×143.4 (H)×384.2 (D) mm
Weight	440

Good lighting and enough natural light:

In order to create a pleasant lighting environment, in the initial design stage of the virtual model, we increased the brightness of the model and used natural light sources instead of ordinary light sources to make the light fill the entire virtual space.

For the requirement of enough natural light, we designed some large floor-to-ceiling windows to replace the walls on either side of the VRWS.

Privacy:

For privacy protection, we designed VRWS that could not share the workspace but could be used one by one. In this process, people can feel like a personal office environment. The related images have been shown in Figure 3.7 and Figure 3.8.

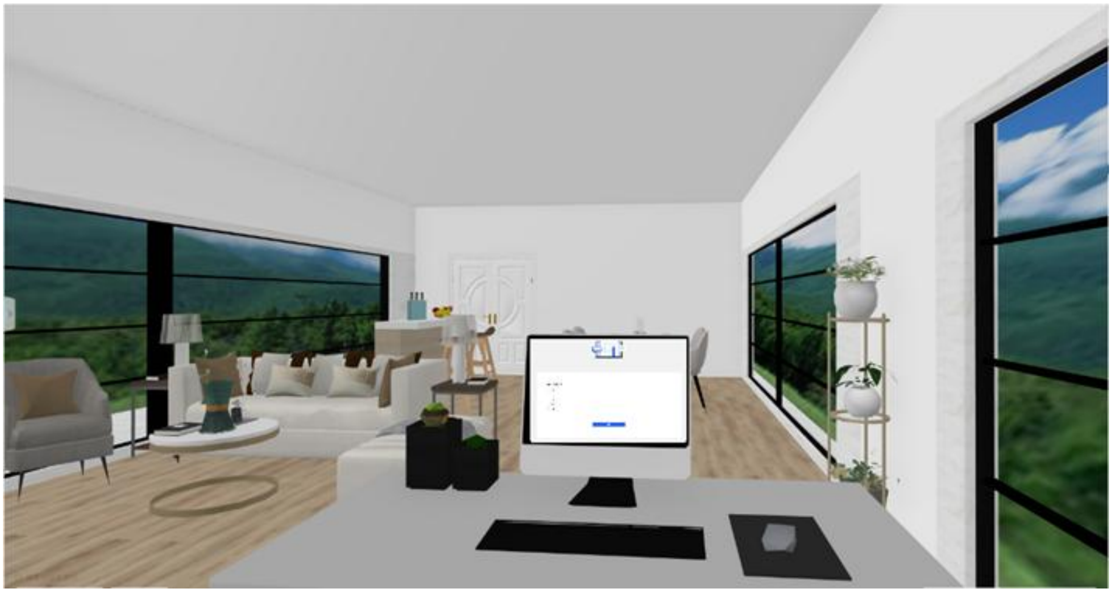


Figure 3.7: User's vision in VRWS

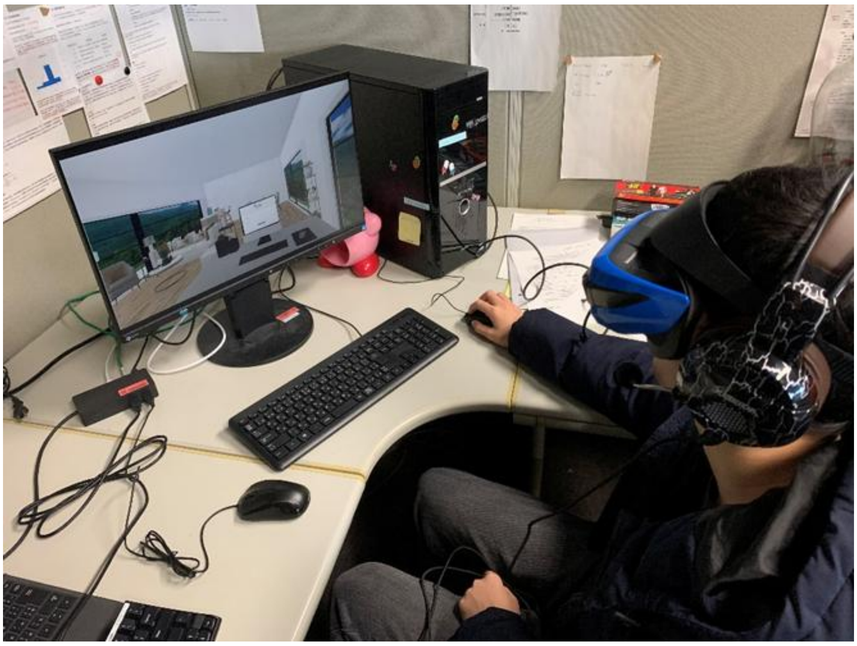


Figure 3.8: The scene of the experiment

In this study, we adopted Unity3D to develop OPWS. Unity3D is a cross-platform 3D engine with a friendly development environment. It is easy to create a virtual environment or virtual model with this powerful engine. Creating a spherical video player that displays a panoramic video is an important step. Then, we generate a panoramic image as a texture and mapping to the sphere. The result is shown in Figure 3.9.

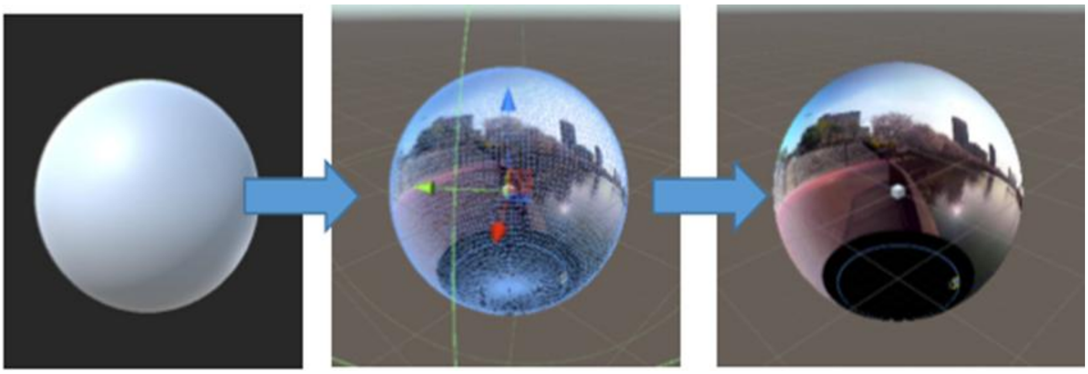


Figure 3.9: Sphere image for VRWS.

In order to maintain the resolution of the panorama, the mesh size of the sphere needs to be meticulous ten times than the standard setting. The left image in Figure 3.10 has been used in this research, and the right-side image is considered in the standard setting.

Figure 3.10: Compare two meshes properties.

3.4 Experiment

3.4.1. Questionnaire

Following research on Emotional Engineering [83] and Versatility of Building Language Description [84], we designed the questionnaire whose questions are measured on the seven levels. A small value is for a positive evaluation, and a large value is for a negative evaluation. The set adjective pairs are shown in Table 2. The reason for choosing these phrases is because they can express people’s feelings where they are at the workspace.

Table 3.2: Adjective pairs for SD evaluation.

	Positive evaluation	Negative evaluation
Q1	Broad view	Narrow view
Q2	Low psychological pressure	High psychological pressure
Q3	Free atmosphere	Non-free atmosphere
Q4	Comfortable	Uncomfortable
Q5	Well-lighted	Ill-lighted
Q6	Not tired	Getting tired
Q7	Natural feeling	Strange feeling
Q8	Grace	Graceless
Q9	Relaxing	Not-relaxing
Q10	Cheerful	Depressed
Q11	Easy to work	Hard to work
Q12	Not noisy in movement	Noisy in movement
Q13	Enjoyable	Not enjoyable
Q14	Not noisy in sound	Noisy in sound
Q15	Motivated	Unmotivated
Q16	Efficient	Inefficient

3.4.2. Experimental Process

In total, 20 postgraduate students participated in the experiment, including 9 females and 11 males, who were between the ages of 24 and 30. All the participants were fluent in English.

Table 3.3: Participant Information.

Participant	Gender	Age.
1	Female	24
2	Male	27
3	Male	25
4	Female	25
5	Female	28
6	Male	27
7	Male	27
8	Female	25
9	Female	25
10	Male	27
11	Male	25
12	Female	28
13	Male	30
14	Male	29
15	Male	26
16	Female	25
17	Female	27
18	Male	24
19	Female	24
20	Male	26

Before the start of the experiment, we assigned all the participants randomly to group A, B, C, and D, as shown in Figure 3.11. Each group consists of five participants. Among them, groups A and C performed OPWS experiments before VRWS experiments. Group B and D performed experiments in the reverse order.

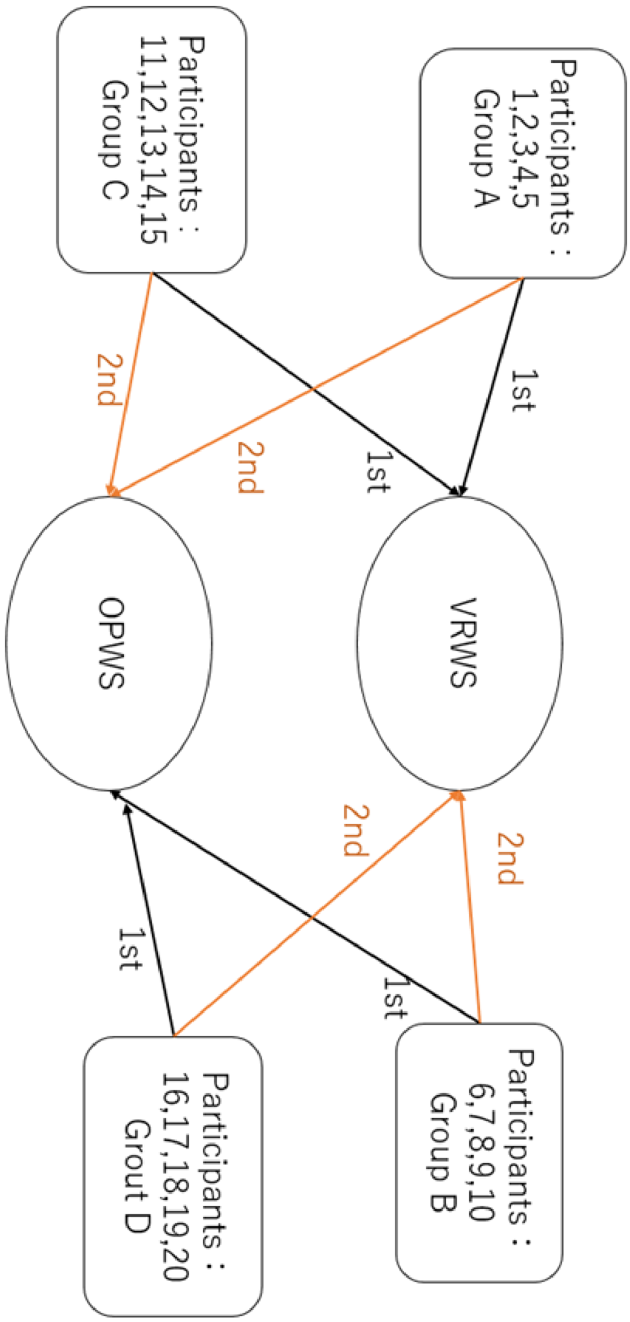


Figure 3.11: Experiment condition for each group.

Each experiment should be controlled within 50 minutes, and after the experiment, a questionnaire was issued. After the experiments, all the data and questionnaires were collected to compare OPWS and VRWS.

In order to rule out errors due to condition differences, the participants were requested not to use all tools except a mouse during the answering process in both settings. However, in the VRWS experiment, an HMD with a computer was provided to the participants to complete the experiment. In the VRWS experiment, each participant experimented alone. The experiments were conducted based on the following process as shown in Figure 3.12.

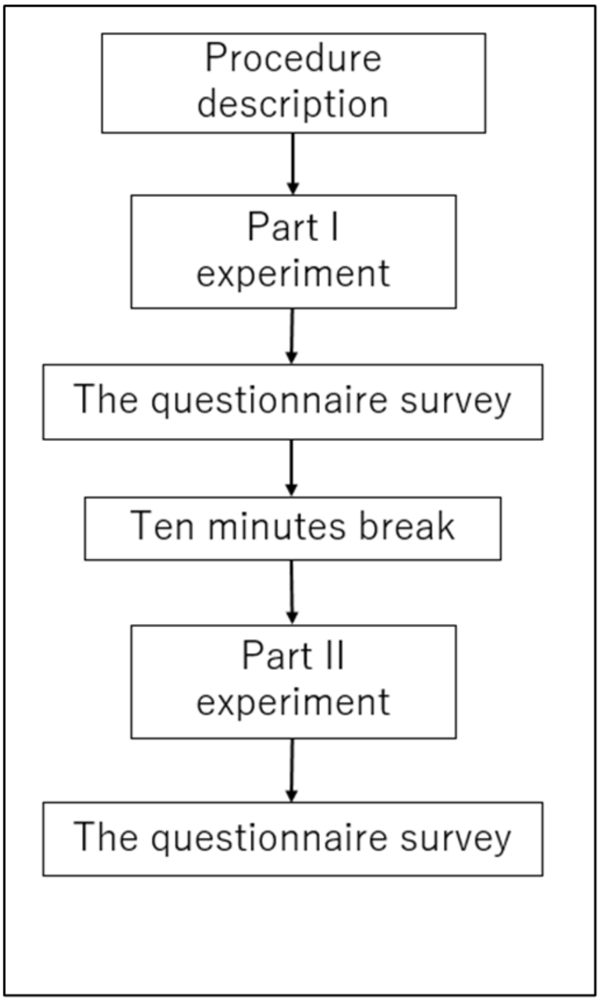


Figure 3.12: Experimental process

3.5 Cognitive Assessment Battery Test

In this experiment, each participant was required to complete his/her "work" in OPWS and VRWS. Therefore, we adopted the CAB (Cognitive Assessment Battery) test consisting of no language-based questions with only numbers and pictures. This test would avoid deviation, such as different understanding speeds and understanding difficulty caused by different languages.

The purpose of the CAB test is to measure people's logical thinking ability. Thus, in this "work" process, the participants were expected to concentrate on solving the test as an essential requirement. We assumed that there was a relationship between the CAB test results and work efficiency.

Every participant received an electronic test containing 45 questions. The questions were designed with reference to some related research [85, 86]. Some examples of the CAB test are shown in Figure 3.13.

The left image in Figure 3.13 shows the mental arithmetic question. Participants need to pay attention to calculate mental math. The middle image shows the logical reasoning question. Here the participants need to focus on searching for the logical relationship among pictures. The right image shows the cipher question. Participants need to apply both local and critical thinking to solve the cipher question. All the types of CAB tests could be answered by mouse operation. Every participant needed to try their best to provide correct answers for all questions provided in the OPWS and VRWS experiments. The participants who answered the maximum number of questions within the shortest time considered that they had more work efficiency.

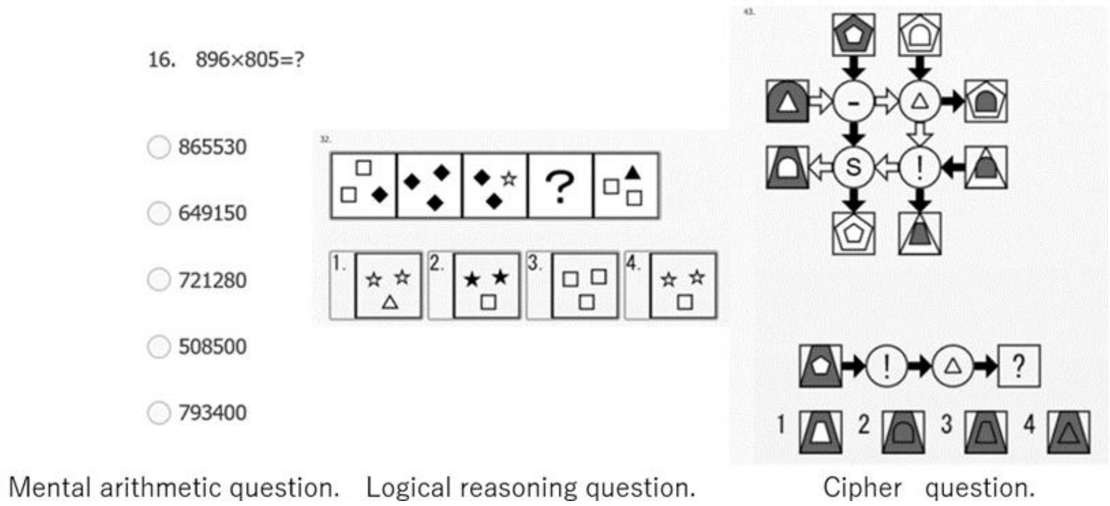


Figure 3.13: Examples of CAB test.

At the same time, these three kinds of test questions would appear in the same proportion in each set of test papers for each participant. The ratio of the test types is shown in Figure 3.14.

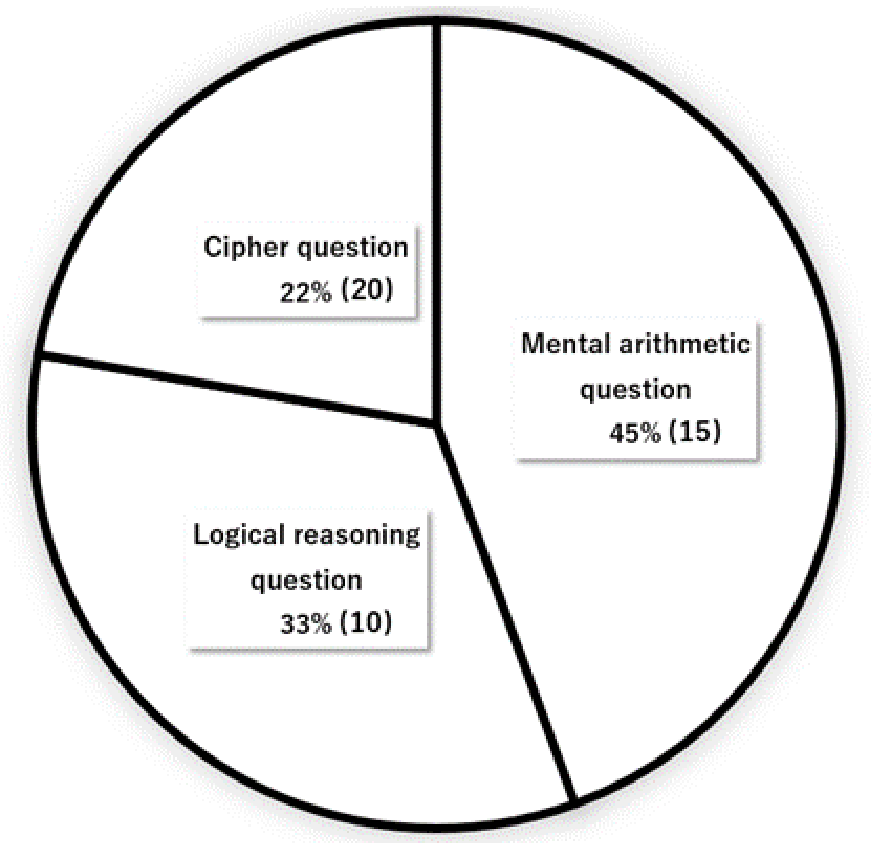


Figure 3.14: The ratio of the 3 types of questions in one CAB test

3.6 Results

The following Table 3.4 displays the total time (minutes) spent by each participant to answer all the questions in each setting.

Table 3.4: Time differences in each setting.

Participants	OPWS (min)	VRWS (min)
1	22	21
2	26	23
3	25	25
4	19	17
5	20	19
6	23	22
7	25	24
8	28	27
9	26	23
10	25	20
11	20	18
12	23	21
13	22	17
14	18	24
15	24	22
16	25	24
17	21	22
18	22	24
19	23	20
20	25	24

We have compared the different times for each participant in OPWS and VRWS, as shown in Figure 3.15.

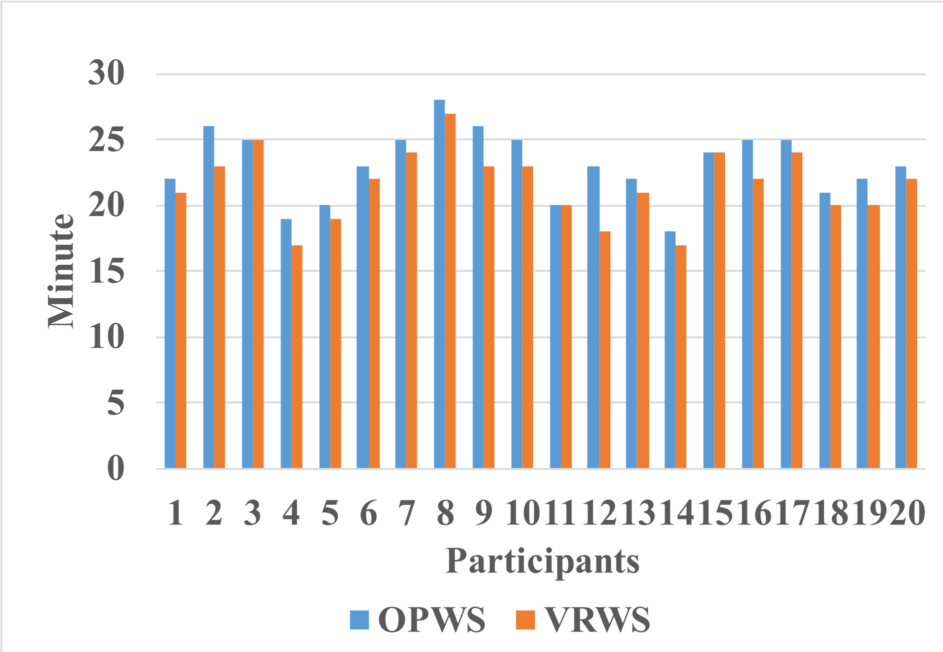


Figure 3.15: Different times for each participant in OPWS and VRWS.

From Figure 3.15, we can see that the line of VRWS is almost all lower than OPWS, but the difference is not apparent. Therefore, it cannot declare VRWS can help participants answer questions faster than OPWS. However, 17 participants spent less time on VRWS. Thus, we probably find significant differences when we get enough experimental data (more questions, more participants).

The questionnaire's results are shown in Table 3.5 and Table 3.6. Moreover, impression evaluation profiles are shown in Figure 3.16 and Figure 3.17.

Table 3.5: The results of OPWS.

