AnkleType: A Hands- and Eyes-free Foot-based Text Entry Technique in Virtual Reality

Xiyun Luo
Department of Computer Science
Shantou University
Shantou, China
24xbluo@stu.edu.cn

Weirong Luo
Department of Computer Science
Shantou University
Shantou, China
24wrluo@stu.edu.cn

Kening Zhu*
School of Creative Media
City University of Hong Kong
HongKong, China
keninzhu@cityu.edu.hk

Taizhou Chen†
Department of Computer Science
Shantou University
Shantou, China
taizhou.chen@my.cityu.edu.hk

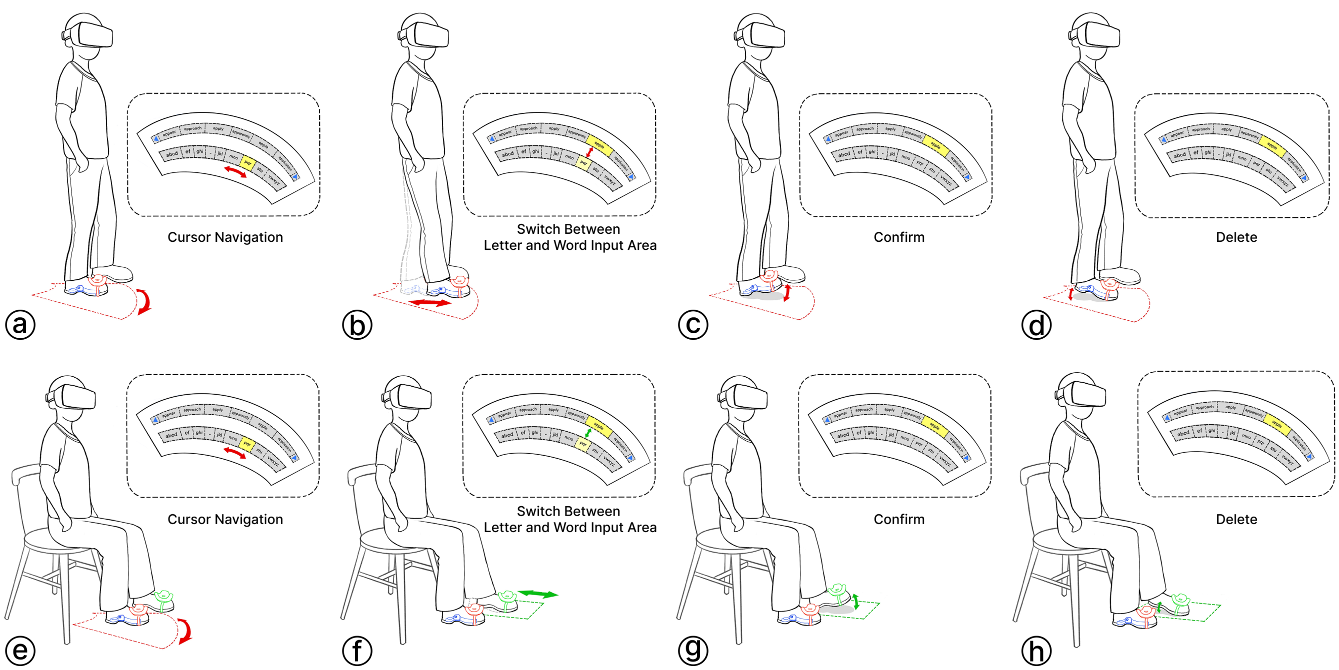


Figure 1: AnkleType enable eye- and hand-free foot-based typing for standing (UPStand ⓐ - ⓓ) and sitting posture (BPSit ⓔ - ⓗ). ⓐⓔ Rotate the ankle for cursor navigation, ⓑⓕ *Foot flat forward* and *Foot flat backward* gesture for switching between the letter input area and word input area, ⓒⓖ *Forefoot single tap* gesture to confirm the input, ⓓⓗ *Rearfoot single tap* to delete the inputted letter or word.