	OPWS															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Broad view	Low psycholo	Free atmosph	Comfortable	Well-lighted	Not tired	Natural feeling	Grace	Relaxing	Cheerful	Easy to work	Not noisy inmovem	Enjoyable	Not noisy in sound	Motivated	Efficient
	Narrow view	High psycholo	Non-free atmosph	Uncomfortable	Ill-lighted	Getting tired	Strange feeling	Graceless	Not-relaxing	Depressed	Hard to work	Noisy in movemen	Not enjoyable	Noisy in sound	Unmotivated	Inefficient
1	2	3	4	2	3	3	2	3	3	2	2	6	2	2	3	2
2	1	2	4	4	3	4	3	2	2	4	4	3	1	1	3	3
3	3	4	1	4	6	4	1	3	1	4	4	1	3	3	4	5
4	4	2	3	4	4	3	3	4	4	4	3	7	1	4	6	4
5	1	2	2	2	5	5	3	3	3	3	3	2	3	1	4	3
6	3	2	2	2	6	2	2	3	3	3	4	4	1	3	3	4
7	2	4	6	2	4	5	3	3	3	5	6	6	3	2	6	6
8	3	5	4	4	6	4	2	3	2	4	5	3	2	4	3	1
9	2	1	3	3	6	3	1	4	4	4	3	4	1	6	4	4
10	3	4	4	2	6	4	2	5	2	2	3	6	4	2	2	2
11	3	1	3	1	5	5	2	5	3	5	4	5	4	3	4	4
12	2	2	3	5	6	3	2	4	3	4	2	3	3	2	3	2
13	2	5	1	5	4	5	3	2	4	3	5	3	3	5	1	1
14	4	2	3	6	6	5	4	1	2	4	6	5	3	1	6	5
15	2	3	2	3	6	3	3	6	3	4	4	4	2	3	3	3
16	2	3	1	3	4	5	2	3	3	4	5	6	4	1	4	4
17	3	4	3	5	4	4	1	4	3	4	4	4	3	3	3	5
18	4	2	1	2	6	1	3	2	1	3	5	7	2	1	2	2
19	4	3	4	3	5	5	1	4	4	3	4	5	1	3	4	2
20	4	4	5	5	5	5	3	2	3	5	4	4	4	2	1	4
AVG	2.70	2.90	2.95	3.35	5.00	3.90	2.30	3.30	2.80	3.70	4.00	4.40	2.50	2.60	3.45	3.30
SD	0.95	1.18	1.36	1.35	1.05	1.14	0.84	1.19	0.87	0.84	1.10	1.59	1.07	1.36	1.40	1.38
Var	0.91	1.39	1.85	1.83	1.10	1.29	0.71	1.41	0.76	0.71	1.20	2.54	1.15	1.84	1.95	1.91

Table 3.6: The results of VRWS.

	VRWS															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Broad view	Low psycholo	Free atmosph	Comfortable	Well-lighted	Not tired	Natural feeling	Grace	Relaxing	Cheerful	Easy to work	Not noisy inmovem	Enjoyable	Not noisy in sound	Motivated	Efficient
	Narrow view	High psycholo	Non-free atmosph	Uncomfortable	Ill-lighted	Getting tired	Strange feeling	Graceless	Not-relaxing	Depressed	Hard to work	Noisy in movemen	Not enjoyable	Noisy in sound	Unmotivated	Inefficient
1	5	7	7	5	5	6	5	4	7	5	4	5	7	7	4	5
2	6	6	6	6	7	3	6	3	5	4	6	4	4	5	3	3
3	7	4	7	3	6	3	5	4	6	6	4	3	5	7	4	5
4	5	5	5	4	7	4	4	5	4	5	2	5	6	6	4	4
5	6	5	5	4	6	5	5	2	5	4	5	5	6	7	5	4
6	5	4	6	3	7	4	5	4	6	5	3	3	7	6	4	5
7	6	5	5	5	6	5	4	4	7	4	5	4	5	6	4	3
8	5	5	6	4	4	5	6	3	3	5	2	5	5	6	5	4
9	7	6	7	5	6	4	5	4	5	2	4	5	4	7	5	4
10	2	4	6	4	6	4	5	3	6	3	5	2	6	6	3	2
11	5	6	3	4	5	3	4	6	6	4	6	4	5	6	5	4
12	3	5	5	3	6	2	6	3	4	2	2	4	5	5	4	5
13	6	6	5	4	6	5	6	5	7	5	5	6	7	5	4	4
14	6	4	4	4	6	4	7	6	6	4	4	6	7	6	5	6
15	4	3	5	3	5	5	7	2	5	6	4	5	7	5	5	4
16	5	5	3	5	5	5	5	5	7	6	5	4	5	5	5	5
17	4	5	5	3	6	2	3	3	5	3	3	6	4	6	4	2
18	6	4	4	5	4	3	3	4	4	4	4	3	6	5	2	4
19	4	6	4	4	4	4	4	3	6	4	2	4	5	6	4	4
20	5	6	6	3	5	5	6	4	6	6	2	5	7	6	4	3
AVG	5.10	5.05	5.20	4.05	5.60	4.05	5.05	3.85	5.50	4.35	3.85	4.40	5.65	5.90	4.15	4.00
SD	1.22	0.97	1.17	0.86	0.92	1.07	1.12	1.11	1.12	1.19	1.31	1.07	1.06	0.70	0.79	1.00
Var	1.49	0.95	1.36	0.75	0.84	1.15	1.25	1.23	1.25	1.43	1.73	1.14	1.13	0.49	0.63	1.00

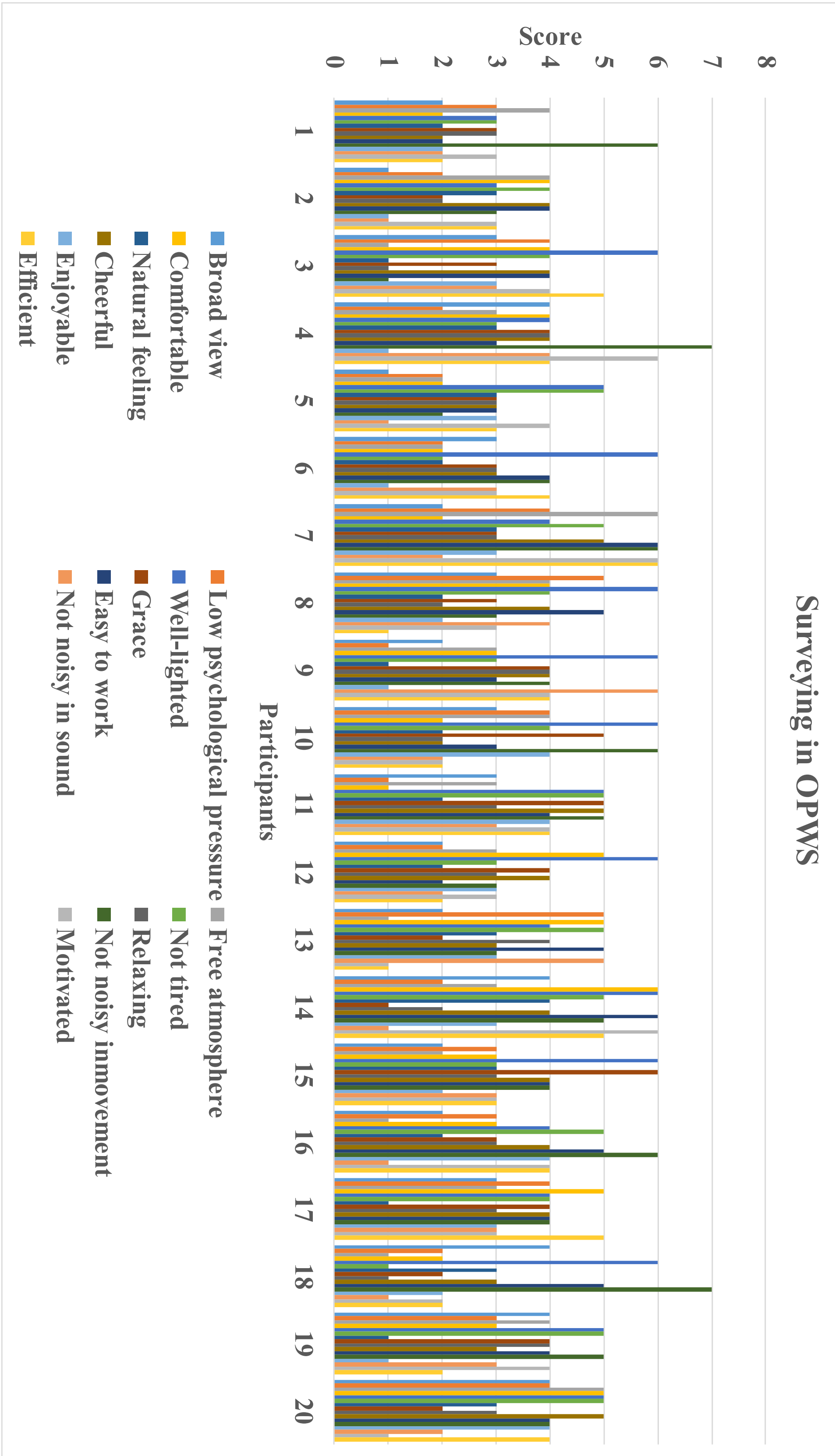


Figure 3.16: Impression evaluation profiles for OPWS

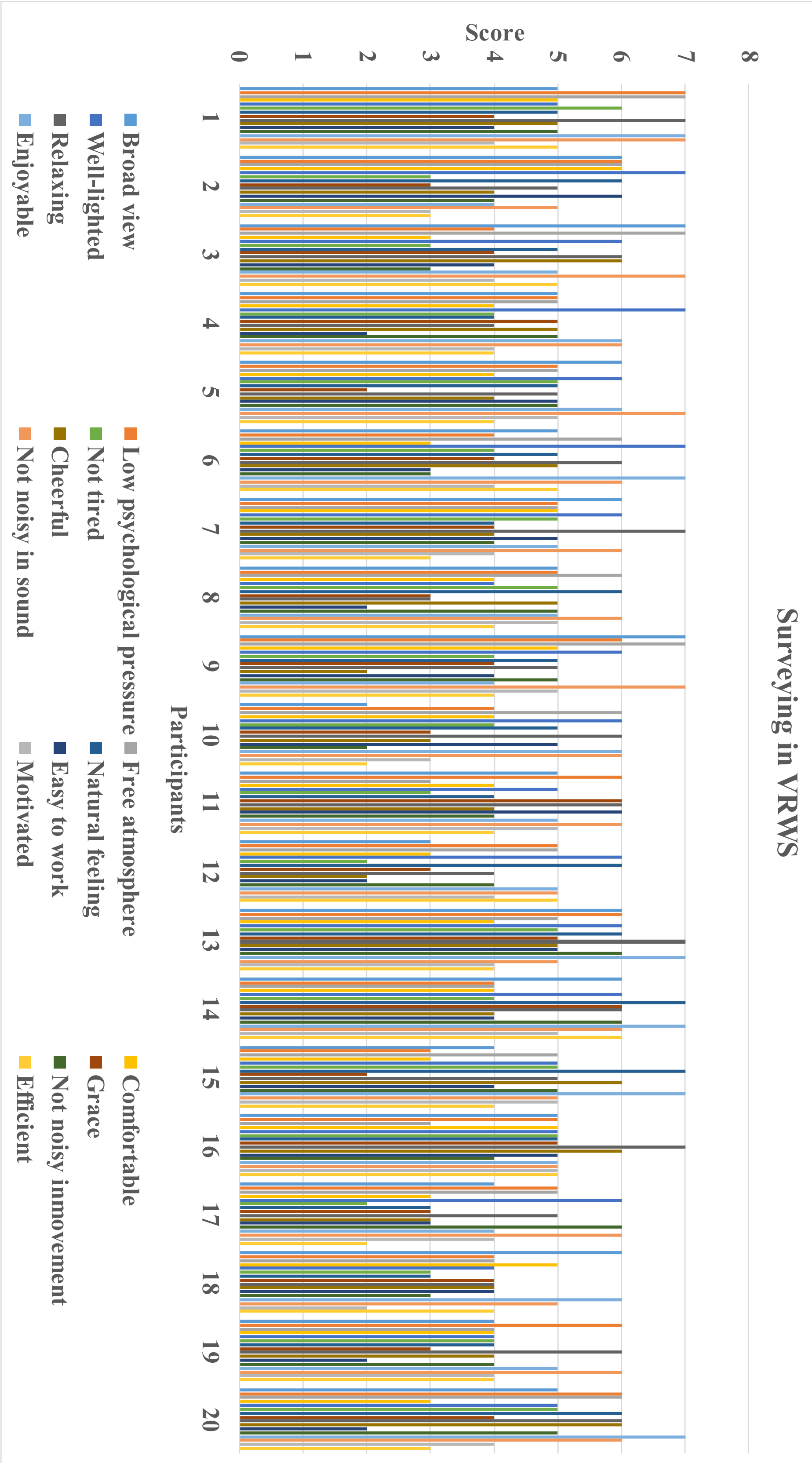


Figure 3.17: Impression evaluation profiles for VRWS

For the questionnaire, the adjective pairs are compared with the average of the two groups' results. As shown in Figure 3.18, lower points are negative evaluations, and higher points are positive evaluations.

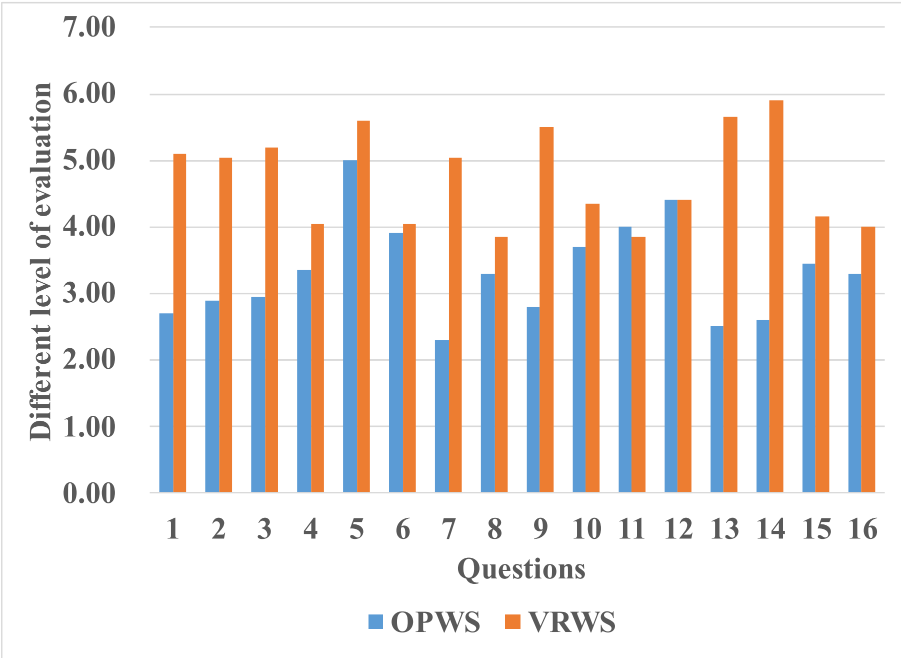


Figure 3.18: The average of the two groups' results.

In order to ensure the validity of this study, a t-test was used to analyze the data further. In this study, SPSSAU was adopted for data analysis. Before performing the t-test, we should confirm the normality of the sample. Because the number of sample data from the CAB test and questionnaire were all less than 50, so the Shapiro-Wilk test was chosen.

Table 3.7: Shapro-Wilk and Kolmogorov-Smirnov.

	Shapro-Wilk	Kolmogorov-Smirnov
Sample data	50 <= a	50 > a

Through the Shapiro-Wilk test, although some sample data were considered to have no normality traits because their P-values were under 0.05. However, their absolute value of Kurtosis are all less than 10, and the absolute values of Skewness are all less than 3. So even some sample data were not the standard normal distribution but basically can accept as a normal distribution. Therefore, all the sample data can be considered to follow the normal distribution. So, we adopted the t-test to analyze the sample data. Normality tests for OPWS and VRWS are shown in Table3.8 and Table 3.9.

Table 3.8: Normality test for OPWS.

Normality test analysis results							
Questionnaire	Sample size	Average	Standard Deviation	Skewness	Kurtosis	Shapro-Wilk	
						Statistic	p
Q1	20	5.3	0.979	0.067	-0.964	0.879	0.017*
Q2	20	5.1	1.21	-0.21	-0.945	0.91	0.063
Q3	20	5.05	1.395	-0.227	-0.281	0.927	0.136
Q4	20	4.65	1.387	-0.214	-0.998	0.925	0.122
Q5	20	3	1.076	0.563	-1.061	0.815	0.001**
Q6	20	4.1	1.165	0.896	0.33	0.844	0.004**
Q7	20	5.7	0.865	0.119	-0.726	0.867	0.010*
Q8	20	4.7	1.218	-0.326	0.085	0.944	0.284
Q9	20	5.2	0.894	0.549	-0.046	0.86	0.008**
Q10	20	4.3	0.865	0.424	-0.105	0.867	0.010*
Q11	20	4	1.124	0	-0.279	0.925	0.126
Q12	20	3.6	1.635	0.164	-0.505	0.954	0.437
Q13	20	5.5	1.1	0.132	-1.259	0.866	0.010*
Q14	20	5.4	1.392	-0.812	0.467	0.897	0.036*
Q15	20	4.55	1.432	-0.291	0.046	0.903	0.048*
Q16	20	4.7	1.418	-0.022	-0.86	0.936	0.205
* p<0.05 ** p<0.01							

Table 3.9: Normality test for VRWS.

Normality test analysis results							
Questionnaire	Sample size	Average	Standard Deviation	Skewness	Kurtosis	Shapro-Wilk	
						Statistic	p
Q1	20	2.9	1.252	0.743	0.754	0.916	0.084
Q2	20	2.95	0.999	0.108	-0.41	0.918	0.091
Q3	20	2.8	1.196	0.221	-0.486	0.922	0.108
Q4	20	3.95	0.887	-0.398	-0.526	0.865	0.010**
Q5	20	2.4	0.94	0.321	-0.577	0.876	0.015*
Q6	20	3.95	1.099	0.372	-0.551	0.908	0.057
Q7	20	2.95	1.146	0.107	-0.474	0.929	0.147
Q8	20	4.15	1.137	-0.323	-0.314	0.92	0.098
Q9	20	2.5	1.147	0.465	-0.399	0.909	0.062
Q10	20	3.65	1.226	0.376	-0.395	0.907	0.057
Q11	20	4.15	1.348	0.128	-1.112	0.891	0.028*
Q12	20	3.6	1.095	0.384	-0.257	0.917	0.089
Q13	20	2.35	1.089	0.021	-1.31	0.86	0.008**
Q14	20	2.1	0.718	-0.152	-0.88	0.812	0.001**
Q15	20	3.85	0.813	0.949	1.184	0.809	0.001**
Q16	20	4	1.026	0.325	0.112	0.906	0.054
* p<0.05 ** p<0.01							

T-test results on the CAB test are shown in Table 3.10, and t-test results on the questionnaire are shown in Table 3.11.

Table 3.10: T-test result on questionnaire

t-test				
Items	Environment (average ± SD)		t	p
	OPWS(N=20)	VRWS(N=20)		
Correct Answer	30.70 ± 3.85	32.25 ± 4.22	-1.214	0.232
Time Difference in Two Experiments	23.10 ± 2.61	21.60 ± 2.66	1.798	0.08

Table 3.11: T-test result on questionnaire

t-test				
Question Number	Environment (average ± SD)		t	p
	OPWS(N=20)	VRWS(N=20)		
Q1	2.70 ± 0.98	5.10 ± 1.25	-6.753	0.000**
Q2	2.90 ± 1.21	5.05 ± 1.00	-6.13	0.000**
Q3	2.95 ± 1.39	5.20 ± 1.20	-5.476	0.00**
Q4	3.35 ± 1.39	4.05 ± 0.89	-1.901	0.066
Q5	5.00 ± 1.08	5.60 ± 0.94	-1.878	0.068
Q6	3.90 ± 1.17	4.05 ± 1.10	-0.419	0.678
Q7	2.30 ± 0.86	5.05 ± 1.15	-8.568	0.000**
Q8	3.30 ± 1.22	3.85 ± 1.14	-1.476	0.148

t-test				
Question Number	Environment (average ± SD)		t	p
	OPWS(N=20)	VRWS(N=20)		
Q9	2.80 ± 0.89	5.50 ± 1.15	-8.301	0.000**
Q10	3.70 ± 0.86	4.35 ± 1.23	-1.938	0.06
Q11	4.00 ± 1.12	3.85 ± 1.35	0.382	0.704
Q12	4.40 ± 1.64	4.40 ± 1.10	0	1
Q13	2.50 ± 1.10	5.65 ± 1.09	-9.098	0.000**
Q14	2.60 ± 1.39	5.90 ± 0.72	-9.424	0.000**
Q15	3.45 ± 1.43	4.15 ± 0.81	-1.901	0.067
Q16	3.30 ± 1.42	4.00 ± 1.03	-1.789	0.082
*p<0.05 **p<0.01				

About the t-test, if the range is outside 5%, it is not a sample from the same population and is a significant difference.

When the confidence interval is set to 95%:

- P > 0.1: non-significant difference;
- 0.05 < P < 0.1: marginally significant difference;
- P < 0.05: significant difference.

Table 3.10 indicates that the Correct Answer (p=0.232) is a non-significant difference, and Time Difference in Two Experiments (p=0.08) is a marginally significant difference.

From Table 3.11 we can find:

- Q1 “Broad view” (p =0.000);
- Q2 “Low psychological” (p =0.000);
- Q3 “Free atmosphere” (p =0.00);
- Q7 “Natural feeling” (p =0.000);
- Q9 “Relaxing” (p =0.000);

- Q13 “Enjoyable” (p =0.000);
- Q14 “Not noisy in sound” (p =0.000).

These p-values are all less than 0.01, so Q1, Q2, Q3, Q7, Q9, Q13, Q14 have significant differences.

- Q4 “Comfortable” (p =0.066);
- Q5 “Well-light” (p =0.068);
- Q10 “Cheerful” (p =0.06);
- Q15 “Motivated” (p =0.067);
- Q16 “Efficient” (p =0.082).

These p-values are between 0.05 and 0.1, so Q4, Q5, Q10, Q15, Q16 are marginally significant differences.

- Q6 “Not tired” (p =0.678);
- Q8 “Grace” (p =0.148);
- Q11 “Easy to work” (p =0.704);
- Q12 “Not noisy in movement” (p =1).

These p-values are all over 0.1, so Q6, Q8, Q11, Q12 are non-significant differences.

3.7 Discussion

3.7.1. Avoiding Visual and Auditory Interferences

The results of Q14 show an effect of sufficient separation of auditory interference by noise-canceling earphones. At the same time, we believe that the no auditory interference environment also has a positive effect on the results of many significant and marginally significant items.

As shown in the results of Q2, Q3, and Q9, compared with the noisy environment of OPWS, the elegant and comfortable virtual environment design and private use features could play a role in preventing psychological pressure.

The participants did not notice the visual interference problem in OPWS from the result of Q12. HMD was a display device wrapped around the eyes of the user, and the user could no longer feel the external visual interference theoretically. Considering that the CAVE system was used to simulate OPWS in the comparative experiment, the busy scene in the noisy OPWS was displayed in 2D by several projection surfaces around the participants in the CAVE system, which might affect the psychological reality of visual interference. Thereby, they reduced the intensity of interference. Furthermore, the contrast effect between OPWS and VRWS in Q12 in the movement was not significant.

3.7.2. Good Lighting and Enough Natural Light

The results of Q1, Q7, Q9, and Q13 indicate the floor-to-ceiling windows significantly improve the participant's vision. The virtual nature environment surrounding the VRWS gave the participants a more natural feeling. Because of the floor-to-ceiling windows, it was easier for natural light through the windows to enter the room.