Abstract

Virtual Reality (VR) emphasizes immersive experiences, while text entry often requires hands or visual attention, which may disrupt the interaction flows in VR. We present AnkleType, a hand- and eye-free text-entry technique that leverages ankle-based gestures for both standing and sitting situations. We began with two preliminary studies: one investigated the movement range of users' ankles, and the other elicited user-preferred ankle gestures for text-entry-related operations. The findings of these two studies guided

*He is also with City University of Hong Kong Shenzhen Research Institute, Shenzhen, China.
†Corresponding author.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790999

our design of AnkleType. To optimize AnkleType's keyboard layout for eye-free input, we conducted a user study to capture the users' natural ankle spatial awareness with a computer-simulated language test. Through a pairwise comparison study, we designed a bipedal input strategy for sitting (**BPSit**) and a unipedal input strategy for standing (**UPStand**). Our first in-VR text-entry evaluation with 16 participants demonstrated that our methods could support the average typing speed from 8.99 WPM (**BPSit**) to 9.13 WPM (**UPStand**) for our first-time users. We further evaluated our design with a 7-day longitudinal study with twelve participants. Participants achieved an average typing speed of 15.05 WPM with **UPStand** and 16.70 WPM with **BPSit** in the visual condition, and 11.15 WPM and 12.87 WPM, respectively in the eyes-free condition.

CCS Concepts

• **Human-centered computing → Virtual reality**; **Text input**; *Keyboards*; *Empirical studies in HCI*.

Keywords

Text Entry, Foot-based Interaction, Virtual Reality

ACM Reference Format:
Xiyun Luo, Weirong Luo, Kening Zhu, and Taizhou Chen. 2026. AnkleType: A Hands- and Eyes-free Foot-based Text Entry Technique in Virtual Reality. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3772318.3790999

1 Introduction

Text entry serves as a fundamental activity in a computing system to support communication, inquiry, annotation, and documentation across a variety of tasks [8]. As the rapid development of hardware capability and the increasing portability of head-mounted display (HMD) has moved Virtual Reality and Augmented Reality (VR/AR) toward everyday applications such as gaming [11, 29, 48], creativity [50], remote collaboration [23, 71], and social applications [12, 25, 34], the ability to input text efficiently and comfortably in a VR/AR system becomes increasingly important. However, conventional methods such as physical keyboards suffer from the reduction of typing performance when adapted to VR environments [24], highlighting the need to investigate more intuitive and expressive text entry techniques for the VR context.

To mitigate this gap, researchers and designers have explored a variety of VR text entry techniques, including the adaptation of handheld controllers [36, 75], head gestures [40, 81], hand gestures [16], virtual keyboards [2, 60], and speech [1]. While these approaches have made significant progress to support text entry in VR, they face limitations when users' hands or visual attention are less available. For example, when a user is holding objects physically or virtually using a controller, their hands become less available for typing (Figure 10). Head-based and eye-gaze interaction, in particular, are prone to induce motion sickness meanwhile causing eye fatigue, which may increase physical and cognitive burden during long-term usage [44, 57, 69]. Beyond these approaches, recent research [68] has pioneeringly explored foot-based text entry techniques for VR, demonstrating the feasibility and potential of leveraging the feet as a text input channel in VR. While this method provides a hands-free typing experience with low visual burden, it still requires users' attention on staring at the keyboard layout during typing and requires a wide range of motion with the user's feet. Furthermore, their study primarily focused on the sitting condition, as they pointed out that users do not prefer typing for a long time while standing. However, for those applications that require a standing posture, such as gaming, it is inevitable to type while standing, highlighting the need to investigate an optimal typing interface for standing scenarios.

In this paper, we propose AnkleType, a novel foot-based text entry technique for hand- and eye-free text entry in VR. The system employs ankle rotations for navigation with forefoot and rearfoot taps to perform control actions such as confirm and cancel. AnkleType is explicitly designed to support both standing and sitting postures, offering an efficient and natural typing experience with only minimal foot movement. To explore the design space of this technique, we focused on two key aspects: keyboard layout and interaction design, and investigated and evaluated them through a series of user studies.

We first conducted a preliminary elicitation study to learn users' preferences and performance for ankle-based foot interaction using unipedal and bipedal input in both standing and sitting postures. Based on the observation of differences in ankle rotation range between postures, we explored two separate keyboard layouts for standing and sitting postures, respectively. To this end, we run a user study with a series of computer-simulated typing tests to compare the performance of various design options, yielding one optimal layout for each posture that balances eye-free feet reachability and language disambiguation. Building on this, we conducted a second study to identify the most effective interaction mechanisms for each posture. Inspired by our preliminary elicitation study, we designed and compared four user-defined mechanisms. The results showed that **unipedal stand (UPStand)** and **bipedal sit (BPSit)** offered the best balance of typing efficiency, comfort, and usability, as the final designs for AnkleType. The results also show that our methods could support the average typing speed from 7.80 WPM to 9.44 WPM for our first-time users. Finally, we run a 7-day longitudinal study to test the system's learnability and the eye-free performance. For the visual condition, result shows that AnkleType achieves an average typing speed of 15.05 WPM for **unipedal stand (UPStand)** and 16.70 WPM for **bipedal sit (BPSit)**, with a total error rate of 3.71% and 2.48%, respectively. For the blind condition, result shows that AnkleType achieves an average typing speed of 11.15 WPM for **unipedal stand (UPStand)** and 12.87 WPM for **bipedal sit (BPSit)**, with a total error rate of 9.91% and 8.80%, respectively. The results also show that AnkleType not only outperformed state-of-the-art hand-free VR text entry techniques, but also achieved competitive performance in the eye-free VR text entry task.

Our contributions are fourfold:

- We propose a novel ankle-based typing technique that enables hands- and eyes-free text entry in VR, supporting both standing and sitting postures.
- We present an elicitation exploration that identifies users' preferences to perform ankle-based foot gestures in both standing and sitting conditions.
- We perform a series of empirical explorations to optimize the keyboard layout and the interaction design of the proposed technique across postures.

- We conduct a longitudinal user study to demonstrate the learnability and the effectiveness of the proposed technique.

2 Related Works

AnkleType is largely inspired by existing works on immersive text entry techniques and foot-based interfaces. In addition, we also refer to existing eye-free text entry techniques.

2.1 Text Entry for Immersive Environment

Text entry in immersive virtual environments has been explored for decades [6]. Previous research reveals that text entry in immersive head-mount display using a conventional keyboard [24] suffers from the low entry speed and high error rate problem. To mitigate this gap, researchers have explored many novel text entry techniques, including handheld controller [36, 75], head gesture [40, 81], hand gesture [16], and virtual keyboard [2, 60]. While the solutions are diverse, we summarized and discussed them from two aspects: controller-based and non-controller-based.

2.1.1 Controller-Based Techniques. Type with VR hand-held controller is one of the straightforward and popular input strategies, attracting various previous explorations. Among them, typing through controller ray casting is the most common input method, but it is less efficient and prone to increasing physical tedium during long-term usage. Therefore, prior research has investigated supporting more expressive and efficient text entry methods using controller motion, such as pointing or tapping. Speicher et al. [59] explored a variety of VR text entry techniques, including a method by reversing and holding the HTC Vive controller for tapping, achieving an input speed of 12.96 WPM. Leng et al. [36] redesigned the keyboard layout for controller typing by arranging letters in a flower-shaped layout to reduce the controller pointing movement. Their method achieved an impressive typing performance of 22.97 WPM after a 6-day practice.

Beyond using controller motions, prior research has leveraged joysticks and touchpads on a conventional VR hand-held controller to support a variety of text entry techniques. PizzaText [82] uses dual joysticks to input on a "pizza" layout, with a speed of 15.85 WPM. HiPad [27] designs a six-key radial layout on the touchpad, achieving an input speed of 13.57 WPM with single-handed input. FanPad [75] maps QWERTY to dual touchpads to support rapid text entry with subtle motions. By optimizing the thumb trajectory, their method achieves a text entry speed of 19.73 WPM.