The CAVE system used in this experiment has good lighting effects, so the participants did not strongly feel the difference in lighting effects between the two experiments from the result of Q5.

3.7.3. Privacy

The result of Q2, Q3, and Q 13 shows significant differences that mean the privacy design of the VRWS provides slight psychological pressure on the workers. Thus, the design lets users feel a free atmosphere and enjoyable work experience. The result from Q4, Q10, Q15, and Q16 indicate marginally significant differences. This kind of privacy design may be comfortable, cheerful, motivational, and efficient for the users. In this case, only Q6 and Q11 have non-significant differences.

3.7.4. Others

A small number of participants could not bear the noisy environment in OPWS. In order to leave as soon as possible, they completed the CAB test at the fastest speed possible while giving the correct answer as much as possible. Therefore, these participants believed that although they could not bear the unbearable interference in OPWS, from the perspective of the results, the work efficiency has improved.

From OPWS to VRWS, although it was more beneficial for participants to answer CAB tests, it was impossible to make difficult questions easier just because the environment changed better, so the Correct Answer was no significant difference.

HMD must be worn when using VRWS. There might be a negative effect on the physical sense, but the impact was not significant from the results of Q4, Q10, Q15, and Q16.

There is no difference between Q15 and Q16 because wearing HMD could be an obstacle to face-to-face communication. As considered by other network communication methods such as e-mail, HMD only caused communication failure in certain situations.

Most of the participants rejected the use of HMD for a long time. The main reasons were the weight and volume of HMD put an extra burden on long-term work, virtual reality might cause vertigo. VRWS did not have sufficient input support and HMD cooling problems. These reasons have led to the results of Q6, Q8, and Q11.

Another noteworthy aspect from our post-experiment discussions with participants is that, while most expressed satisfaction with their overall experience of working in VR, they expressed a negative attitude toward VRWS in terms of efficiency. A key contributing factor was the difficulty of engaging in prolonged and consistent touch typing while wearing an HMD. Since the majority of participants were unable to touch type, they frequently had to remove and re-wear the HMD to verify hand positioning and ensure that the input displayed on the VR screen was accurate. This repetitive process not only disrupted their workflow but also significantly reduced work efficiency and enthusiasm for using VRWS in an office setting.

Therefore, we recognize that for the majority of potential VRWS users who cannot touch type, the inability to type effectively in VR is a critical and urgent issue that needs to be addressed. Based on this understanding, we have decided to focus our subsequent research on how to enable users who cannot touch type to efficiently complete typing tasks in VR, thereby enhancing the usability and user experience of VRWS.

Chapter 4

Low-Jitter Hand Tracking System

An excellent VRWS should not only have a well-designed virtual office environment but also feature a robust typing assistance function. Regarding typing in VR, in the lecture review section, we have summarized the drawbacks of not supporting a physical keyboard and relying on additional auxiliary devices. To thoroughly address the challenges of typing in VR, we have decided to use only the camera mounted on the HMD to capture typing actions and reproduce them in real-time in VR. This approach eliminates the need for additional auxiliary devices and allows users wearing HMD to confirm the position of their typing hands in VR. Figure 4.1 provides an overview of our idea.

4.1 Data Collection

Typing in VR presents unique challenges. If using the HMD's camera to capture the typing hand position, from the perspective of the HMD, the fingers of the typing hand are often obscured by the palm, making it difficult to obtain a complete hand contour. Therefore, accurately capturing the typing position is challenging. As existing hand databases primarily feature complete hand images, there is a need for an "obscured typing hand" dataset to train hand tracking models. Moreover, typing actions are often subtle and delicate, show in Figure 4.2. Training a model using non-targeted datasets makes it hard to trust the model's performance.

A dataset of "obscured typing hand" was required to train the hand tracking model, as existing hand databases predominantly feature complete hand images. Due to the limited availability of such data, we conducted an independent data collection process as follows. Figure 4.3 show de difference between the "obscured typing hand" and other complete hand image.

A total of eleven students from our graduate university participated in the data collection phase, including four females, aged between 25 and 31, and they all possess a certain level of typing skills. The participants were instructed to use a wearable camera while typing on a computer. The camera device, a 4K high-definition camera worn on the ear, was used to capture images of the “obscured typing hands, as shown in Figure 4.4.

We downloaded CNN news from the internet and split the news into sentences of varying lengths. Participants were required to input paragraphs of varying lengths using the QWERTY keyboard based on prompts. The UI is shown in Figure 4.5. We developed a small program to monitor the participants' keypress states, recording the time of keypress events. After the experiment, participants uploaded video footage from a wearable camera. Using the recorded keypress times, we automatically extracted images before and after each keypress event. This approach helps avoid entering invalid content into the database, such as moments of distraction, rest, or contemplation.

Additionally, to ensure that each key on the keyboard has a minimum number of keystrokes, we manually selected certain sentences to control the occurrence frequency of specific letters.

In the experiment, each participant engaged in a one-hour typing session, resulting in a total of 21,900 images collected. Subsequently, following the steps outlined in related research [87], we employed OpenCV to apply image processing techniques for data augmentation. Specifically, we adjusted the hand color and brightness of these images to create variations.

By employing the HSV model, we randomly varied the values of H (Hue) and V (Value brightness). Consequently, we generated a dataset comprising 438,000 images, approximately 20 times larger than the original dataset. To ensure an unbiased evaluation, we randomly partitioned the dataset into training and testing sets, allocating 80% of the images for training and the remaining 20% for testing purposes.

After the experiment, manual bounding box annotation of the data was performed by human annotators using media pipe as an assistive tool. This was done to extract useful portions of images featuring the typing hand from the wide-angle wearable camera. This is shown in Figure 4.6.

Table 4.1: Participant Information.

Participant	Gender	Age.	Typing proficiency (self-evaluation)
1	Female	29	Average
2	Female	27	Touch typing
3	Male	26	Proficient
4	Male	25	Proficient
5	Female	28	Proficient
6	Male	27	Proficient
7	Male	27	Proficient
8	Male	29	Proficient
9	Female	31	Touch typing
10	Male	31	Proficient
11	Male	25	Rusty
Skill: Poor<Rusty<Average<Proficient<Touch typing			

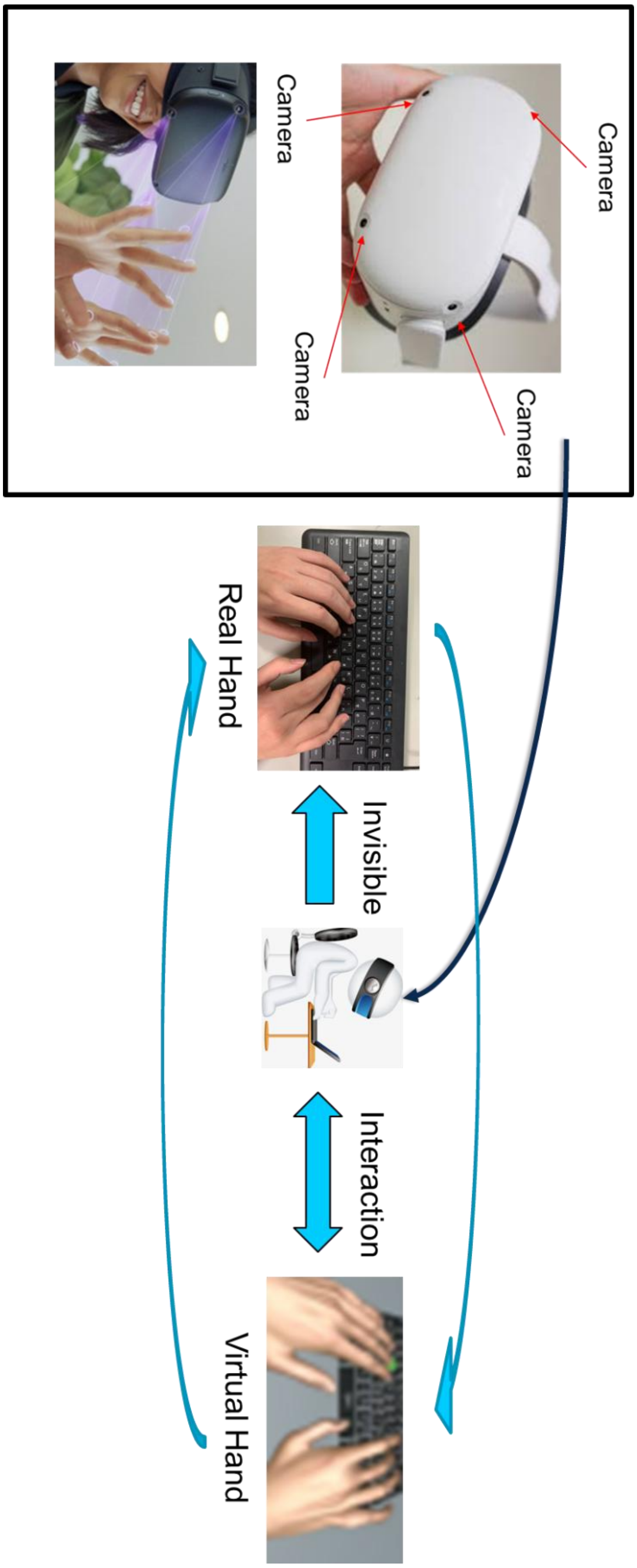


Figure 4.1: Use camera which is on HMD wo tracking typing hand, show the hand position in VR in real time.

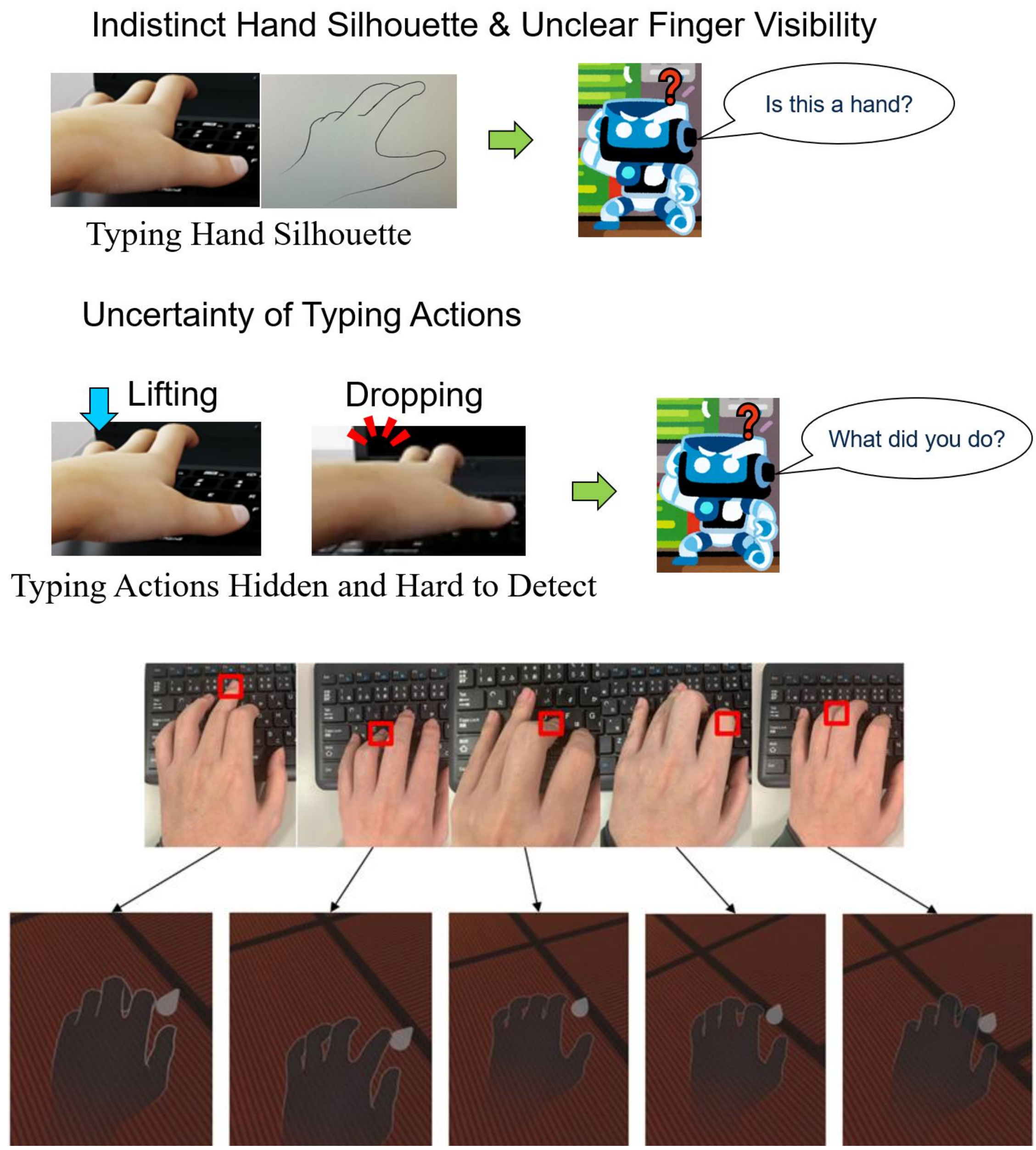


Figure 4.2: The typing actions are very subtle, making them challenging to detect. It is also difficult to predict the position of the typing hand through subtle changes in the contour. In the examples shown in the lower part of the figure, it is evident that even with different typing positions, there is not a significant difference in the position of the VR hands.

Figure 4.3: The comparison shows the difference between "obscured typing hand" and other complete hand image. The left one is collected from the perspective of HMD, right one is from KBH [88] dataset. The bottom one in the middle is MSU. [84]

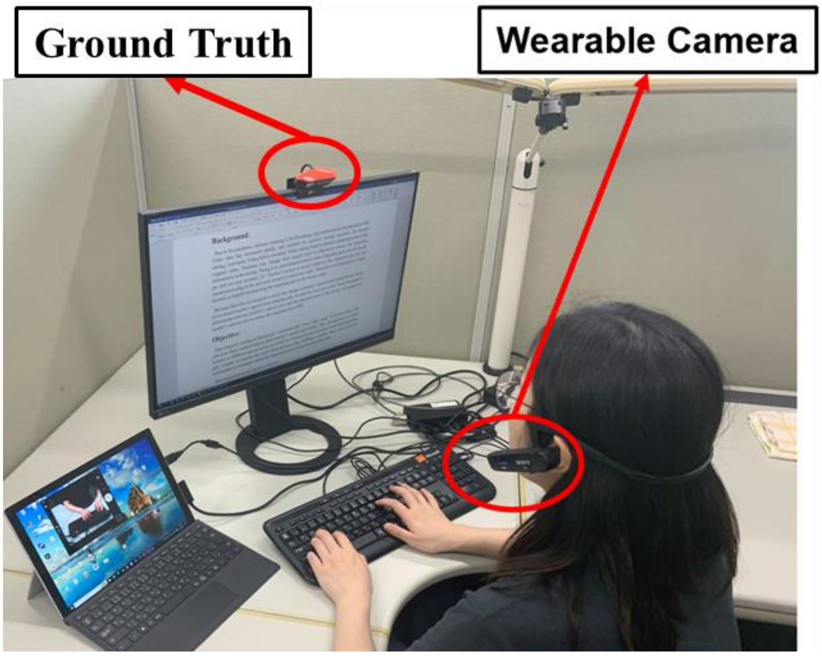


Figure 4.4: Typing scene with a wearing camera.

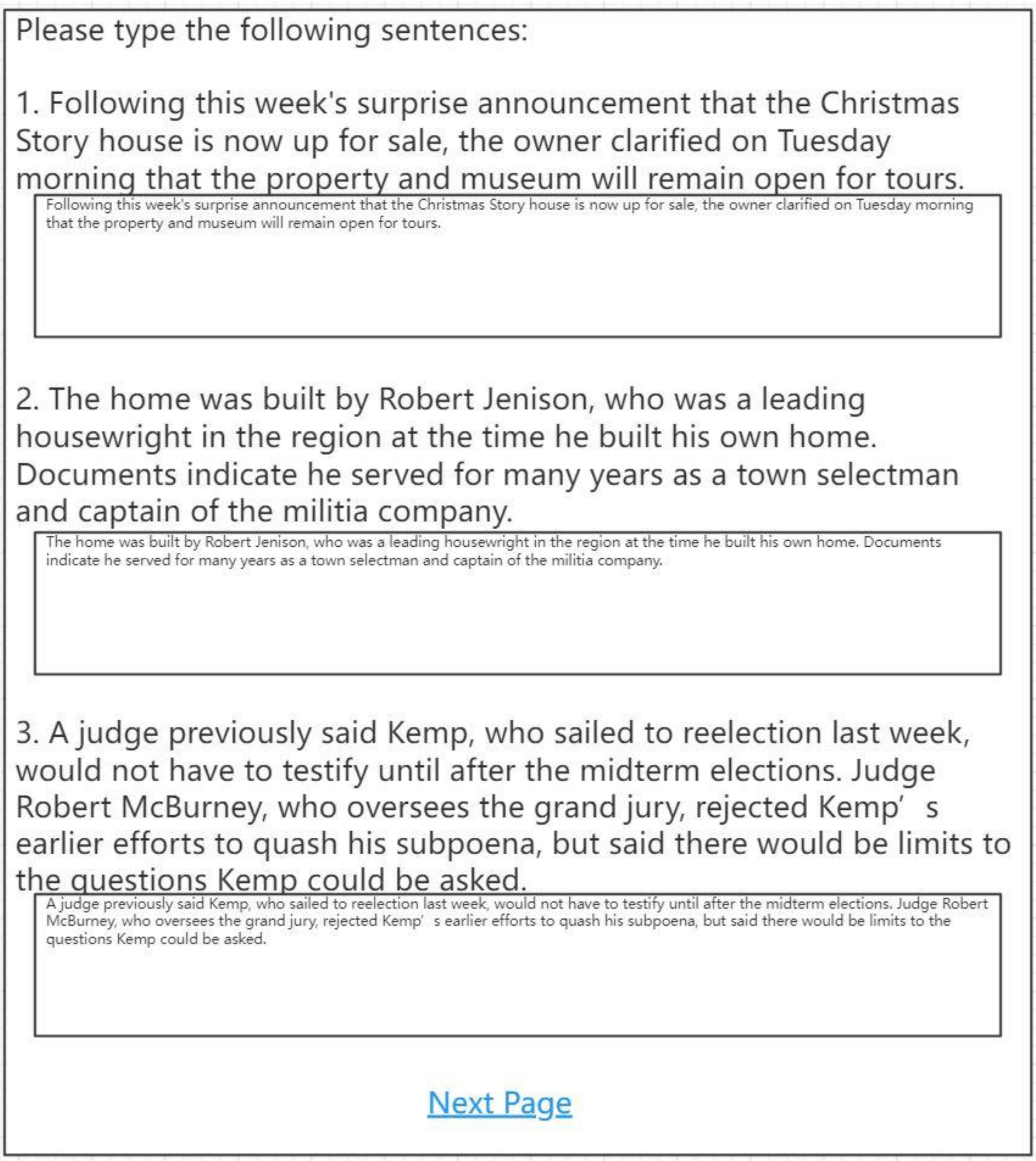


Figure 4.5: Example of typing task.

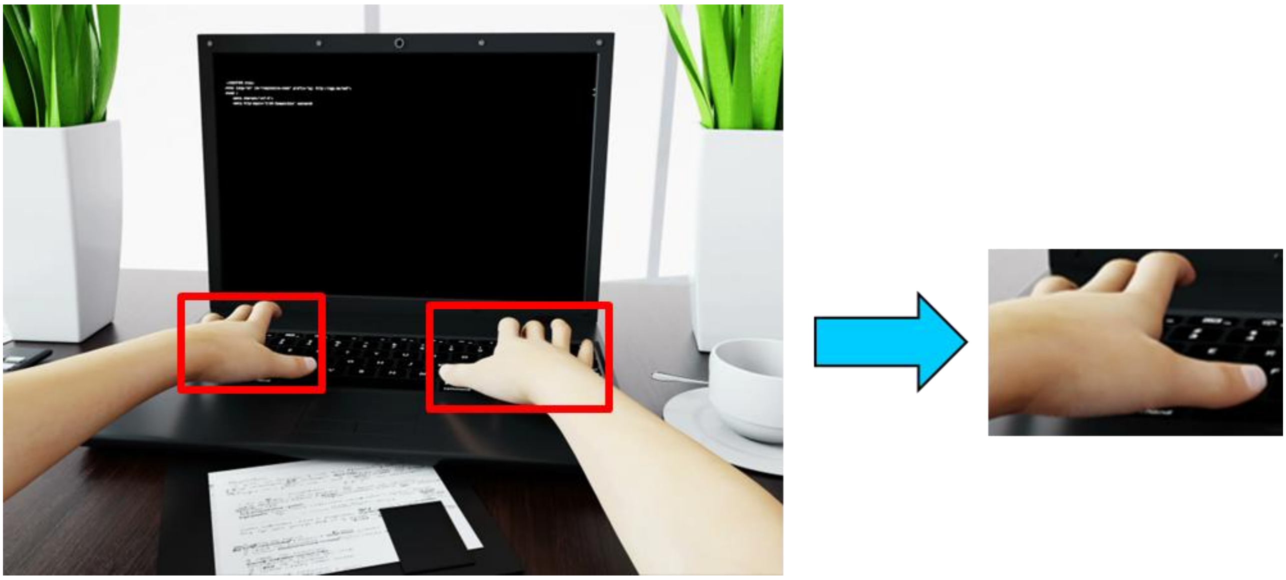


Figure 4.6: Bounding box annotation and cropping of the data were conducted.

4.2 Motion History Image

Motion History Image (MHI) is a valuable concept in computer vision, specifically designed for capturing and representing temporal information in video sequences [89]. It plays a crucial role in motion analysis, allowing for the extraction of meaningful patterns related to object movements over time. MHI is a chronological representation of motion in a sequence of images, emphasizing the recency of pixel changes. It assigns higher pixel values to regions where motion has occurred more recently, creating a visual representation of the temporal evolution of movement within a video. MHI is a chronological representation of motion in a sequence of images, emphasizing the recency of pixel changes. It assigns higher pixel values to regions where motion has occurred more recently, creating a visual representation of the temporal evolution of movement within a video. The formula as follow [89]:

$$H_\tau(x,\quad y,\quad t) = \begin{cases} \tau & \text{if } \Psi\ (x,y,t) = 1 \\ \max(0, H_\tau(x,y,t-1) - \delta) & otherwise \end{cases} \tag{4.1}$$

In the formula, (x, y) and t represent the pixel's position and time, respectively. τ represents the duration, determining the temporal scope of the motion from the frame perspective. δ is the decay parameter. *Ψ (x, y, t)* is the updating function, which can be defined by frame difference:

$$\psi(x,y,t) = \begin{cases} 1 & if\ D(x,y,t) \geq \xi \\ 0 & otherwise \end{cases} \tag{4.2}$$

Where

$$D(x,y,t) = \ |I(x,y,t) - I(x,y,t \pm \Delta)| \tag{4.3}$$

Here, *I (x, y, t)* is the intensity value of the pixel at coordinates (x, y) in the video image sequence at frame *t*, delta is the frame interval, and ξ is a manually set difference threshold adjusted with changes in the video scene.