2.1.2 Non-Controller-Based Techniques. Numerous researchers have dedicated themselves to exploring non-controller-based text entry techniques, such as using hand or head gestures. Mid-air typing, as one of the intuitive text entry strategies in immersive environments that adapt to users' traditional behaviors, has been widely explored by the research community. However, although natural, air typing is suffering from a problem of lacking support and tactile feedback, which can easily lead to arm fatigue and limit input efficiency [2, 80]. Dudley et al. [15] mitigated this problem by projecting a virtual keyboard onto a physical surface, which significantly improved the text entry speed by around 50% compared to mid-air typing. TouchInsight [60] further improved the typing performance on a physical surface by refining the touch area tracking accuracy, and achieved a typing speed of 37.54 WPM. Beyond fixed surfaces, researchers explored transforming tangible props [18, 61], arbitrary surfaces[14], and the body [13, 31, 35] into tactile typing interfaces.

Researchers have also explored integrating physical keyboards into immersive environments. Grubert et al. [24] explored adapting standard keyboards, including physical keyboards and tablet soft keyboards, for use with an immersive head-mounted display. Their findings revealed a certain gap in typing speed compared to typing in the real world. Researchers have explored combining with other input modalities, such as gaze and speech, to mitigate this problem. Adhikary et al. [1] combine speech with physical keyboard input and showed a significant improvement in input speed and error rate, but voice-based input is prone to suffer from privacy problems. Kalamkar et al. [28] combined gaze with a physical keyboard to improve the efficiency of special character input, but it increases the user's visual burden. HawKEY [51] further investigated improving the flexibility of VR physical keyboard typing with a hawker's tray to support typing while standing.

Beyond the physical keyboard, researchers also investigated hands-free typing techniques, such as using eye movements. Eye-based typing shows an advantage in HMD typing, as it relies only on natural eye movements without requiring additional physical effort. However, gaze-based typing normally requires a dwell action to trigger the input event, which would reduce the input speed [40, 81]. RingText [76] mitigated this problem by exploring a circular keyboard layout to improve the text entry speed with gaze without dwell. Hu et al. [26] further explored the dwell-free eye typing technique on a virtual QWERTY keyboard in mixed reality. iText [39] pushed the limit of eye-based typing techniques by exploring typing on an invisible keyboard. Their solution achieved a text entry speed of 13.77 WPM without requiring too much eye attention. Apart from eye-based typing, Wan et al. [68] recently pioneered the use of foot-based gestures for text entry in VR. Their study focused on evaluating tap-based and swipe-based text entry techniques, and the results show that bipedal tap-based input yields the best performance of 11.12 WPM in sitting posture.

AnkleType follows Wan et al. [68]'s concept of foot-based typing, and extends this concept from two perspectives. First, we consider that VR applications, such as gaming, often require a standing posture. Therefore, we optimized the foot-based typing experience to support both sitting and standing conditions. Second, we consider that VR applications mostly demand users' visual attention. Therefore, our explorations focus on enabling eye-free text entry in virtual scenes.

2.2 Foot-based Interfaces

We also drew inspiration from existing foot-based interfaces in the design of AnkleType. Foot-based input had been widely explored as an alternative interaction modality, particularly in scenarios where both hands are occupied[67, 83]. Prior research had examined diverse approaches, including pressure-based interaction [30, 62], heel on-floor rotation [85], ankle mid-air rotation [56], and foot tapping interaction [54]. Yasmin et al. [17] further conducted a gesture elicitation study on foot-based gestures to identify the potential of foot-based gestures to support avatar control and GUI control.