Building upon this foundation, a more advanced approach involves using optical flow to define *Ψ (x, y, t)* [90].

$$E(x,y,t) = \ s\ (x,y,t) + E\ (x,y,t-1) \cdot \alpha \tag{4.4}$$

where s (*x, y, t*) denotes the optical flow length corresponding to pixel (*x, y*) at time frame *t*.

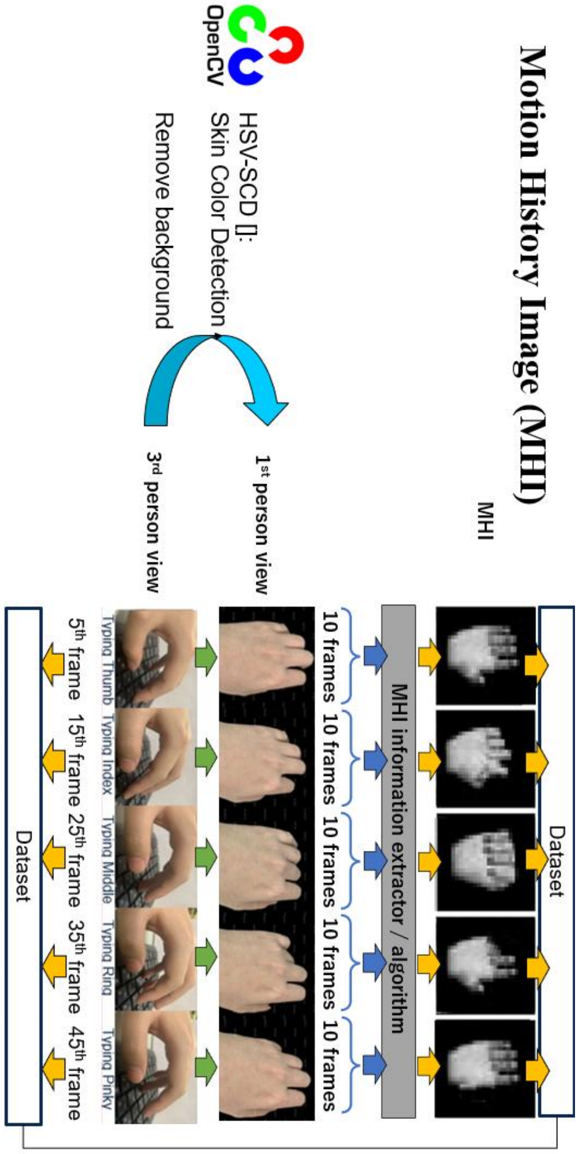


Figure 4.7: Optical flow based MHI.

4.3 Network Architecture

The study explores the factors influencing typing efficiency in VRWS. During this investigation, specific typing behavior characteristics, such as fingers being blocked by the palm or subtle typing finger actions, will be thoroughly examined. By leveraging or addressing these characteristics, we propose the following network architecture.

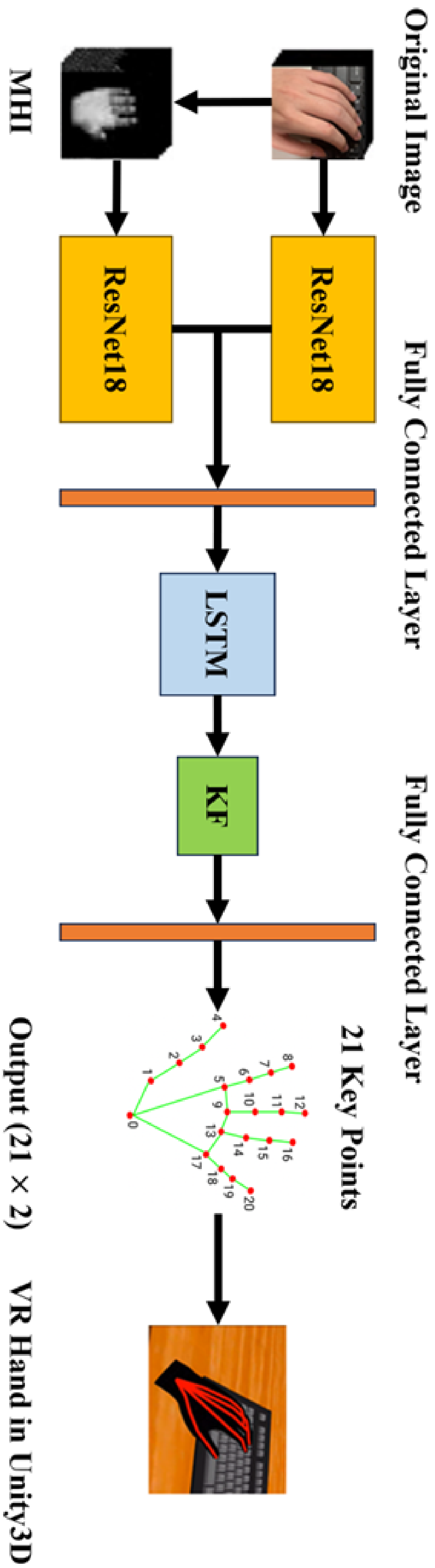


Figure. 4.8: The overview of 2S-LSTM network.

4.3.1. Two Stream ResNet18

Before delving into ResNet, it is essential to understand its predecessor, VGG. VGG standing as the precursor to ResNet, characterized by its simplicity, utilizing small-sized convolutional kernels with a focus on deeper networks, specifically the widely recognized VGG16 and VGG19 configurations [91].

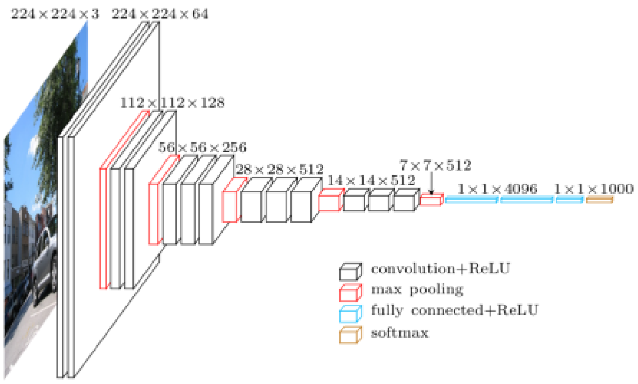


Figure. 4.9: VGG16 architecture [91].

ResNet is a convolutional neural network (CNN) architecture, built upon the foundation laid by VGG while introducing innovative residual connection structures [92]. In addition to these advancements, ResNet retains classical features inherited from VGG, such as the use of small convolutional kernels. ResNet takes this a step further by addressing the challenge of training very deep networks. The introduction of residual connections allows the model to bypass certain layers, mitigating the vanishing gradient problem and enabling the successful training of networks with an unprecedented depth.

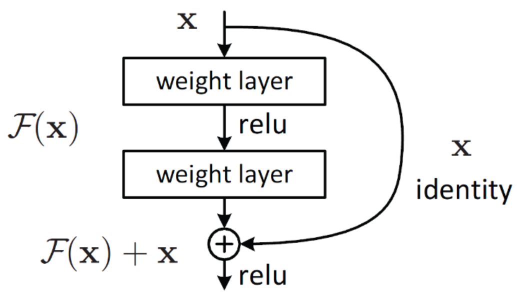


Figure. 4.10: Residual Connection [92].

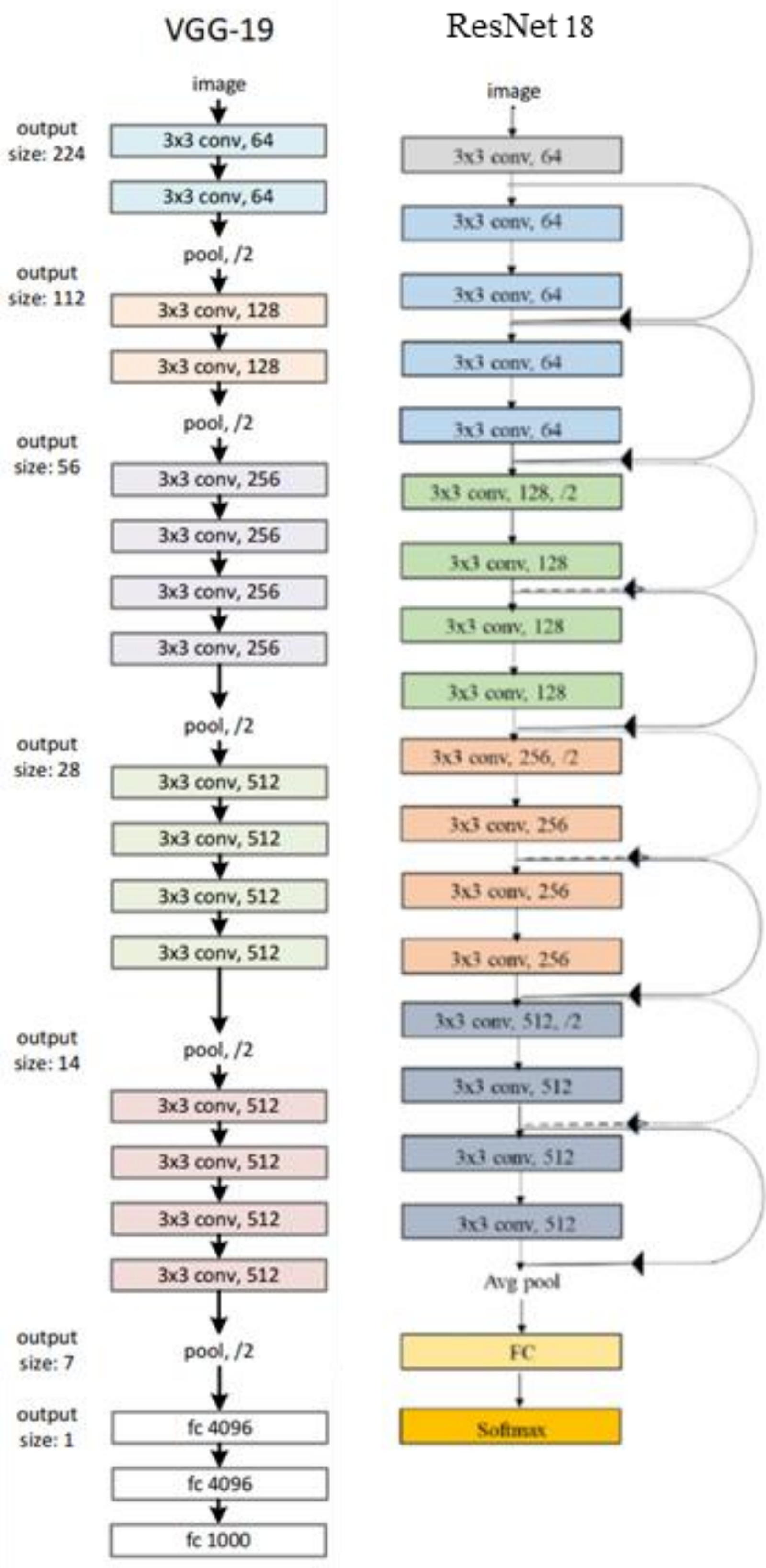


Figure. 4.11: ResNet18 compared with VGG19 [91,92].

ResNet 18, a variant of ResNet that stands out for its smaller size compared to its counterparts. ResNet 18 is particularly well-suited for deployment in environments with resource constraints, such as HMD. The advantages of ResNet18 include its relatively compact architecture while retaining the benefits of the residual connections. This smaller size ensures that deploying ResNet18 on an HMD does not introduce significant latency, making it an optimal choice for real time applications.

In this research, the training sequence of length τ is 10. For each τ, we use the hand position labels $y_{1:\tau}$ and two input streams: original image $I_{1:\tau}$ and MHI $X_{1:\tau}$, are separately processed through a ResNet18 network to extract visual features. Subsequently, a fully connected layer is used to combine two visual features into a unified visual feature ϕ. Following this, the visual feature sequence $\phi_{1:\tau}$ is fed into an LSTM layer to extract temporal feature sequence $\psi_{1:\tau}$.

4.3.2. LSTM

Long Short-Term Memory (LSTM) is a specialized recurrent neural network (RNN) architecture designed to address challenges in capturing long-term dependencies within sequential data [93]. Unlike traditional RNNs, LSTM introduces a memory cell equipped with gating mechanisms, allowing it to selectively store, forget, and update information over extended sequences. This design overcomes issues like vanishing and exploding gradients, making LSTM particularly effective for tasks involving sequential data analysis. With advantages such as the ability to maintain context over extended periods and selective information retention, LSTM has become a cornerstone in diverse applications, including natural language processing and time series prediction. The architecture's key features include memory cells, gating mechanisms, and hidden states, governed by mathematical formulations involving input gates, forget gates, cell states, output gates, and hidden states. These equations, characterized by weight matrices, biases, and activation functions, enable LSTM to excel in capturing intricate temporal patterns, making it a pivotal technology in the realm of deep learning.

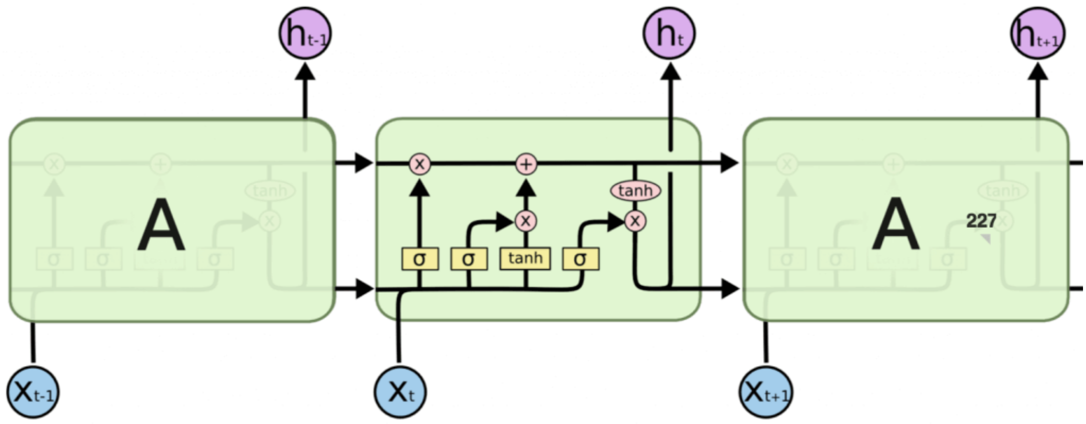


Figure. 4.12: LSTM architecture [93].

Attention mechanisms play a crucial role in enhancing the capability of neural networks to focus on relevant information while processing input sequences. One notable implementation of attention is the Multi-Head Attention mechanism, a key component in transformer architectures. The attention mechanism allows the model to assign varying degrees of importance to different parts of the input sequence, dynamically adjusting the focus during processing.

In the Multi-Head Attention mechanism, the input is transformed by multiple sets of linear projections, each referred to as a "head." These heads operate in parallel, capturing different aspects and relationships within the data. The outputs from all heads are then concatenated and linearly transformed to produce the final attention output. This parallelization enables the model to capture various features simultaneously, promoting richer and more nuanced representations.

The attention mechanism itself involves calculating attention scores for each element in the input sequence concerning other elements. These scores are determined through a combination of query, key, and value transformations. The attention weight for each element signifies its relevance to others, allowing the model to weigh the contributions of different parts of the sequence dynamically.

Given the distinctive characteristics of LSTM, we opt to employ LSTM to establish a connection with the Two Stream ResNet18, aiming to extract temporal feature sequence $\psi_{1:\tau}$.

4.3.3. Kalman Filter

Kalman filtering is a recursive algorithm designed for estimating the state of a system [94]. This filtering method excels in dealing with dynamic systems characterized by uncertainties and measurement noise. One of its notable advantages is the ability to provide accurate estimates of the system state by fusing information from both the system model and actual measurements. The fundamental idea behind Kalman filtering is to iteratively update the estimate of the system state, considering prior knowledge and real-time measurement data, thus yielding a more reliable state estimation.

One key feature of Kalman filtering is its applicability to the continuous monitoring of dynamic systems. Its scope spans various fields, including navigation, control systems, and signal processing. The algorithm's success lies in its ability to dynamically adapt and refine state estimates based on incoming information, making it a versatile tool for real-world applications.

Kalman filter is a mathematical technique that can estimate the state of a dynamic system from noisy measurements. Kalman filter has two steps: prediction and update. In the prediction step, the filter uses a motion model to predict the next state based on the previous state and the control input. In the update step, the filter uses a measurement model to correct the prediction based on the observation and the measurement noise.

The combination of LSTM and Kalman filter [95] can be used for pose regularization, state estimation and traffic flow forecasting. The idea is to use LSTM to learn a rich and dynamic representation of the motion and noise models from data, and then use Kalman filter to recursively update the state based on the LSTM output. This way, the LSTM-Kalman filter can capture complex and nonlinear dynamics that are difficult to model explicitly.

For example, in pose estimation tasks such as body joint localization, camera pose estimation and object tracking, one-shot methods are generally noisy and temporal filters are needed for regularization. However, traditional Kalman filters require specifying a priori motion and measurement models that are often crude approximations of reality. By using LSTM to learn these models from data, the LSTM-Kalman filter can achieve state-of-the-art performance .

LSTM-KF integration capitalizes on the strengths of both Kalman filtering, which excels in handling uncertainties and noise, and LSTM, renowned for capturing temporal dependencies in sequential data.In conclusion, the combination of Kalman filtering and LSTM holds significant potential to reduce jitter in virtual reality systems. Given that typing behavior is a continuous and linear process, the introduction of Kalman filtering is expected not only to minimize jitter but also to enhance the accuracy of recognizing the position of the typing hands.

The KF serves to stabilize the sequence of features extracted by the network, enhancing the accuracy and robustness of hand position estimation, especially in the presence of occlusions and complex backgrounds. Then, the output is passed through another fully connected layer. This step serves to map the temporal feature to the estimated position of the typing hands$\tilde{y}_{1:T} = f(I_{1:\tau}, X_{1:\tau})$.

4.3.4. Key Point

We referred the design of BlazePalm [86], each hand position label includes 42 key points (21 key points in one hand). The key points show in Figure. 4.8 and Figure 4.13.

To visualize $\tilde{y}_{1:\tau}$, we implemented a hand simulator using Unity3D. This simulator can map $\tilde{y}$ to a both hand model consisting of 42 key points. By associating these key points with $\tilde{y}$, we are able to dynamically reproduce and simulate the movements and positions of typing hands in real time.

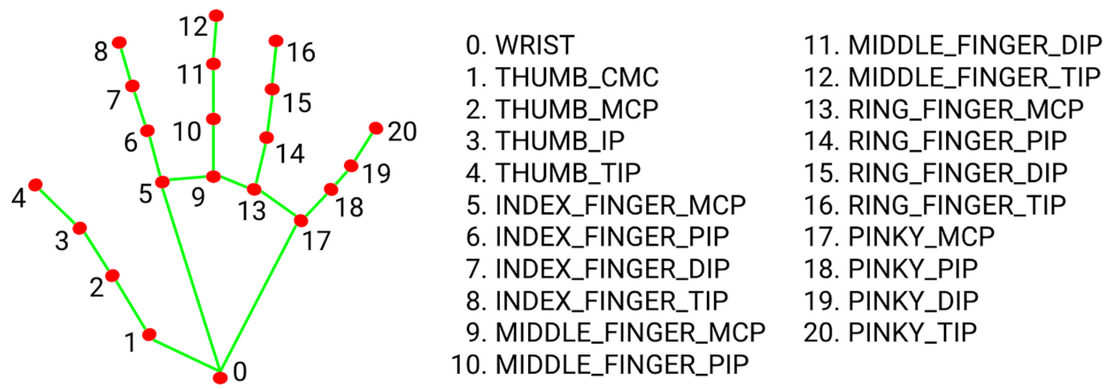


Figure 4.13: 21 key points for one hand [61].

Chapter 5

Performance

This chapter evaluates the performance of the proposed network model. It consists of two sections: Ablation Study and Performance Comparison.

5.1 Ablation Study

While the structural framework of 2S-LSTM has been elucidated, the affirmative impact of the newly proposed modules in 2S-LSTM remains uncertain. Therefore, we aim to ascertain the efficacy of 2S-LSTM through this Ablation Study.

We systematically removed each proposed structural element from the architecture, starting from the 2S-LSTM network and progressing through VGG16 [91] and ResNet18 [92]. The comparative analysis evaluated the two-stream LSTM-KF network against the standard LSTM network. The considered architectures are: 2S ResNet18+LSTM+KF (RGB, MHI), 2S ResNet18+LSTM (RGB, MHI), ResNet18 + LSTM (RGB), ResNet18 (RGB), and VGG16 (RGB). Here, "2S" denotes 2-stream, and "Ours" represents 2S ResNet18+LSTM+KF (RGB, MHI). The evaluation utilized our dataset, and the "accuracy of hand positions" for each model was calculated.

Specifically, for the "accuracy of hand positions" calculation, considering the size of the letter key on the keyboard as approximately 1.5cm × 1.5cm, typing errors were assumed to occur when the distance between the fingertip and the center of the target key on the keyboard exceeded half the length of a key (0.75cm). Thus, the threshold was set at 0.75cm, considering a key point as valid if the distance to the ground truth was less than the threshold. The "accuracy of hand positions" is determined by the total number of valid key points divided by the total number of key points. The results indicate that the 2S-LSTM-KF architecture outperforms the others.

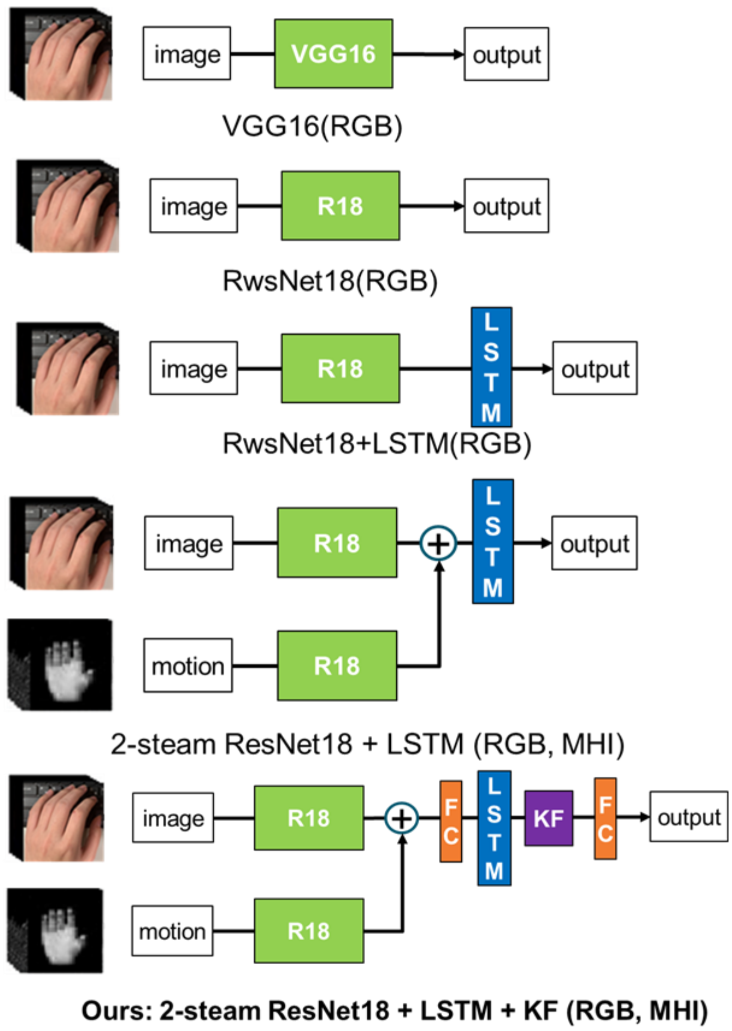


Figure. 5.1: Comparative architectures in ablation study.

5.2 Performance Comparison

To identify the optimal network framework for VR typing tasks, we compared multiple models, focusing on latency, accuracy, and jitter. This comparison aimed to determine which model provides good overall performance.

5.2.1. Participants and Equipment

Latency, accuracy, and jitter are influenced primarily by the performance of hardware and algorithms. Therefore, we standardized the hardware across all conditions, using the

HTC VIVE Pro paired with our developed VR typing interface. The only variable across conditions was the hand-tracking model employed. To gain insights into real-world user experience, we recruited three participants (two males and one female) with normal vision for the comparison.

Latency data were captured using VRScore [96], a widely used VR performance assessment tool. Accuracy was evaluated using the test set from the internal dataset described in Section 3. Jitter was quantified by comparing the positions of real and virtual hands.

5.2.2. Comparison Conditions

We tested the following models in the VR typing environment, where 2S denotes a 2-stream network architecture, and KF represents Kalman filtering:

Condition 1: HTC VIVE Pro built-in gesture detection;

Condition 2: TSSequencer [97];

Condition 3: 2S-TSSequencer;

Condition 4: 2S-TSSequencer-KF;

Condition 5: PatchTST [98];

Condition 6: 2S-PatchTST;

Condition 7: BNNActionNet [99];

Condition 8: 2S-BNNActionNet;

Condition 9: 2S-BNNActionNet-KF;

Condition 10: LSTM [100];

Condition 11: 2S-LSTM;

Condition 12: 2S-LSTM-KF (Ours).

Each condition differed only in the model used, with all other factors, such as refresh rate, kept consistent to ensure that performance differences were attributed solely to the models.

5.2.3. Metrics and Data Collection

For an optimal VR typing experience, latency, accuracy, and jitter are all crucial evaluation metrics. We chose to collect data on all three metrics across the conditions to make a comprehensive assessment and identify the best-performing model.

Latency:

Latency is an important evaluation metric, as high latency can induce motion sickness in users [101]. While an ideal latency is below 20 ms [102], most VR systems struggle to maintain stability within this range due to various factors like graphical rendering, signal transmission, and computational load. Individual sensitivity to latency varies, with some users perceiving delays as short as 3–4 ms [103]. We recorded the minimum, maximum, and average latency over 10 min intervals for each model. Participants provided feedback on their perceived latency and were allowed to switch between conditions for better comparison.

Accuracy:

The accuracy for each model was measured using the test set from our internal dataset after training with the training set. This allowed us to evaluate each model's effectiveness in accurately recognizing hand movements during the typing task.

Jitter:

Jitter was evaluated as the stability of hand positions by measuring discrepancies between real and virtual hand positions at 21×2 key points. Points with a discrepancy exceeding a threshold were counted as contributing to jitter, while points below this threshold were not. The threshold value was established based on criteria published in our prior work at TENCON2023 [104].

5.3 Result

5.3.1. Result of Ablation Study

Table 5.1 presents the comparative results of the ablation study. The analysis revealed that the full 2S ResNet18+LSTM+KF (RGB, MHI) architecture achieved the highest accuracy, outperforming all other configurations. Specifically, the 2S structure contributed to capturing the temporal and spatial features of typing gestures effectively, while the Kalman filter reduced jitter and enhanced tracking stability. In contrast, models without the 2S structure or KF exhibited a noticeable decline in performance, highlighting their critical roles in achieving high accuracy. The results confirm that the proposed 2S-LSTM-KF architecture is the most effective for accurately tracking hand positions in typing tasks.

Table 5.1 Result of ablation study.

Architecture	Accuracy of hand positions
VGG16 (RGB)	0.46
ResNet18 (RGB)	0.48
ResNet18 + LSTM (RGB)	0.53
2-steam ResNet18 + LSTM (RGB, MHI)	0.79
Proposed	0.86

5.3.2. Result of Performance Comparison

The latency measurements for different conditions are summarized in Table 5.2 below. The table presents the minimum, maximum, and average latency values derived from the total 10 min of latency data collected for each condition.

Table 5.2. Result of Performance Comparison.

Condition	Latency (10 min)			Accuracy (%)	Jitter (Number of Point)
	Min.	Max.	Avg.		
HTC VIVE Pro built-in gesture detection	39 ms	73 ms	48 ms	65.05%	3565
TSSequencer [97]	41 ms	83 ms	62 ms	80.25%	2687
2S-TSSequencer	45 ms	107 ms	61 ms	78.80%	2606
2S-TSSequencer-KF	45 ms	111 ms	62 ms	80.45%	2049
PatchTST [98]	117 ms	250 ms	201 ms	83.73%	2389
2S-PatchTST	151 ms	297 ms	274 ms	83.19%	2710

BNNActionNet [99]	29 ms	75 ms	51 ms	77.47%	2194
2S-BNNActionNet	29 ms	91 ms	57 ms	80.81%	2124
2S-BNNActionNet-KF	30 ms	105 ms	61 ms	81.15%	1989
LSTM [35]	41 ms	77 ms	52 ms	69.75%	3134
2S-LSTM	44 ms	99 ms	58 ms	77.00%	3111
Ours	44 ms	112 ms	59 ms	78.88%	1974

Concerning latency, as shown in Table 5.2, most models demonstrated acceptable latency compared to the baseline (HTC VIVE Pro built-in gesture detection), with only PatchTST and 2S-PatchTST showing significantly higher latency. This increased latency may lead to user discomfort, such as dizziness, making these models less suitable for VR typing tasks. Participants reported feeling very uncomfortable and restless after using PatchTST and 2S-PatchTST for a period of time, which differed from their experiences in other conditions.

Regarding accuracy, Table 5.2 indicates that all models, except LSTM, achieved respectable accuracy. Excluding the high-latency PatchTST and 2S-PatchTST, the highest accuracy was observed with 2S-BNNActionNet-KF, which outperformed our proposed model by 2.27%. However, this difference was not substantial enough to noticeably affect typing performance, as participant feedback confirmed that users could not perceive a clear difference in accuracy among the top-performing models.

In terms of jitter, as shown in Table 5.2, comparing 2S-TSSequencer with 2S-TSSequencer-KF, 2S-BNNActionNet with 2S-BNNActionNet-KF, and 2S-LSTM with 2S-LSTM-KF (Ours), it is evident that the models with KF exhibit smaller jitter values com-pared to their non-KF counterparts. Additionally, participants reported being generally satisfied with the jitter performance of the conditions which have KF.

5.4 Discussion

5.4.1. Discussion on Ablation Study

The findings from the ablation study underscore the importance of integrating the two-stream structure and Kalman filter in the proposed 2S-LSTM architecture. The 2S structure, leveraging both RGB and MHI, provided complementary data streams that enhanced the model's ability to capture subtle typing gestures. By including MHI, the temporal dynamics of typing actions were preserved, enabling better prediction of finger movements compared to single-stream configurations.

The Kalman filter also played a pivotal role by mitigating jitter, which is particularly critical in VR typing tasks. Typing involves highly repetitive, small, and precise movements, which are prone to being smoothed out excessively by conventional filtering techniques like low-pass filters. Predictive filters, on the other hand, struggled with high similarity among successive typing movements, leading to prediction errors. The Kalman filter, with its adaptive smoothing and noise reduction capabilities, effectively balanced motion fidelity and tracking stability, making it the optimal choice for this application.

The backbone networks, ResNet18 and VGG16, also demonstrated varying impacts on accuracy. While VGG16 provided a baseline for feature extraction, ResNet18 significantly improved performance due to its ability to capture deeper and more representative features. However, the two-stream approach amplified this advantage further by integrating RGB and MHI data, which are essential for modeling typing movements.

Overall, this ablation study confirms that the 2S ResNet18+LSTM+KF (RGB, MHI) architecture is uniquely suited for VR typing tasks, providing the highest accuracy for hand position tracking while addressing the unique challenges of typing in immersive environments. These results establish a strong foundation for further refinement and application of this architecture in VRWS.

5.4.2. Discussion on Performance Comparison

The latency results showed that while the vast majority of models (except for PatchTST and 2S-PatchTST) exhibited slightly higher latency than the baseline condition (Condition 1), this increase of a few milliseconds to over ten milliseconds remained within an acceptable range. Participant feedback indicated that the slight increase in latency brought by these models was imperceptible compared to Condition 1. Consequently, due to excessive latency, both PatchTST and 2S-PatchTST can be excluded from consideration, and we believe that the computational heaviness of PatchTST, which is based on the Transformer architecture, is a key factor contributing to its significant latency issues.

In terms of accuracy, 2S-BNNActionNet-KF emerged as the top performer, while the 2S-LSTM-KF model trailed by 2.27%. The 2S-TSSequencer-KF also performed admirably, leading 2S-LSTM-KF by just 1.57%. Given the nature of typing actions, which involve subtle movements and rapid finger lifts, the task of identifying typing fingers may not necessitate complex long-range dependency modeling, thereby limiting the advantages of the TSSequencer. Furthermore, the TSSequencer model might require larger and higher-quality datasets to fully realize its strengths. However, the dataset used in this study was self-made under limited conditions and funding, potentially constraining the performance of the TSSequencer. The results show that 2S-BNNActionNet-KF is a promising solution, especially in terms of accuracy. However, LSTM performed slightly better in terms of latency and jitter. Some previous research reported that BNNActionNet has the advantage with lower computing resources, but that LSTM achieves higher accuracy, especially in applications that require capturing subtle temporal variations [105]. As the computing re-sources of new HMDs improve in the future, these results may change.

Jitter analysis showed that 2S-LSTM-KF performed the best, followed by 2S-BNNActionNet-KF and 2S-TSSequencer-KF, which also demonstrated solid results. When comparing models with and without KF, the KF-enhanced versions consistently showed improved jitter performance. This suggests that incorporating KF benefits jitter reduction not only in 2S-LSTM-KF but across other models as well.

After considering latency, accuracy, and jitter performance, we believe that both 2S-BNNActionNet-KF and 2S-LSTM-KF are optimal choices. Given that 2S-BNNActionNet does not significantly outperform 2S-LSTM-KF across all metrics and considering the author's extensive experience in deploying LSTM on VR devices, we have decided to use 2S-LSTM-KF for this experiment. In our future work, we will further explore and investigate the potential applications of 2S-BNNActionNet.

Chapter 6

Typing Experiment

A comparative experiment was conducted to assess the developed assistance solution (2S-LSTM) in comparison to two existing solutions, Oculus Quest 2 and Leap Motion. The primary objective was to validate the effectiveness of the proposed method in enhancing typing efficiency. All experiments conducted in this study received approval from the JAIST Life Sciences Committee (H04-032).

6.1 Experiment Design

6.1.1. Participants

A total of 24 participants were recruited, comprising 23 right-handed individuals and 1 left-handed individual (16 males and 8 females, with an average age M=26), all with normal or corrected-to-normal vision. Among the participants, 7 had prior VR experience. We balanced the 6 participant groups by gender and experimental order. All participants demonstrated a certain level of English proficiency and were not required to possess advanced touch-typing skills.

Table 6.1: Participant Information.

Participant	Gender	Age.	English proficiency
1	Female	26	Fluent
2	Male	25	Native-level proficiency

3	Male	26	Fluent
4	Male	25	Fluent
5	Male	26	TOEIC Score: Exceeds 800 points
6	Male	29	Native English Speaker
7	Male	26	Native-level proficiency
8	Male	25	Fluent
9	Female	26	TOEIC Score: Exceeds 500 points
10	Male	26	Fluent
11	Male	25	TOEIC Score: Exceeds 900 points
13	Female	26	Native English Speaker
14	Female	27	Fluent
15	Male	26	TOEIC Score: Exceeds 500 points
16	Female	26	Fluent
17	Male	26	Native-level proficiency
18	Male	27	Native-level proficiency
19	Female	26	Fluent
20	Male	25	Fluent
21	Male	26	Fluent
22	Male	26	Fluent
23	Female	26	Native-level proficiency
24	Male	26	Fluent

6.1.2. Equipment

The experiment was conducted on a desktop PC with an NVIDIA GeForce GTX 1080 Ti graphics card. The 2S-LSTM network was applied using an HTC VIVE Pro Eye headset, while Oculus Quest 2 and Leap Motion served as baseline solutions. The VR environment and other VR models utilized in the experiment were developed in VRWS design phase. Various USB cameras were employed to record experimental data from the participants.

6.1.3. Experimental Conditions

- **Regular Typing (Normal):**

  Participants initially completed typing tasks without wearing the HMD for 30 minutes. This condition served as a baseline to assess participants' regular typing ability.

- **HMD Typing:**

  Participants wore the HMD and performed typing tasks using three distinct typing assistance solutions—Oculus Quest 2, Leap Motion, and the developed 2S-LSTM solution. Each task was conducted for 30 minutes. The order of the solutions was counterbalanced among participants to mitigate potential order effects.

6.1.4. Experiment Procedure

- **Pre-Experiment Session:**

  Participants underwent a brief training session to acquaint themselves with the HMD and the typing assistance solutions. This session ensured participants' comprehension of task requirements and their ability to perform typing tasks comfortably.

- **Typing Tasks:**

The above regular and HMD Typing were performed as the Typing Task, with breaks.

- **Breaks and Comfort:**

  Participants had the flexibility to take breaks at any point during the experiment to ensure their comfort and prevent symptoms such as "VR sickness."

- **Typing Hands Position:**

  The experimental setup involved recording participants' typing actions using a combination of a USB camera and a virtual camera within the real and VR environments. These cameras captured the real hand position and the virtual hand positions when participants pressed keys on the keyboard. The dataset for each typing session was created by combining these recordings. High hand tracking accuracy and minimal jitter were expected to result in closely resembling typing postures of real and virtual hands. The comparison of typing postures assessed the level of fidelity and jitter in replicating hand movements in the virtual environment.

As shown in Figure 6.1, experiment order for each group A, B, and C stands for Oculus Quest 2, Leap Motion, and the developed 2S-LSTM solution.

6.1.5. Data Collection

During typing tasks, the following data were collected:

- **Total number of words (NoW)** entered (including errors) in normal, Oculus Quest 2, Leap Motion, and 2S-LSTM conditions. The quantity of NoW (Number of Words) within a unit of time can also serve as a measure for typing speed and fluency.

- **Number of errors (E)** in normal, Oculus Quest 2, Leap Motion, and 2S-

LSTM conditions.

- **Error rate (ER)** in normal, Oculus Quest 2, Leap Motion, and 2S-LSTM conditions.

- **Difference (Diff.)** of hand positions in HMD typing conditions. The difference between real and virtual hand positions was quantified at 21 * 2 key points of the hand, and the differences were summed for 100 inputs.

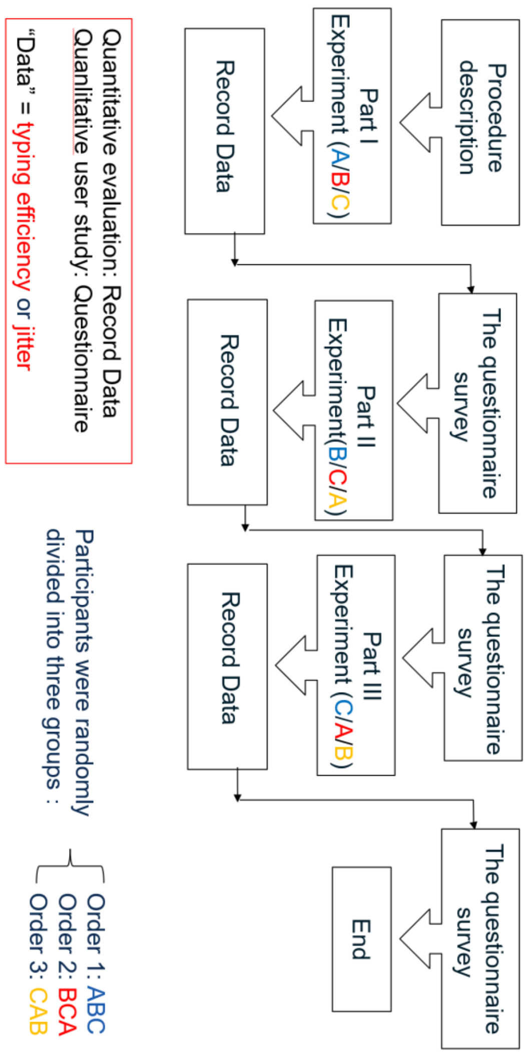


Figure 6.1: Experiment order for each group. A, B, C stand for Oculus Quest 2, Leap Motion, and the developed 2S-LSTM solution.

6.2 Questionnaire

After each typing task, participants were asked to complete a questionnaire. Each question of the questionnaire consisted of a 7-point Likert scale, ranging from 1 (negative) to 7 (positive). The questionnaire is shown in Table 6.2. The questionnaire was provided after each of the four experimental conditions: Normal, Oculus Quest 2, Leap Motion, and 2S-LSTM. The questions which were marked as "only for VR typing" were not asked in the normal condition. By administering a questionnaire to participants after the typing experiment, we aim to collect subjective evaluations of the different typing methods. Statistical analysis of the questionnaire results will also be conducted to validate potential strengths and weaknesses.

Table 6.2. Questionnaire.

1.	Did you perform at your normal typing efficiency during this typing session?
2.	How fatigued did you feel during the typing session?
3.	To what extent did the virtual hands replicate real hand position during this typing session? (Only for VR typing)
4.	Would you be willing to replace traditional typing with this typing scheme? (Only for VR typing)
5.	Please evaluate the level of jitter in this VR typing system. (Only for VR typing)
6.	How much did Jitter have a negative impact on you in last task? (Only for VR typing)

7. How much dizziness did you experience during the typing task? (Only for VR typing)
8. How comfortable did you find in the last typing task? (Only for VR typing)
9. Were there any times during the typing when you just wanted to give up?
10. Would you like to use this typing system again in the future? (Only for VR typing)
11. Please evaluate your level of focus during the typing process.
12. Did your typing fluent in last task?

6.3 Result

In order to evaluate the impact of the factors on user performance, we conducted statistical tests using SPSS software. First, we conduct tests to examine the normality and homogeneity of variance of all the collected data.

Typing Data:

The average results of typing data are shown in Fig. 3. We conducted tests for normality and homogeneity of variances. Since the sample size for all the collected data is less than 50, the Shapiro-Wilk (S-W) test was used for the normality test. The results show that E and ER for all conditions followed the normal distribution (P-values of E: 0.421, 0.137, 0.188, 0.484 respectively; P-values of ER: 0.082, 0.138, 0.338, 0.344 respectively). However, the tests for homogeneity of variances indicated that E (P = 0.011) and ER (P = 0.000**) did not meet the assumption of equal variances.

Moreover, none of the conditions exhibited normal distributions for Now and Diff. values (P values of NoW: 0.001, 0.012, 0.011, 0.001 respectively; P values of Diff.: 0.001, 0.013, 0.011 respectively). Therefore, non-parametric tests were employed to analyze the total number of words typed, number of errors, error rates, and Diff. values. Since there were more than two conditions, the Kruskal-Wallis test was used to examine the differences among conditions. The results indicated significant differences among the conditions for the NoW, E, ER, and Diff. (P values are all less than 0.05).

We conducted multiple comparisons using the Mann-Whitney U test with Bonferroni's adjustment. For NoW, the comparison of 2S-LSTM and Leap Motion is no significant difference (P = 0.357). For E, the comparison of 2S-LSTM and Leap Motion also no significant difference (P = 0.313). For other comparisons, the p-values are all less than 0.05. In summary, the number of NoW is Normal > 2S-LSTM = Leap Motion > Oculus, the number of E is Oculus > Leap Motion = 2S-LSTM > Normal, and the number of Diff. is Oculus > Leap Motion > 2S-LSTM, respectively.

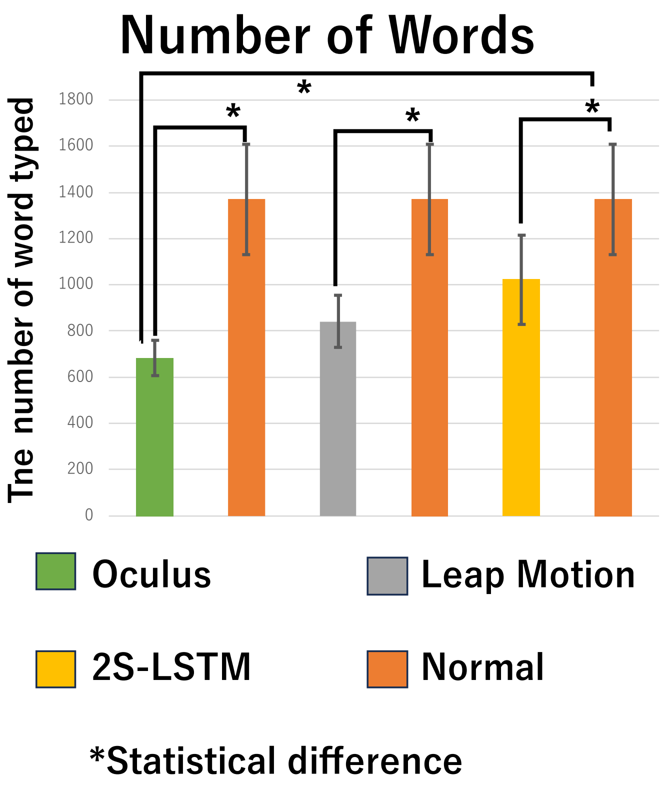


Figure 6.2: The result of Now.

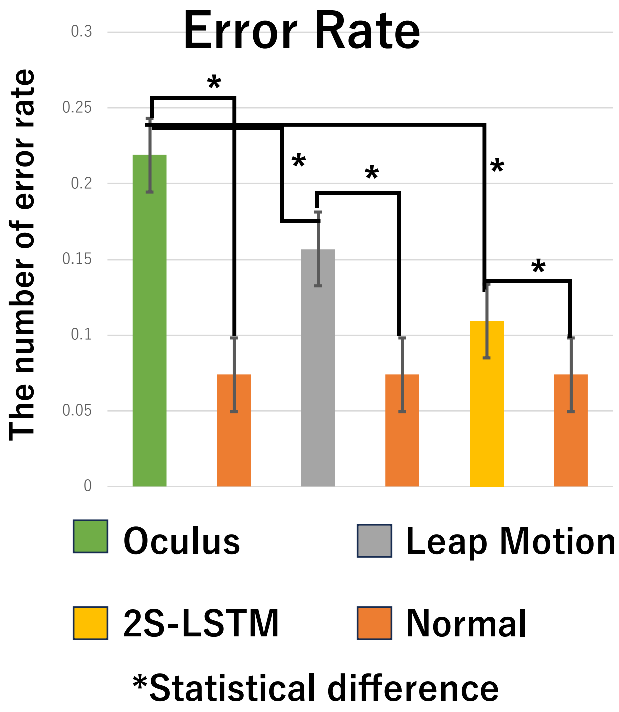


Figure 6.3: The result of ER.