In the VR context, users' hands are often occupied by the handheld controller or other hand-based tasks such as gesturing. Therefore, foot-based interactions were considered suitable for VR applications and had been explored in recent years' research. Christopher et al. [4] conducted a gesture elicitation to explore user-defined foot gestures for AR map interactions, while Shih et al. [58] designed a floor-projected radial menu interaction with semi-transparent avatars to address occlusion. Similarly, Müller et al.[46] investigated foot-tapping input for HMDs by comparing direct floor projections with indirect floating interfaces. Their findings reveal that although direct interaction yields higher accuracy, users prefer indirect interaction due to improved comfort and reduced neck fatigue. Recently, Chan et al. [9] proposed Seated-WIP, a simple yet inspiring VR locomotion technique using foot stepping. Their design distinguished forefoot stepping and rearfoot stepping to support moving forward and backward with low fatigue required and high input efficiency.

AnkleType is largely inspired by the aforementioned literature focusing on foot-based interaction. On one hand, the interface design of AnkleType draws inspiration from Foot Menu [85], where they design a radial-based foot menu to support heel rotation interaction. On the other hand, we referred to Seated-WIP's [9] hardware solution when implementing AnkleType.

2.3 Eye-free Text Entry Techniques

We also refer to prior eye-free text entry research when designing AnkleType. Eye-free text entry refers to entering text without requiring visual attention to the keyboard or finger/cursor movement [41, 63, 78]. Literature has explored supporting eye-free text entry across various contexts, such as smartphone [79, 86], tablet [37], VR [19], and wearable devices [77, 78]. For example, Zhu et al. [86] explored supporting typing on a smartphone with an invisible keyboard, achieving an average typing of 37.8 WPM after 3 days of training. Li et al. [37] extend this concept to tablets and present ResType, a system that adaptively adjusts the invisible keyboard placement to facilitate eye-free text entry, with a text entry speed of 41.6 WPM. Gil et al. [19] further extend this concept to VR and propose an invisible keyboard to support eye-free mid-air typing, with an average input speed of 41.6 WPM. Eye-free typing has also been explored for wearable devices, as mobile application contexts often require users' attention on their surroundings. TipText [78] and BiTipText [77] support eye-free thumb-to-index finger typing on a QWERTY layout keyboard, achieving 13.3 WPM and 25 WPM, respectively. Furthermore, TypeAnywhere [84] enables ten-finger QWERTY typing on any surface by decoding finger-tap sequences, achieving 70.6 WPM. Other approaches, such as BlindType [41] and i'sFree [87], support eye-free remote typing by using a handheld touchpad or a smartphone as a typing proxy without requiring users' attention on it.

The design concepts and approaches of AnkleType are largely inspired by the aforementioned literature. In particular, we adopt a similar approach proposed by TipText [78] and BiTipText [77], where we optimize our keyboard layout by combining a user study on users' natural ankle spatial awareness with a computer-simulated language model to reduce word ambiguity.

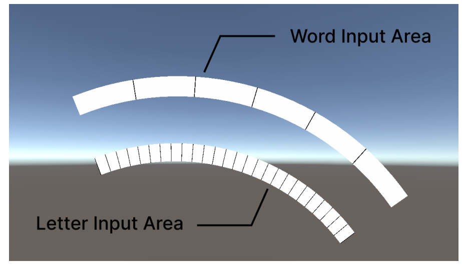


Figure 2: Illustration of our interface design concept of using two concentric radial areas for letter and word input areas separately.

3 Design Space and Research Questions

This section elaborates on the design space of the ankle-based typing technique, including interface design and interaction design, and defines our research questions.

3.1 Interface Design

Our design follows a general principle of ergonomic interaction design [58, 66, 85] where we adapt a radial menu layout controlled by ankle rotation. In the design context of text entry interfaces, following prior works [38, 70, 75, 76], we organize the interface into two concentric radial areas for letter and word input separately, as shown in Figure 2.

Importantly, our design aims to support ankle-rotation typing in both standing and sitting postures, with the cursor's position mapping to users' ankle movement. This raises a critical design consideration to investigate whether there is any significant difference in ankle rotation range across different postures, to determine if it is necessary to design distinct keyboard layouts for each posture.