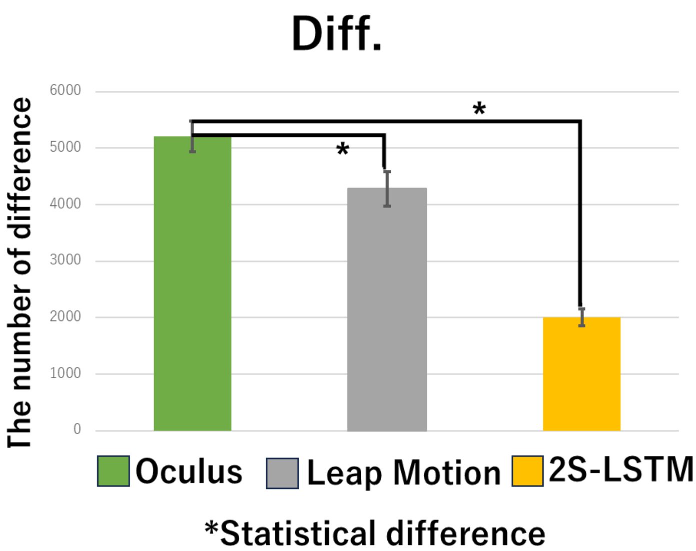


Figure 6.4: The result of Diff.

Questionnaire:

The sample sizes for all questions are less than 50, the Shapiro-Wilk (S-W) test was used. However, the data for all these questions did not exhibit normal distribution characteristics (P values are all less than 0.05). Therefore, non-parametric tests were employed. Since there are more than two experimental conditions, the Kruskal-Wallis test statistic was used for analysis. The results showed that there was no significant difference among the different conditions for Question 11 ($H= 0.446$, $p = 0.93$). For other questions, the different typing conditions demonstrated significant differences (all p values are less than 0.05). The average score of each question in different conditions is shown in Figure. 6.5.

We also performed multiple comparisons using the Mann-Whitney U test with Bonferroni's adjustment. For Question 2, the comparison of 2S-LSTM and Normal is no significant difference ($P = 0.514$). For Question 7, Question 8, and Question 12, the comparison of Leap Motion and Oculus is no significant difference ($P = 0.445$, $P = 0.102$, $p = 0.101$). For Question 9, the comparison of 2S-LSTM and Leap Motion is no significant difference ($P = 0.054$). The results were (Normal >) 2S-LSTM > Leap Motion > Oculus for most questions.

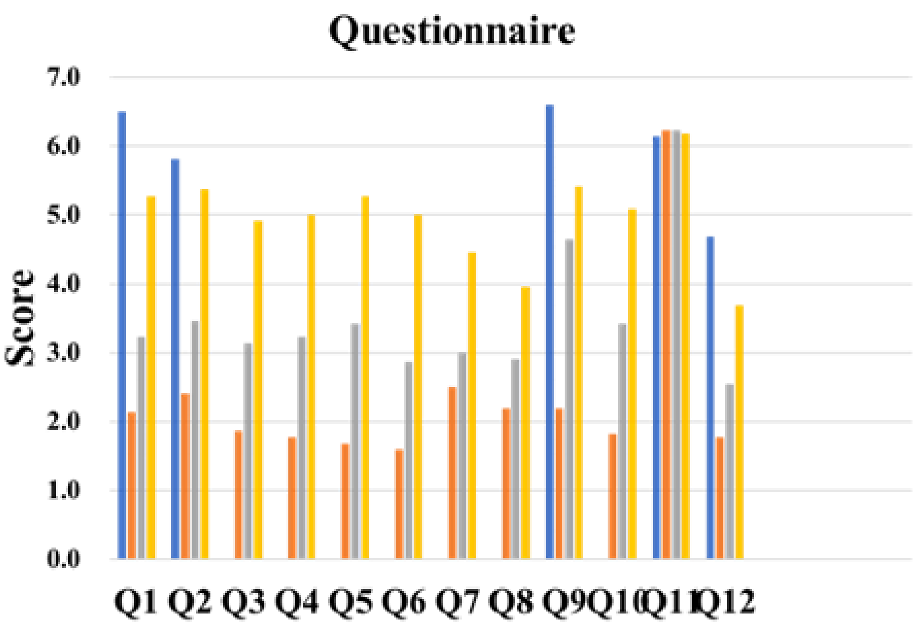


Figure 6.5: The result of questionnaire.

6.4 Discussion

Typing Data:

Notably, it is evident from our statistical analysis that the 2S-LSTM outperformed the Oculus Quest 2 and Leap Motion. These findings highlight the importance of considering the specific typing scheme when evaluating typing efficiency, error rates, and Diff. values. The Mann-Whitney U test with Bonferroni's adjustment was conducted to obtain these results.

From the results of the Mann-Whitney U test with Bonferroni's adjustment, we can conclude that there is no significant difference between 2S-LSTM and Leap Motion in the number of inputs and errors quantity per unit time. Our method utilizes a regular RGB camera on HMD, while Leap Motion employs a depth camera. Therefore, achieving similar results to Leap Motion by using a regular device is still considered a positive outcome. Additionally, there is a significant difference between 2S-LSTM and the other methods in Diff.. This result indicates that using the original image and MHI, combined with the implementation of KF to reduce jitter, indeed leads to a reduction in the Diff.. Considering the deployment cost and the other results obtained in this research, we have reasons to believe that our approach is superior to the Leap Motion and Oculus solutions.

Questionnaire:

The questionnaire focused on various aspects such as typing efficiency, fatigue, replication of hand position, the willingness to replace traditional typing, evaluation of jitter, negative impact of jitter, dizziness, comfort, willingness to continue using the system, focus level, and typing fluency. The statistical analysis involved non-parametric tests due to the data not exhibiting normal distribution characteristics. The questionnaire results indicated no significant difference among the different conditions for Question 11, which evaluated the level of focus during the typing process. This suggests that the different typing conditions, including the use of 2S-LSTM, Oculus Quest 2, and Leap Motion, did not significantly affect the participants' focus level.

Notably, it can be observed that the 2S-LSTM condition generally outperformed the

Oculus Quest 2 and Leap Motion conditions in terms of typing efficiency, fatigue, replication of hand position, willingness to replace traditional typing, evaluation of jitter, negative impact of jitter, dizziness, comfort, and typing fluency. These findings suggest that the 2S-LSTM typing solution showed promising results in various aspects compared to the existing solutions of Oculus Quest 2 and Leap Motion. The 2S-LSTM condition exhibited higher typing efficiency, lower fatigue levels, better replication of hand position, and a more positive user experience.

From the results of the Mann-Whitney U test with Bonferroni's adjustment, for question 2: "How fatigued did you feel during the typing session?", there was no significant difference between 2S-LSTM and Normal. This result shows 2S-LSTM performs excellently in the VR typing task, and users do not experience additional fatigue from VR. For Question 7: "How much dizziness did you experience during the typing task?", Question 8: "How comfortable did you find the last typing task?", and Question 12: "Did your typing feel fluent in the last task?", there was no significant difference between Leap Motion and Oculus. However, our approach showed significant differences in these questions compared to Leap Motion and Oculus, and the questionnaire results are more positive. This outcome suggests that, compared to the existing VR systems, our approach is hard to make user fill dizziness in VR typing tasks and also superior in comfort and typing fluency. Furthermore, for Question 9: "Were there any times during the typing which you just wanted to give up?", there is no significant difference between our approach and Leap Motion. Although our approach performed better in reducing dizziness, improving comfort, and better typing fluency, users still want to give up while using our approach to type. We suggest that there might be some hidden flaws in our approach that lead to user dissatisfaction. Therefore, further discussions and investigations regarding this issue are essential for future improvements.

There are still some limitations in this research. Firstly, the sample size for the questionnaire was limited to a specific number of participants. Expanding the sample size and including a more diverse group of participants could enhance the generalizability of the findings. Additionally, the study focused on specific typing tasks and conditions, and further investigation is needed to evaluate the solution's performance in different contexts and for various user profiles. In conclusion, the results of the questionnaire highlighted significant differences among the different typing conditions, with the 2S-LSTM solution demonstrating superior performance compared to the Oculus Quest 2 and Leap Motion solutions. These findings support the effectiveness of the developed solution in improving typing efficiency, reducing fatigue, and providing a more comfortable and satisfactory typing experience.

Chapter 7

Typing Behavior

In a previous experiment, we observed variations in finger usage among participants under different experimental conditions. Specifically, we noticed that participants' typing habits were influenced by changes in the experimental setup. Based on these observations, we hypothesize the following:

- The more effective a VR typing solution is, the less it affects the user, resulting in a smaller difference in typing habits compared to normal typing.

7.1 Preliminary experiment

We employed distinct colors to represent different fingers and utilized color blocks to indicate the corresponding striking positions. The size of these color blocks was employed to signify the frequency of finger strikes. On a given key, a more prominent color block denoted a higher frequency of strikes from the corresponding-colored finger, while a smaller block indicated a relatively lower frequency, the result show in Figure. 7.1. In instances where the total number of strikes for a specific finger on a key was less than 25% of the overall total strikes, it was considered an exception and was not included in the statistical analysis.

Among the four participants, the most notable differences in color block movement and size changes in comparison to normal typing were observed with the Oculus system. In contrast, both Leap Motion and our proposed solution exhibited relatively minor variations in movement and size compared to the baseline of normal typing.

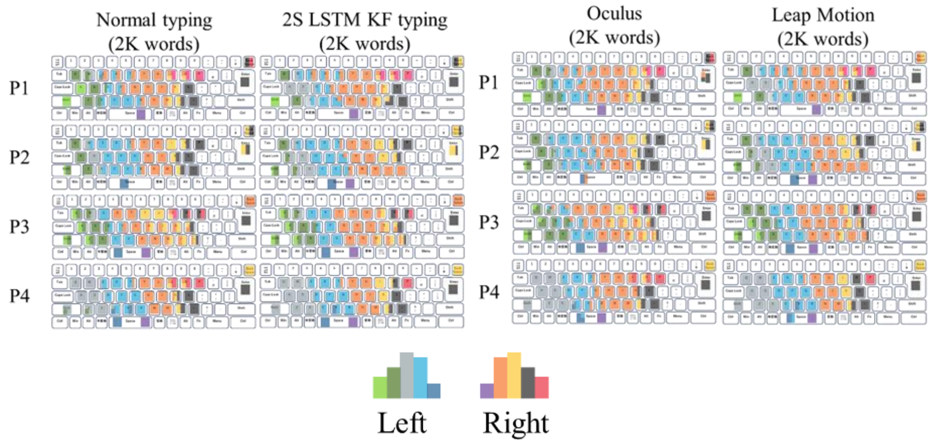


Figure 7.1: Users' typing behaviors in different conditions. The colors represent the use of different fingers.

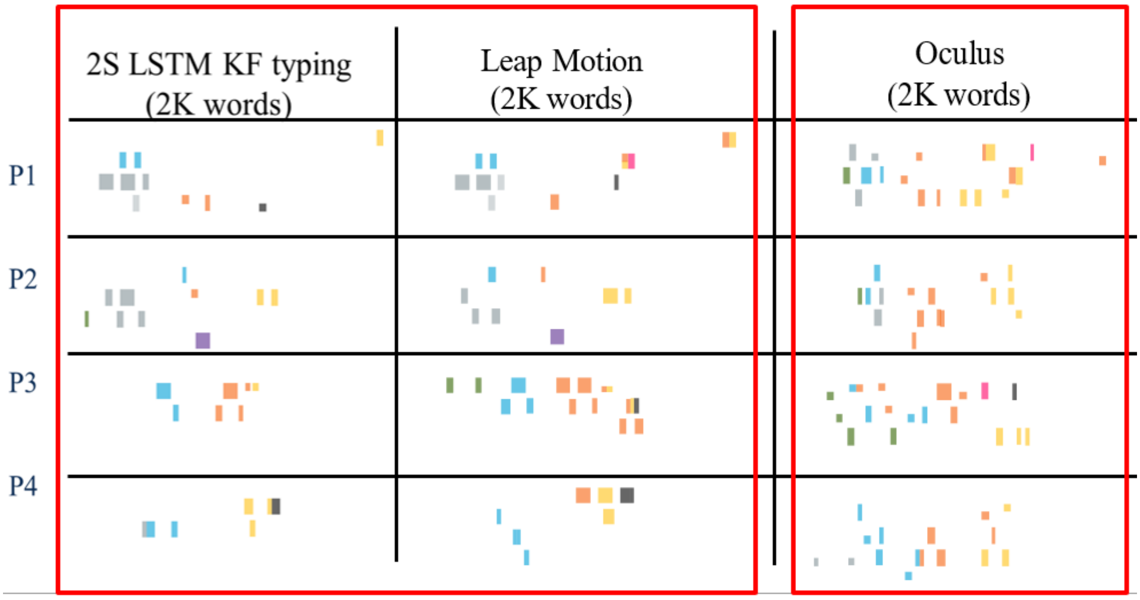


Figure 7.2: Remove the repeating colors, and the remaining colors represent different typing behaviors.

During the earlier experimental phase, we observed that when VR typing was not smooth, users tended to more frequently resort to using their dominant fingers for key presses in the VR environment. In this experiment, we represented finger usage with different colors. To enhance clarity in interpreting the data, we excluded other colors and focused solely on the color representing the use of both index fingers. The results are depicted in Figure 7.2. It can be observed that there is not a significant difference between 2S-LSTM and normal conditions. However, under the Oculus and leap motion condition, the color area for the right index finger is larger than others. Considering all participants are left-handed, this does provide some insights into certain aspects of the issue.

7.2 Typing Behavior Experiment

To further evaluate our proposed solution and verify this hypothesis, we decided to conduct a detailed analysis of the typing habit data collected under four different conditions again: regular typing, Oculus, Leap Motion, and our solution.

The overall experimental design in this experiment, including the settings for participants, equipment, experimental conditions, and experimental procedure, remains largely consistent with the typing experiment detailed in Section 6.1. To avoid redundancy, only the aspects that differ from the previous experiment will be explicitly introduced in this section. Commonalities will not be reiterated.

7.2.1. Participants

A total of 22 participants were recruited in this experiment. Unlike the previous experiment (5.1 Typing Experiment), where prior VR experience was considered, all participants in this study were VR novices, having never engaged in typing within a VR environment before. The participant group included 21 right-handed individuals and 1 left-handed individual, with 15 males and 7 females, maintaining an average age of M = 26.

All participants were proficient in English, ensuring that typing in English posed no challenges. In contrast to the previous experiment, where advanced touch-typing skills were not required, this study imposed no restrictions on participants' typing skills, allowing individuals with advanced touch-typing abilities to participate as well.

7.2.2. Equipment

The equipment setup was largely consistent with the previous experiment, with the addition of a camera and the use of Media Pipe to accurately record which keys each finger pressed during typing. This addition was specifically implemented to capture and analyze participants' typing habits more accurately.

7.2.3. Experimental Conditions and Procedure

The experimental conditions and procedures in this section were identical to those outlined in Section 6.1.3 and Section 6.1.4 of the typing experiment. All participants underwent the same pre-experiment training session, followed the same typing tasks, and had the same flexibility to take breaks. Typing hands were recorded using the same methods, with no additional modifications to the setup.

7.2.4. Data Collection

To investigate whether the participants' typing habits changed under different VR typing conditions, we collected the following data:

- Typing habit data: We extracted the number of times each participant used each finger in four different conditions from the typing experiment.
- Typing habit difference data: We calculated the differences in typing habits by comparing the three VR typing conditions with the normal condition.

Subsequently, we performed cluster analysis and statistical analysis to determine whether the typing conditions influenced participants' typing habits and to clarify the specific nature of these changes. The typing habit data are recorded in Appendix A and show in Table A1.

It is important to note that during the actual typing tasks, participants did not use their thumbs to type on keys other than the spacebar. Therefore, we focused only on the usage of the eight fingers, excluding the thumbs. The fingers are named from the left pinky to the left index and the right index to the right pinky: L1, L2, L3, L4, R4, R3, R2, R1.

7.2.5. Use Typing Habit Data to Cluster

For all participants' typing habit data, we used k-means clustering. We used k-means clustering for two primary reasons: (1) k-means is not very sensitive to outliers in the data; and (2) k-means is well-known and easy to implement. Table 7.1 shows the sum of squares due to error (SSE) and average silhouette width (ASW).

Table 7.1: SSE and ASW values for different cluster numbers.

Cluster Number	SSE (the Sum of Squares Due to Error)	ASW (Average Silhouette Width)
2	425.782	0.380
3	379.603	0.407
4	318.158	0.495
5	301.294	0.508
6	301.862	0.509

Through practical observation, two clusters are the most suitable. One cluster consists of typists who use five fingers on each hand (referred to as “balance typists”), while the other cluster consists of typists who use only two or three fingers on each hand (referred to as “crab typists”). Although the SSE and ASW values for the four-cluster solution are better than those for the two-cluster solution, some clusters in the four-cluster solution are too small, making the two-cluster solution more practical. Details of the two-cluster solution and four-cluster solution are shown in Tables 7.2 and Table 7.3.

Table 7.2: Two-cluster solution details.

Clustering Category	frequency	Percentage (%)
Cluster_1 (crab typist)	9	40.91%
Cluster_2(balance typist)	13	59.09%
Sum	22	100%

Table 7.3: Four-cluster solution details.

Clustering Category	frequency	Percentage (%)
Cluster_1	2	9.09%
Cluster_2	13	59.09%
Cluster_3	4	18.18%
Cluster_4	3	13.64%
Sum	22	100%

Figure 7.3 illustrates L1 to R1 fingers usage by 22 participants under different conditions. The usage of L1 to R1 fingers in different conditions was visualized using Python, based on the typing habit data collected from participants. The data includes the frequency of L1 and R1 finger usage across four typing conditions. This visualization confirms distinct differences in typing behaviors between two clusters: Cluster_1 (crab

typists) and Cluster_2 (balance typists). Notably, balance typists show relatively stable usage of L1 and R1 across different conditions, whereas crab typists exhibit an increased usage trend under the Leap and Oculus conditions compared to the normal and 2S conditions. This pattern highlights the influence of VR typing conditions on finger usage, a point further explored in the Discussion section to understand the adaptive responses of crab typists in varied VR environments.

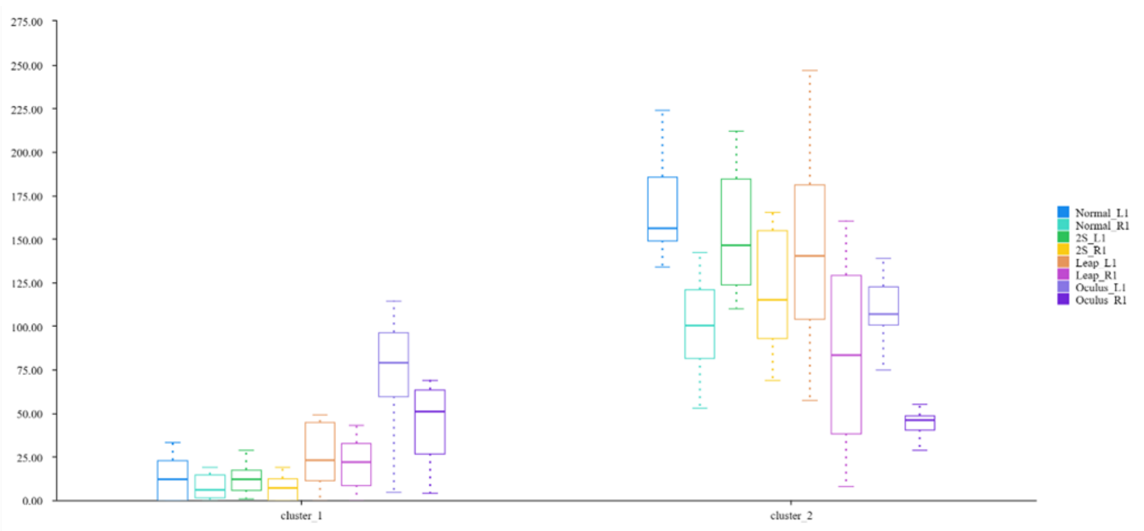


Figure 7.3: Users' typing behaviors in different conditions. The colors represent the use of different fingers.

7.2.6. Use Typing Habit Data to Re-clustering

By clustering the 22 participants, we identified two clusters representing crab typists and balance typists, which aligns with our actual observations of all participants during the typing tasks. Next, we re-cluster the typing habits of these two types of typists under the four typing conditions to clarify their more detailed typing characteristics.

It is important to note that there are 9 participants in cluster 1 and 13 participants in cluster 2, which is consistent with the actual situation. Therefore, we re-clustered the typing habits of the 9 participants in cluster_1 under 4 conditions and the 13 participants in cluster_2 under 4 conditions. This results in 9 participants × 4 conditions = 36 data points for cluster 1, and 13 participants × 4 conditions = 52 data points for cluster 2.

1. Crab typists.

We used k-means to re-cluster the typing habits. The results of the re-clustering are shown in Table 7.4. The variance analysis results are shown in Table 7.5.

Table 7.4: Re-clustering results for crab typists.

Clustering Category	Frequency	Percentage (%)
Cluster 1_1	11	30.56%
Cluster 1_2	25	69.44%
Sum	36	100%

Table 7.5: Comparison results of variance analysis of clustering categories.

	Mean ± Standard Deviation		F	*p*
	Cluster_1 (n = 11)	Cluster_2 (n = 25)		
L1	72.27 ± 23.90	12.88 ± 11.02	106.175	0.000 **
L2	203.00 ± 44.58	149.08 ± 80.95	4.262	0.047 *
L3	444.18 ± 72.11	443.76 ± 90.00	0.000	0.989
L4	540.09 ± 63.17	535.12 ± 70.24	0.041	0.842
R4	557.18 ± 69.57	626.16 ± 61.57	8.865	0.005 **
R3	489.27 ± 85.11	588.92 ± 65.83	14.616	0.001 **
R2	147.82 ± 71.48	135.32 ± 67.24	0.254	0.617
R1	46.18 ± 15.78	8.76 ± 7.45	95.158	0.000 **

* $p < 0.05$; ** $p < 0.01$.

From Table 7.5, the items L1, R4, R3, and R1 exhibit highly significant differences ($p < 0.05$ or $p < 0.01$), reflecting notable changes in usage patterns. These significant results suggest that crab typists vary their usage of L1, R4, R3, and R1 across different conditions, possibly adapting these finger movements to accommodate VR-related constraints.

2. Balance typists.

We still used k-means to re-cluster the typing habits for balance typists. The results of the re-clustering are shown in Table 7.6. The variance analysis results are shown in Table 7.7.

Table 7.6: Re-clustering results for balance typists.

Clustering Category	Frequency	Percentage (%)
Cluster 2_1	22	42.31%
Cluster 2_2	30	57.69%
Sum	52	100%

Table 7.7: Comparison results of variance analysis of clustering categories.

	Mean ± Standard Deviation		F	*p*
	Cluster_1 (n = 22)	**Cluster_2 (n = 30)**		
L1	141.77 ± 51.52	144.60 ± 33.62	0.057	0.812
L2	146.55 ± 25.01	202.80 ± 30.25	50.625	0.000 **
L3	491.73 ± 73.70	486.30 ± 44.10	0.110	0.742
L4	598.36 ± 52.23	499.87 ± 56.76	40.851	0.000 **
R4	606.64 ± 34.30	525.23 ± 44.44	51.308	0.000 **
R3	279.68 ± 69.46	324.10 ± 114.05	2.617	0.112
R2	153.91 ± 44.45	226.33 ± 52.56	27.373	0.000 **
R1	81.36 ± 49.77	90.77 ± 34.92	0.642	0.427

* $p < 0.05$; ** $p < 0.01$.

Items L2, L4, R4, and R2 have p-values below 0.01, indicating significant variance across clusters. This finding implies that balance typists demonstrate notable differences in the usage of L2, L4, R4, and R2, highlighting the impact of VR environments on their typing patterns for these specific fingers.

7.2.7. Use Typing Habit Difference Data to Cluster

Typing habit difference data represents the differences in finger usage between VR conditions and normal condition. Similar to previous steps, we used k-means clustering on this data for the 22 participants. The SSE and ASW values are shown in Table 7.8.

Table 7.8: SSE and ASW values for different cluster numbers.

Cluster Number	SSE (the Sum of Squares Due to Error)	ASW (Average Silhouette Width)
2	369.933	0.292
3	290.189	0.368
4	278.493	0.387

We rely on the SSE and ASW values to determine the optimal number of clusters. As shown in Table 7.8, a cluster number of 3 shows an optimal inflection point for both SSE and ASW. Hence, we chose a cluster number of 3. Details of the three-cluster solution are shown in Table 7.9, with variance analysis results in Table 7.10. Here, N-2S represents the difference in typing habits between 2S-LSTM and Normal conditions, N-Le represents the difference between Leap Motion and Normal conditions, and N-Oc represents the difference between Oculus Quest 2 and Normal conditions. L1 to R1 represents different fingers.

Table 7.9: Three-cluster solution details.

Clustering Category	frequency	Percentage (%)
Cluster_1	5	22.73%
Cluster_2	9	40.91%
Cluster_3	8	36.36%
Sum	22	100%

Considering the results in Table 7.9 and the actual types of typists, we found that the 9 crab typists were still clustered into one group, while the 13 balance typists were clustered into two groups. This indicates that the changes in typing habits among crab typists tend to be consistent, whereas the changes in typing habits among balance typists fall into two distinct categories.

Table 7.10: Comparison results of variance analysis of clustering categories.

	Mean ± Standard Deviation			F	*p*
	Cluster_1 (n = 5)	Cluster_2 (n = 9)	Cluster_3 (n = 8)		
N-2SL1	6.20 ± 17.28	0.00 ± 14.70	19.13 ± 15.21	3.304	0.059
N-2SL2	32.00 ± 14.65	5.78 ± 10.40	21.25 ± 12.10	8.294	0.003 **
N-2SL3	36.40 ± 38.55	−28.89 ± 56.62	−38.00 ± 40.05	4.228	0.030 *
N-2SL4	−26.20 ± 38.32	18.78 ± 58.52	47.00 ± 40.70	3.491	0.051
N-2SR4	8.40 ± 17.01	−8.89 ± 32.98	−14.88 ± 16.45	1.385	0.274
N-2SR3	−27.80 ± 24.89	5.11 ± 27.71	−10.25 ± 15.67	3.259	0.061
N-2SR2	−3.80 ± 14.25	7.00 ± 10.90	−14.00 ± 12.75	6.128	0.009 **
N-2SR1	−25.20 ± 19.31	1.11 ± 4.76	−10.25 ± 15.53	6.348	0.008 **
N-LeL1	−27.80 ± 12.87	−13.89 ± 16.36	48.88 ± 34.34	20.602	0.000 **
N-LeL2	63.80 ± 10.89	−5.11 ± 32.26	10.63 ± 28.76	10.199	0.001 **
N-LeL3	34.80 ± 39.91	39.78 ± 51.98	−41.63 ± 22.61	9.739	0.001 **
N-LeL4	−67.40 ± 38.40	−41.78 ± 49.96	−3.63 ± 24.97	4.243	0.030 *
N-LeR4	−57.80 ± 18.02	2.11 ± 41.14	−28.50 ± 21.93	6.246	0.008 **
N-LeR3	11.40 ± 19.22	8.00 ± 37.68	−29.00 ± 25.60	4.074	0.034 *
N-LeR2	65.80 ± 10.89	23.22 ± 28.34	2.63 ± 28.21	9.412	0.001 **
N-LeR1	−22.80 ± 12.87	−12.33 ± 11.00	40.63 ± 36.99	14.168	0.000 **
N-OcL1	62.60 ± 42.32	−60.89 ± 34.57	55.75 ± 33.53	29.283	0.000 **
N-OcL2	71.20 ± 20.89	−46.89 ± 76.19	71.25 ± 40.37	11.830	0.000 **
N-OcL3	−78.80 ± 52.35	−48.67 ± 80.77	−75.00 ± 69.15	0.408	0.670
N-OcL4	−51.80 ± 54.61	13.78 ± 99.35	−108.00 ± 67.18	4.892	0.019 *
N-OcR4	−91.20 ± 60.13	62.00 ± 52.23	−88.63 ± 40.72	24.255	0.000 **
N-OcR3	−79.80 ± 43.34	112.11 ± 114.65	0.38 ± 127.28	5.373	0.014 *
N-OcR2	103.40 ± 74.19	5.22 ± 92.13	91.00 ± 60.74	3.618	0.047 *
N-OcR1	64.40 ± 21.31	−36.67 ± 22.01	53.25 ± 19.96	53.285	0.000 **

* $p < 0.05$ ** $p < 0.01$.

From Table 7.10, the majority of items exhibit significant differences (p < 0.05 or p < 0.01). Under the N-Le and N-Oc conditions, nearly all items display significant differences (with only N-OcL3 showing no significance), whereas only half of the items under the N-2S condition show significant differences. This reflects variations in typing habits across different VR modes. These important findings indicate that typists adjust the usage of almost all their fingers under the Leap Motion and Oculus conditions, while only half of the finger usage patterns show changes under the 2S-LSTM condition. This further supports the idea that typists modify their finger movements to adapt to constraints specific to each VR condition.

7.2.8. Statistical Test

To identify the specific changes in finger usage for crab typists and balance typists under different conditions, we conducted statistical tests to analyze their typing habit difference data.

1. Compare crab typist's typing differences in different conditions.

Because some data lack normality and homogeneity of variance, we used Welch ANOVA, a robust alternative to standard ANOVA when assumptions of normality and homogeneity of variance are violated. This method accommodates unequal variances across groups and reduces the risk of Type I error under these conditions, making it suitable for our dataset. The results are shown in Table 7.11. The normality and homogeneity of variance test results are recorded in Appendix B.

From Tables A2 and A3, we can see that some data do not have normality and homogeneity of variance. Therefore, we used Welch ANOVA in the next step, the results shown in Table 7.11.

Table 7.11: The result of welch ANOVA for crab typists.

	Condition (Standard Deviation)					
	Normal (n = 9)	**2S-LSTM (n = 9)**	**Leap Motion (n = 9)**	**Oculus Quest 2 (n = 9)**	***Welch F***	***p***
L1	12.33 ± 12.05	12.33 ± 8.28	26.22 ± 18.27	73.22 ± 32.92	10.054	0.001 **
L2	154.00 ± 76.93	148.22 ± 77.21	159.11 ± 82.56	200.89 ± 65.46	1.010	0.411
L3	434.44 ± 77.22	463.33 ± 115.12	394.67 ± 47.75	483.11 ± 65.76	3.582	0.036 *
L4	534.33 ± 64.59	515.56 ± 64.57	576.11 ± 81.64	520.56 ± 47.50	1.151	0.356

	Condition (Standard Deviation)					
	Normal (n = 9)	2S-LSTM (n = 9)	Leap Motion (n = 9)	Oculus Quest 2 (n = 9)	*Welch F*	*p*
R4	618.89 ± 61.50	627.78 ± 74.65	616.78 ± 69.67	556.89 ± 64.79	2.015	0.148
R3	589.78 ± 62.61	584.67 ± 77.05	581.78 ± 67.33	477.67 ± 85.53	3.693	0.032 *
R2	148.00 ± 75.02	141.00 ± 77.18	124.78 ± 61.84	142.78 ± 65.95	0.196	0.898
R1	8.22 ± 6.89	7.11 ± 6.79	20.56 ± 14.17	44.89 ± 21.92	9.235	0.001 **

* $p < 0.05$; ** $p < 0.01$.

It can be concluded that samples with different conditions do not show significant differences in terms of L2, L4, R4, and R2. However, samples with different conditions show significant differences in terms of L1, L3, R3, and R1. The analysis and comparison results of all fingers under the four conditions are shown in Figure 7.4.

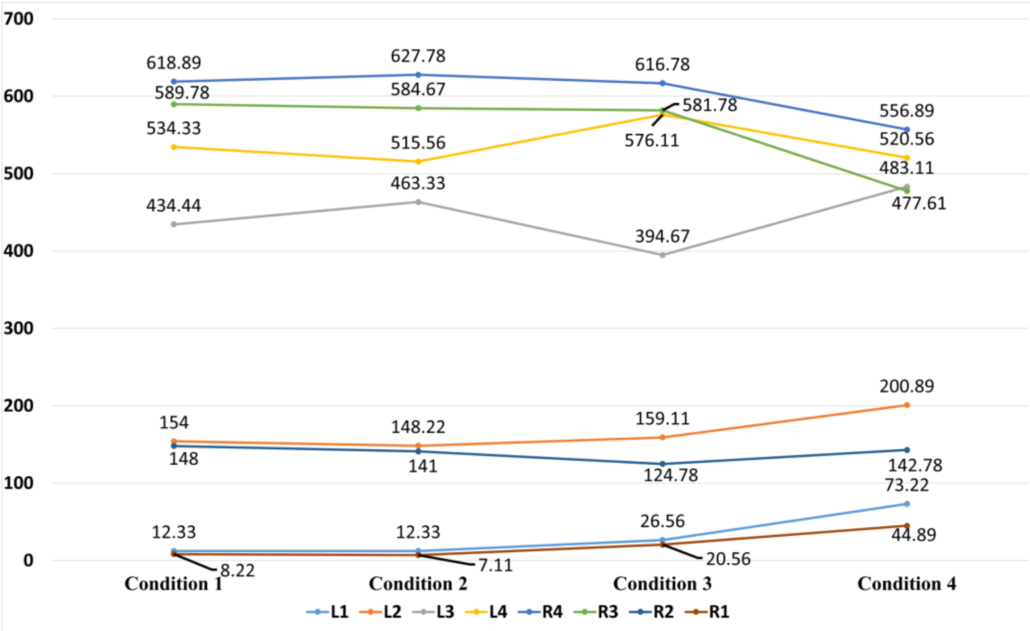


Figure 7.4: Comparison of finger usage analysis under four conditions.

From the above analysis, it can be concluded that crab typists exhibit different typing styles in L1, L3, R3, and R1 fingers under different conditions.

2. Compare balance typist's typing differences in different conditions.

Following the analysis of typing habit differences for crab typists, we conducted a similar analysis for balance typists. Similar to the previous step, because some data do not meet the assumptions of normality and homogeneity of variance, we used Welch ANOVA. This approach is specifically recommended for datasets with unequal variances and non-normal distributions, allowing for more accurate comparisons across the groups in question. The results are shown in Table 7.12, with normality and variance homogeneity test results recorded in Appendix B.

Table 7.12: The result of welch ANOVA for balance typists.

	Condition (Standard Deviation)					
	Normal (n = 9)	**2S-LSTM (n = 9)**	**Leap Motion (n = 9)**	**Oculus Quest 2 (n = 9)**	***Welch F***	***p***
L1	166.38 ± 25.62	152.23 ± 33.98	147.00 ± 56.55	108.00 ± 18.64	15.970	0.000 **
L2	210.92 ± 29.10	185.54 ± 33.11	179.85 ± 36.06	139.69 ± 24.17	15.560	0.000 **
L3	464.08 ± 37.30	473.46 ± 64.19	476.31 ± 59.60	540.54 ± 34.44	10.843	0.000 **
L4	517.62 ± 49.81	498.77 ± 74.43	545.77 ± 65.77	604.00 ± 60.23	6.764	0.002 **
R4	525.85 ± 49.87	531.77 ± 49.64	565.62 ± 47.87	615.46 ± 33.21	13.369	0.000 **
R3	290.08 ± 113.94	307.08 ± 106.05	303.54 ± 122.54	320.54 ± 46.98	0.308	0.819
R2	223.85 ± 48.35	233.92 ± 47.20	196.92 ± 52.32	128.08 ± 31.35	20.486	0.000 **
R1	101.23 ± 25.70	117.23 ± 32.82	85.00 ± 48.33	43.69 ± 7.51	38.024	0.000 **

* $p < 0.05$ ** $p < 0.01$.

It can be concluded that samples with different conditions show significant differences in all terms except R3. The analysis and comparison results of all fingers under the four conditions are shown in Figure 7.5.

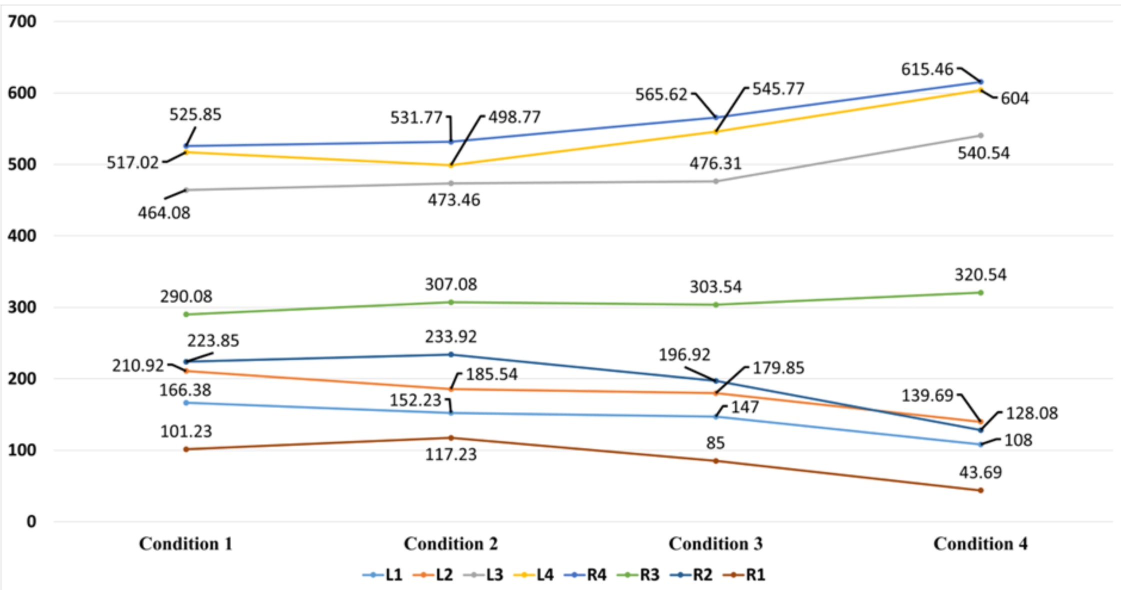


Figure 7.5: Comparison of balance typists' finger usage analysis under 4 conditions.

From the above analysis, we can see that in different conditions, balance typists will not change their typing habit in R3 but exhibit different typing styles in other fingers.

7.2.9. Result Summary

From the clustering of normal data combined with practical experience, it is evident that there are two types of typists: crab typists and balance typists (Sections 7.2.5–7.2.7). Both types of typists have distinct typing habits, and these habits change differently in various VR environments (Section 7.2.8). According to the actual data, compared with normal and 2S conditions, crab typists will increase the use of L1 and R1 and decrease the use of L3 and R3 in Leap and Oculus conditions (degree of change: normal <= 2S < Leap < Oculus). Conversely, balance typists will change their typing habits in a more chaotic manner (degree of change: normal < 2S < Leap < Oculus).

7.2.10. Discussion

The analysis reveals that both crab and balance typists exhibit changes in their typing habits under different VR conditions. However, the nature and extent of these changes vary between the two groups. From Section 7.2.7 we can know that crab typists show a more consistent pattern of change, while balance typists exhibit a more unpredictable alteration in their typing habits. This insight could inform the design of VR typing systems to better accommodate different typing styles and enhance user experience.

1. Behavior of balance typists.

Balance typists displayed a systematic change in their typing habits across different VR environments. Both initiative and passive changes were noted:

- Initiative changes: Balance typists consciously reduced the use of error-prone fingers (R1, R2, L1, and L2) and increased reliance on other fingers (R4, L3, and L4) to maintain typing efficiency. The reason for the change is that R1, R2, L2, and L1 are error-prone, and changes are made to maintain typing efficiency.
- Passive changes: The same shift in finger usage occurred reactively, as balance typists compensated for errors by using more reliable fingers for corrections. The reason for the change is that R1, R2, L2, and L1 are error-prone; to edit errors, use other fingers to re-type.

Interview feedback confirmed these findings, with balance typists reporting an awareness of their changing habits. They attributed these adjustments to the higher error rates and the need to maintain their overall typing speed and accuracy in VR.

2. Behavior of crab typists.

Crab typists, who typically do not use their pinkies, exhibited a unique pattern of adaptation:

- Increased pinky usage: Despite their usual reluctance, crab typists increased their use of pinkies in VR, particularly with the Oculus system. This increase, ranging from one to three times their normal usage, though still less frequent than balance typists, suggests a significant behavioral shift.
- Unawareness of changes: Unlike balance typists, crab typists often did not perceive their habits as having changed. This lack of awareness indicates an unconscious adaptation process, likely driven by the VR system's feedback mechanisms rather than a deliberate strategy.

Interviews highlighted the challenges crab typists faced, with many reporting unexpected difficulties and a heightened impact of VR hand motion accuracy. Despite these challenges, the increased pinky usage suggests that the VR environment might implicitly encourage (or force) a more balanced finger usage.

3. Common factors and additional insights.

Both groups noted the substantial impact of VR hand motion accuracy on their typing experience. This feedback aligns with the broader observations of adaptation and change in typing behavior:

- Perception of VR Tools: Many participants felt they were typing with a VR controller rather than their hands. This perception can be compared to the "fake hand experiment," where the brain is tricked into perceiving a fake hand as part of the body. In VR, if the hand models are highly realistic and closely mimic human hands, users can more easily adapt and integrate their virtual hands as part of their body. Conversely, suppose the hand models are less realistic or resemble controllers rather than hands. In that case, it becomes difficult for users to feel a natural connection, leading to disconnection and impacting their typing behavior.
- Adaptation over time: Some participants reported that the feeling of using a

controller persisted throughout the experiment, while others adapted over time, suggesting that familiarity with the VR setup could reduce the sense of disconnection and lead to more stable typing habits.

Chapter 8

Conclusion

8.1 Findings

This research sets out to explore how VRWS can address work efficiency challenges and identify factors that influence productivity in virtual environments. By investigating the potential of VRWS and proposing a novel typing solution, this study has laid the groundwork for the development of more effective and immersive virtual office environments.

Findings Related to VRWS Design and Work Efficiency:

In the initial stages, we hypothesized that a VRWS incorporating the best characteristics of OPWS—such as minimal visual and auditory interference, adequate natural light, privacy protection, and ergonomic design—could maintain or even improve work efficiency. Through experiments, it was found that the conclusions drawn from OPWS-related research could serve as a foundational design standard for VRWS. Using these principles, VRWS demonstrated several significant advantages compared to OPWS:

- VRWS was rated more positively in terms of being relaxing, enjoyable, and quieter.
- Participants reported reduced psychological pressure, a freer atmosphere, and greater enjoyment during work.
- VRWS provided a more comfortable experience, with better lighting effects, generating positive emotions and increasing work enthusiasm.
- These factors contributed to overall improvements in productivity and user satisfaction.

This study represents one of the first attempts to establish a design standard for VRWS. The proposed standard not only enhances the practicality of VRWS but also provides a framework for other researchers and developers to create more efficient and user-friendly virtual workspaces.

Findings Related to Typing in VR:

While VRWS shows great promise in improving work environments, a critical challenge identified during this research was the lack of robust typing support. Typing is a fundamental activity in office settings, and the limitations of existing text input solutions in VR hinder the realization of efficient office work. Addressing this gap, we proposed the 2S-LSTM typing solution, which leverages machine learning techniques to enhance typing performance in immersive environments. Key findings include:

- The 2S-LSTM solution, which utilizes images of the back of the hand, demonstrated superior performance compared to existing solutions like Oculus Quest 2 and Leap Motion.
- It significantly improved typing efficiency, reduced fatigue and system jitter, accurately replicated hand positions, and provided a more positive user experience.
- Experiments revealed minimal differences in users' typing behaviors when using the 2S-LSTM solution compared to traditional typing methods, indicating that the solution maintains users' natural typing habits.

Through additional performance comparisons, the 2S-LSTM-KF model showed strong performance across metrics such as latency, accuracy, and jitter. While other advanced models, like 2S-BNNActionNet-KF, demonstrated similar potential, the 2S-LSTM-KF model was ultimately chosen due to its balance of performance and deployment feasibility.

Practical Contributions:

This research has practical implications for various fields, including distance learning, telecommuting, and virtual collaboration. By addressing the challenges of text entry in VR, the proposed solutions can facilitate the adoption of VR technology across diverse applications. Moreover, the insights gained from the experiments provide valuable guidance for designing more effective VR environments that prioritize user comfort, productivity, and satisfaction.

In summary, this study not only establishes a design standard for VRWS but also introduces a robust solution for text entry in VR. These findings highlight the potential of VRWS to transform virtual office experiences, making them more efficient and enjoyable for users while paving the way for future advancements in VR technology.

8.2 Limitations

This study has several limitations that need to be addressed, which will guide future research directions.

a) Duration of Experiments:

   In the VRWS experiments, each session was conducted over a relatively short period. This limited timeframe may not have been sufficient to observe significant differences in the number of correct answers or time differences between the experimental conditions. Longer sessions could provide more comprehensive insights into user behavior and performance under varying conditions.

b) Sample Size:

   The number of participants in this study was limited to a small group of 20 participants. This relatively small and homogeneous sample size restricts the generalizability of the findings to broader populations. A larger and more diverse participant pool is necessary to validate the results and ensure their applicability across different demographics, such as varying age groups, typing proficiency levels, and VR experience.

c) Simulated Environment:

   The use of the CAVE system to simulate a noisy OPWS environment introduces limitations. While the CAVE system effectively mimics certain characteristics of OPWS, it may not fully replicate the real-world distractions and dynamics present in actual OPWS. This reduced realism may influence the ecological validity of the results. Future research should compare simulated VRWS and real-world OPWS environments to better understand the differences and validate findings.

d) HMD and Noise-Cancelling Headphones as Combined Variables:

   In this study, the HMD and noise-cancelling headphones were used simultaneously as part of the experimental setup. Their effects were not evaluated as separate variables, which limits the understanding of how each component individually contributes to user performance and comfort. For example, it remains unclear whether the results were more influenced by the visual isolation provided by the HMD, or the auditory isolation provided by the headphones. This lack of separation limits the interpretability of the findings and should be addressed in future experiments.

e) HMD Design and Weight:

The HMD used in the experiment was relatively heavy, which may have contributed to user discomfort and potentially affected performance. Lightweight and ergonomically designed HMDs could mitigate this issue in future studies.

f) Lack of Focus on Communication:

While VRWS aims to simulate and enhance office work environments, this study did not address communication aspects such as interactions between participants in the VR environment. Communication is a critical component of remote work, and its exclusion represents a significant limitation. Future research should integrate and evaluate communication tools within VRWS to provide a more comprehensive analysis.

g) Algorithm Generalizability:

In the 2S-LSTM component, the sample size for the performance evaluation of the typing solution was also limited. Additionally, the dataset used for training and testing the model was relatively specific, which may limit the generalizability of the findings. Expanding the dataset to include a broader range of scenarios and user typing behaviors would strengthen the robustness of the results.

8.3 Future Work

Building on the limitations identified, future research will focus on addressing these challenges and exploring additional opportunities to enhance VRWS and typing solutions.

a) Extended Experiment Duration:

To observe long-term effects and provide more accurate assessments, future studies will conduct experiments over extended periods. This will help reveal how user performance and adaptation evolve over time in VRWS.

b) Larger and More Diverse Participant Pool:

Expanding the participant pool to include a larger and more diverse group will ensure findings are generalizable across different demographics. This will also allow for more nuanced analyses of factors such as age, typing proficiency, and familiarity with VR technology.

c) Comparison of Simulated and Real OPWS:

Future studies will compare real-world OPWS environments with VRWS simulations to evaluate how closely the latter can replicate real-world dynamics. This will provide insights into the practicality and effectiveness of VRWS for professional use.

d) Independent Evaluation of HMD and Noise-Cancelling Headphones:

Future experiments will separate the effects of HMD and noise-cancelling headphones as independent variables. By isolating their contributions, researchers can better understand how visual and auditory isolation individually influences user performance, comfort, and overall experience in VRWS. For example, future work could involve experiments with:

- ✧ HMD only (without noise-cancelling headphones),
- ✧ Noise-cancelling headphones only (without HMD), and
- ✧ Both devices are combined. These comparisons will offer a more detailed understanding of the role each component plays in immersive office environments.

e) Improved HMD Design:

Future work will explore the use of lightweight HMDs or alternative designs that reduce physical strain while maintaining high imaging performance. This will address user comfort and potentially improve performance outcomes.

f) Integration of Communication Tools:

Communication is a fundamental aspect of collaborative work. Future research will explore how VRWS can facilitate real-time communication among users, incorporating tools such as virtual meeting spaces, chat systems, and collaborative whiteboards. This will provide a more holistic evaluation of VRWS as a remote work solution.

g) Algorithm Refinement:

Building on the success of the 2S-LSTM-KF model, future work will investigate the potential of integrating more advanced machine learning techniques, such as 2S-BNNActionNet-KF, to further improve typing accuracy and efficiency. Additionally, adapting algorithms to account for individual differences in typing habits will enhance the user experience.

h) Exploration of Long-term Adaptation:

While this study focused on short-term adaptation to VR typing, future research will investigate how typing habits evolve over extended periods of VR use. Understanding long-term learning effects will provide valuable insights into designing more intuitive and efficient typing systems.

i) Transition to ARWS:

Future developments may explore transitioning VRWS to ARWS (Augmented Reality Workspace) by incorporating cameras into the HMD. This would allow users to seamlessly interact with physical objects like keyboards, paper, and pens, enhancing the practicality of the system without compromising immersion.

j) Incorporation of other Non-Keyboard Input Support:

To address the demands of teamwork and discussions, future research will explore integrating non-keyboard input support into VRWS. For instance, developing functionalities that detect participants' gaze directions during discussions in real time could enhance collaboration efficiency and interaction experience. These improvements will further expand VRWS's applications and enhance its overall efficiency.

By addressing these limitations and pursuing the proposed directions, future research can further advance VRWS and typing solutions, making them more practical, efficient, and accessible to a wider range of users.

Bibliography

[1] D. M. Cooper. Remote laboratories in teaching and learning - issues impinging on widespread adoption in science and engineering education. Int. J. Online Eng., vol. 1, (2005)

[2] Y. Elawady, A. Tolba. Educational Objectives of Different Laboratory Types: A Comparative Study. ArXiv, (2009)

[3] F. Brooks. What's real about virtual reality. IEEE Computer Graphics and Applications, vol. 19, no. 6, pp. 16-27, (1999)

[4] M. Shirer, S. Soohoo. Worldwide Spending on Augmented and Virtual Reality Forecast to Deliver Strong Growth Through 2024, According to a New IDC Spending Guide, (2020)

[5] C. Imperatori, A. Dakanalis, and B. Farina. Global storm of stress-related psychopathological symptoms: a brief overview on the usefulness of virtual reality in facing the mental health impact of COVID-19. Cyberpsychology, Behavior, and Social Networking, vol. 23, no. 11, pp. 782-788, (2020)

[6] Q. Zhao. A survey on virtual reality. Sci China Ser F: Inf Sci 2009, 52(3), pp. 348–400, (2009)

[7] G. Burdea, P. Coiffet. Virtual reality technology. New York: Wiley-IEEE Press, (2003)

[8] Mazuryk T, Gervautz M. Virtual reality-history, applications, technology and future. Vienna: Institute of Computer Graphics, (1996)

[9] M. Mihelj, D. Novak, V. Begus. Virtual Reality Technology and Applications. Berlin: Springer, (2014)

[10] M. Kizil, J. Joy. What can virtual reality do for safety. St. Lucia QLD: University of Queensland, (2001)

[11] P. Milgram, F. Kishino. A taxonomy of mixed reality visual displays. IEICE Trans. Inf & Syst., vol. 77, no. 12, pp. 1321-1329, (1994)

[12] P. Stothard, A. Squelch, and R. Stone. Taxonomy ofinteractive computer-based visualisation systems and content for the miningindustry – part 2. Mining Technology, vol. 124, no. 2, pp. 83-96, (2015)

[13] M. Slater, M. Sanchez-Vives. Enhancing Our Lives with Immersive Virtual Reality, Frontiers Robotics AI, vol. 3, pp. 74, (2016)

[14] F. Brooks. What's real about virtual reality. IEEE Computer Graphics and Applications, vol. 19, no. 6, pp. 16-27, (1999)

[15] S. Chan, L. Qiu, G. Esposito, K. Mai, K. Tam, and J. Cui. Nature in virtual reality improves mood and reduces stress: evidence from young adults and senior citizens, Virtual Reality, pp. 1-16, (2021)

[16] G. Reese, J. Stahlberg, and C. Menzel. Digital shinrin-yoku: do nature experiences in virtual reality reduce stress and increase well-being as strongly as similar experiences in a physical forest? Virtual Reality, vol. 26, pp. 1245-1255, (2021)

[17] M. Davis, D. Can, J. Pindrik, B. Rocque, and J. Johnston. Virtual Interactive Presence in Global Surgical Education: International Collaboration Through Augmented Reality, World Neurosurgery, vol. 86, pp. 103-111, (2016)

[18] R. Jackson, E. Fagan. Collaboration and learning within immersive virtual reality, Proceeding of the International Conference on Collaborative Virtual

Environments, pp. 83-92, (2000)
[19] R. Rahman, M. Islam. VREd: A Virtual Reality-Based Classroom for Online Education Using Unity3D WebGL. ArXiv, (2023)
[20] B. Cho, J. Ku, D. Jang, S. Kim, Y. Lee, I. Kim, J. Lee, and S. Kim. The Effect of Virtual Reality Cognitive Training for Attention Enhancement, Cyberpsychology & Behavior, vol. 5, no. 2, pp. 129-137, (2002)
[21] X. Yang, L. Lin, P. Cheng, X. Yang, Y. Ren, and Y. Huang. Examining creativity through a virtual reality support system, Educational Technology Research and Development, vol. 66, pp. 1231-1254, (2018)
[22] D. Feuereissen. VR: Getting the Reality Part Straight, *Does Jitter and Suspension of the Human Body Increase Auditory Circular Vection*, Department of Computer Science in Media, (2008)
[23] P. Caserman, A. Garcia-Agundez, A. Zerban, and S. Göbe. Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. Virtual Reality, vol. 25, pp. 1153-1170, (2021)
[24] M. Mcgill, R. Murray-Smith, D. Boland, and S. Brewster. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, (2015)
[25] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. Kristensson. Effects of Hand Representations for Typing in Virtual Reality. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp.151-158, (2018)
[26] J. Stauffert, F. Niebling, and M. Latoschik. Effects of Latency Jitter on Simulator Sickness in a Search Task, 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),pp. 121-127, (2018)
[27] Y. Tanaka. VR Easy to Make Friends and Useful for Work? What I Realized by Having Meetings in VR. Next-System, 2018.
[28] Y. Fushimi. Japanese Input in VR Using Hand Controllers. Qiita, https://qiita.com/yutoVR/items/a80bd382f9b52b786a8a, [retrieved: 11, 2023]
[29] J. Gruber et al. Effects of Hand Representations for Typingin Virtual Reality. IEEE VR 2018 publication, pp. 151-158, (2018)
[30] https://store.steampowered.com/app/438100/VRChat/, [retrieved: 11, 2023]
[31] https://www.moguravr.com/virtual-desktop-4/, [retrieved: 11,2023]
[32] W. Ang, P. Pradeep, and C. Riviere. Active tremor compensation in handheld instrument for microsurgery. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA, (2004)
[33] R. Elble, R. Sinha, and C. Higgins. Quantification of tremor with a digitizing tablet. J Neurosc Methods vol. 32, no. 3, pp. 193-198, (1990)
[34] H. Hefter, V. Homberg, and K. Reiners. Stability of frequency during long-term recordings of hand tremor. Electroencephalogr Clin Neurophysiol, vol. 67, no. 5, pp. 439-446, (1987)
[35] Fang W, Zheng L, Deng H, Zhang H. Real-time motion tracking for mobile augmented/virtual reality using adaptive visual-inertial fusion. Sensors, vol. 17, no. 5, pp.1037, (2017)
[36] O. Junho, L. Sanghwa, L. Boohwan, and P. Jongil Park JI. Probability analysis of position errors using uncooled IR stereo camera. Infrared Physics &

Technology, vol. 76, pp. 346-352, (2016)
[37] J. Guna, G. Jakus, M Pogacnik, S. Tomazic, and J. Sodnik. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors, vol. 14, no. 2, pp. 3702-3720, (2014)
[38] C. Xi, J. Chen, C. Zhao, Q. Pei, L. Liu. Real-time hand tracking using kinect. Proceedings of the 2nd international conference on digital signal processing, pp. 37-42, (2018)
[39] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text entry in immersive head-mounted display-based virtual reality using standard keyboards, 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 159-166, (2018)
[40] D. Bowman, C. Rhoton, and M. Pinho. Text input techniques for immersive virtual environments: An empirical comparison. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol 46, pp. 2154-2158, (2002)
[41] D. Bowman, E. Kruijff, and J. LaViola. 3D User Interfaces: Theory and Practice. Addison Wesley Longman Publishing Co., Inc, (2004)
[42] C. Roelofsen. Performance loss in open-plan offices due to noise by speech. Journal of Facilities Management, vol. 6, no. 3, pp. 202-211, (2008)
[43] O. Vinesh. Why your office could be making you sick. Asia-Pacific Journal of Health Management , (2009)
[44] T. Julian. The 4 ways sound affects us. TED.com, TED Talks, TED, https://www.ted.com/talks/julian_treasure_the_4_ways_sound_affects_us, [retrieved: 11,2023]
[45] Xymax Real Estate Institute. Office tenants of office buildings in Tokyo 23 Wards. Research Report Xymax Real Estate Institute, (2016)
[46] G. Evans, D. Johnson. Stress and open-office noise. Journal of Applied Psychology, vol. 85, no. 5, pp. 779-783, (2000)
[47] M. Humphries. Quantifying occupant comfort: Are combined indices of the indoor environment practicable? Building Research and Information, vol. 33, no. 4, pp. 317-325, (2005)
[48] J. Veitch, K. Charles, and G. Newsham. Workstation characteristics and environmental satisfaction in open-plan offices: COPE Field Findings. 65th Annual Conference of the Canadian Psychological Association, pp 1-4, (2004)
[49] R. Karasek, T. Theorell. Healthy work: Stress, productivity and the reconstruction of working life. New York: Basic Books. (1990)
[50] M. Kompast, I. Wagner. Telework: managing spatial, temporal and cultural boundaries. Teleworking, pp. 115-137, (2002)
[51] V. Biener, D. Schneider, and T. Gesslein. Breaking the Screen: Interaction across Touchscreen Boundaries in Virtual Reality for Mobile Knowledge Workers. IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 12, pp. 3490-3502, (2020)
[52] E. Kim, G. Shin, G. User discomfort while using a virtual reality headset as a

personal viewing system for text-intensive office tasks. Ergonomics, vol. 64, no. 7, pp. 891-899, (2021)

[53] R. Shen, D. Weng, and S. Chen. Mental fatigue of long-Term office tasks in virtual environment. Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2019, pp. 124-127, (2019)

[54] C. Boletsis, S. Kongsvik. Text Input in Virtual Reality: A Preliminary Evaluation of the Drum-Like VR Keyboard, Technologies, vol. 7, no. 2, pp. 31, (2019)

[55] A. Otte, D. Schneider, and T. Menzner. Evaluating Text Entry in Virtual Reality using a Touch-sensitive Physical Keyboard, 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 387-392 (2019)

[56] N. Fourrier, G. Moreau, and M. Benaouicha. Handwriting for Efficient Text Entry in Industrial VR Applications: Influence of Board Orientation and Sensory Feedback on Performance. IEEE Transactions on Visualization and Computer Graphics, vol. 29, no. 11, pp. 4438-4448 (2023)

[57] M. Meier, P. Streli, and A. Fender. TapID: Rapid Touch Interaction in Virtual Reality using Wearable Sensing, 2021 IEEE Virtual Reality and 3D User Interfaces (VR), pp. 519-528, (2021)

[58] H. Gil, I. Oakley. ThumbAir: In-Air Typing for Head Mounted Displays, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, pp. 1-30, (2023)

[59] K. Taejun, A. Karlson, and A. Gupta. STAR: Smartphone-analogous Typing in Augmented Reality. UIST '23: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1-13, (2023)

[60] D. Hwang, K. Aso, and H. Koike. MonoEye: Monocular Fisheye Camera-based 3D Human Pose Estimation. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp.988-989, (2019)

[61] B. Gavin. Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. Virtual Reality and Human Behaviour, vol 2, (2021)

[62] H. Wang, B. Wang , and B. Liu. Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle. Robotics and Autonomous Systems, vol. 88, pp.71-78, (2017)

[63] L. Abraham, A. Urru, and N. Normani. Hand Tracking and Gesture Recognition Using Lensless Smart Sensors. Special Issue Emerging Algorithms and Applications in Vision Sensors System based on Artificial Intelligence, (2018)

[64] U. Kim , S. Yoo , and J. Kim. I-Keyboard: Fully Imaginary Keyboard on Touch Devices Empowered by Deep Neural Decoder, IEEE Transactions on Cybernetics, vol. 51, no. 9, (2021)

[65] M. Zaman, C. Shekhar, and J. Hwang. Effects of Virtual Hands on Physical Demands and Task Performance for Typing in Virtual Reality, Conference: The XXXIVth Annual International Occupational Ergonomics and Safety Conference, (2022)

[66] R. Teather, A. Pavlovych, and W. Stuerzlinger. Effects of tracking technology, latency, and spatial jitter on object movement In 3D User Interfaces, 2009.

3DUI 2009. IEEE Symposium on, pp.43–50, (2009)
[67] A. Pavlovych and W. Stuerzlinger. The tradeoff between spatial jitter and latency in pointing tasks. In proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems, pp.187-196. ACM, (2009)
[68] A. U. Batmaz, M. R. Seraji, J. Kneifel, and W. Stuerzlinger, “No jitter please: Effects of rotational and positional jitter on 3D mid air interaction,” In Future Technologies Conference, vol. AISC 1289. Springer, (2020).
[69] M. Mughrabi, A. Mutasim, and W. Stuerzlinger. My Eyes Hurt: Effects of Jitter in 3D Gaze Tracking. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshop, pp.12-16, (2022)
[70] A. U. Batmaz, M. R. Seraji, and J. Kneifel. No jitter please: Effects of rotational and positional jitter on 3D mid air interaction, In Future Technologies Conference, vol. AISC 1289 (2020)
[71] M. Ribo, A. Pinz, and A. Fuhrmann. A new optical tracking system for virtual and augmented reality applications. In Proceedings of IEEE Instrumentation and Measurement Technology Conference (IMTC), vol. 3, pp. 1932-1936, (2001)
[72] R. Teather, W. Stuerzlinger. Pointing at 3D targets in a stereo head-tracked virtual environment.In Proceedings of IEEE Symposium on 3D User Interfaces (3DUI), pp. 8-94, (2011)
[73] C. P. G. Roelofsen, "Performance loss in open-plan offices due to noise by speech", Journal of Facilities Management, vol. 6, no. 3, pp. 202-211 (2008)
[74] Vinesh Oommen, "Why your office could be making you sick", Asia-Pacific Journal of Health Management, (2009)
[75] Treasure, Julian. "The 4 ways sound affects us". TED.com. TED Talks. TED. [retrieved: 11, 2023]
[76] C. Candido, P. Chakraborty, and D. Tjondronegoro. The Rise of Office Design in High-Performance, Open-Plan Environments. Buildings 2019, vol. 9, no. 4, pp. 100, (2019)
[77] Xymax Real Estate Institute, "Office tenants of office buildings in Tokyo 23 Wards", Research Report| Xymax Real Estate Institute, September 21, 2016.
[78] https://www.theguardian.com/technology/2017/jan/05/i-tried-to-work-all-day-in-a-vr-headset-so-you-never-have-to, [retrieved: 11, 2023]
[79] https://www.youtube.com/watch?v=Y0yOTanzx-s, [retrieved: 11, 2023]
[80] C. Osgood. Semantic differential technique in the comparative study of cultures. American Anthropologist, vol. 66, no. 3, pp. 171-200, (1964)
[81] N. Mituo, S. Isao, and I. Rituko. The study of emotional engineering, human engineering, (1972)
[82] S. Ellbin, N. Engen, and I. Jonsdottir. Assessment of cognitive function in patients with stressrelated exhaustion using the Cognitive Assessment Battery (CAB). Journal of Clinical and Experimental Neuropsychology, vol. 40, no. 6, pp. 567-575, (2018)
[83] A. Nordlund, L. Pahlsson, and C. Holmberg. The Cognitive Assessment Battery (CAB): a rapid test of cognitive domains. International Psychogeriatrics, vol. 23, no. 7, pp. 1144-1151, (2011)
[84] W. Wang, K. Yu, and J. Hugonot. Beyond One Glance: Gated Recurrent Architecture for Hand Segmentation, arxiv, (2018)

[85] A. Bobick, J. Davis. The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257-267, (2001)
[86] D. Tsai, W. Chiu, and M. Lee. Optical Flow-Motion History Image (OF-MHI) for Action Recognition. Signal, Image and Video Process, vol. 9, pp. 1897-1906, (2015)
[87] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, pp. 1-14, (2014)
[88] K. He, X. Zhang, and S. Ren. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, (2016)
[89] S. Hochreiter, J. Schmidhuber. Long Short-Term Memory. Neural Computation, vol. 9, no. 8, pp. 1735-1780, (1997)
[90] G. Welch, G. Bishop. An Introduction to the Kalman Filter. University of North Carolina: Chapel Hill, (1995)
[91] H. Coskun, F. Achilles, and R. DiPietro. Long Short-Term Memory Kalman Filters: Recurrent Neural Estimators for Pose Regularization. 2017 IEEE International Conference on Computer Vision (ICCV), pp. 5525-5533, (2017)
[92] F. Zhang, V. Bazarevsky, and A. Vakunov. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv 2020, arXiv:2006.10214, (2020)
[93] GPUScore Legacy Products. Available online. https://www.gpuscore.com/benchmarks/legacy-products/, [retrieved: 11,2023]
[94] Y. Tatsunami, M. Masato Taki. Sequencer: Deep LSTM for Image Classification. arXiv 2020, arXiv:2205.01972, (2020)
[95] Y. Nie, N. Nguyen, and P. Sinthong. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. arXiv 2022, arXiv:2211.14730, (2022)
[96] F. Fontana,A. Matteo, and L. Cinque. BNNAction-Net: Binary Neural Network on Hands Gesture Recognitions. In Proceedings of the ACM SIGGRAPH 2024 Posters (SIGGRAPH'24), no. 52, pp. 1-2, (2024)
[97] S. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Computation. 1997, 1735–1780.
[98] D. Simon, N. Keith, and N. Eugene. A Systematic Review of Cybersickness. In Proceedings of the 2014 Conference on Interactive Entertainment, pp. 1-9, (2014)
[99] X. Hou, Y. Lu, and S. Dey. Wireless VR/AR with Edge/Cloud Computing. In Proceedings of the 2017 26th International Conference on Computer Communication and Networks (ICCCN), pp. 1-8, (2017)
[100] J. Jerald. Scene-Motion- and Latency-Perception Thresholds for Head Mounted Displays. Ph.D. Thesis, University of North Carolina, (2009)
[101] T. Xu, W. Gu, and K. Ota. A Low-Jitter Hand Tracking System for Improving Typing Efficiency in Virtual Reality Workspace. In Proceedings of the TENCON 2023—2023 IEEE Region 10 Conference (TENCON), pp. 1-6, (2023)

Publications

T. Xu, W. Gu, K. Ota, and S. Hasegawa, "Development and Evaluation of Low-Jitter Hand Tracking System for Improving Efficiency in Virtual Reality Workspace," Multimodal Technologies and Interaction, vol.9, no. 1, pp. 4, 2025.

T. Xu, and S. Hasegawa, "A Study on Virtual Reality Work-Space to Improve Work Efficiency," The Thirteenth International Conference on Advances in Computer-Human Interactions (ACHI 2020,), Valencia, Spain, pp. 304-309, 2020.

T. Xu, W. Gu, K. Ota, and S. Hasegawa, "A Low-Jitter Hand Tracking System for Improving Typing Efficiency in Virtual Reality Workspace," TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), Chiang Mai, Thailand, pp. 1-6, 2023.

T. Xu, and S. Hasegawa, "A Study on Virtual Reality Work-Space to Improve Work Efficiency," Japanese Society for Information and Systems in Education (JSISE 2018), pp. 23-24, 2018.