In this context, another design factor, keyboard layouts, was identified with considerations of balancing the pointing accuracy and the text input efficiency under eye-free conditions. Keyboard layout here refers to the number, size, and spatial arrangement of keys, as well as their corresponding letter assignments. Specifically, layouts with larger keys may facilitate pointing speed and accuracy, while layouts with smaller keys allow alphabets to be assigned more separately across different keys, which may improve the language disambiguation and typing efficiency. Accordingly, we raise the following two research questions to guide our exploration on the interface design of AnkleType:

- **RQ1** Are there significant differences in ankle rotation range between standing and sitting postures, and if so, how do these differences appear?
- **RQ2** How to find an optimal keyboard layout considering the balance between pointing accuracy and input efficiency under the eye-free condition?

3.2 Interaction Design

With the proposed interface, our goal is to investigate a user-friendly and efficient interaction design to support the text entry task in the VR context. We divide the interaction task into two subtasks: navigation and control, where the navigation task focuses on controlling cursor movement, and the control task focuses on command execution. For the navigation input, our goal is to provide an ergonomic input solution while minimizing users' foot movement, as suggested by prior works [68]. Prior research on foot-based interaction revealed that heel-pivot rotation is efficient and user-friendly [4, 56], which has been widely adopted to foot-based HCI systems [30, 38, 70]. Building on this foundation,we adopt Norman's design principle of natural mapping [49], by explicitly mapping the cursor position to the direction in which the user's toe points, to provide an intuitive and naturally controllable interaction experience.

Building on it, our exploration focuses on the interaction design for control inputs, including commands for basic input tasks such as confirm and delete, and commands for switching between letter-input and word-input. In addition, we also investigate typing with unipedal and bipedal input, in both standing and sitting postures. This is motivated by prior research [68] suggesting that using both feet in a sitting posture would improve typing efficiency through alternating foot usage, while we hypothesize that typing with both feet in a standing posture may increase users' fatigue. To find an optimal trade-off between typing efficiency and users' preferences across postures, we formulate the following two research questions to guide our exploration on the interaction design of AnkleType:

- **RQ3** What are users' preferences for different input commands in the ankle-based text entry task, considering both unipedal and bipedal interaction?
- **RQ4** How do different input strategies (unipedal vs. bipedal) under postures (standing vs. sitting) affect users' typing performance and their subjective preference in ankle-based text entry tasks?

4 Preliminary Explorations: Understanding Users' Performance and Preference of Ankle-based Foot Interaction

Our exploration starts with two preliminary studies. One is a within-subject quantitative analysis of ankle horizontal rotation range in both standing and sitting postures across users (**RQ1**). Another one is an elicitation study to generate user-defined ankle-based foot gestures for each of the input actions (**RQ3**).

4.1 Preliminary Study 1: Investigation of Ankle Horizontal Rotation Range across Posture

4.1.1 Participants. We recruited 22 participants (13 male, 9 female), aged 18-26 years ($M = 22.00, SD = 2.83$), through social media. Their shoe sizes (EU) ranged from 41 to 44 ($M = 42.00, SD = 0.96$) for male and from 36 to 39 ($M = 37.72, SD = 1.15$) for female. All participants were right-handed users.

4.1.2 Apparatus. To accurately track users' ankle horizontal pointing angle, we developed a customized shoe with an HTC Vive Tracker 2.0 that was mounted on the toe cap (Figure 4a), paired with an HTC Vive Pro 2 HMD for tracking. We built a customized program using Unity3D (2023.2.20f1c1) with SteamVR Unity plugin (version 2.8.0) on a desktop PC with an i9-14900K CPU, 64 GB RAM, and an NVIDIA GeForce RTX 4060 GPU for the experiment. An HTC Vive controller was used to trigger the data recording. The program recorded a 3-DoF rotation data of the tracker only when the controller trigger was pressed. We provided a 42 cm-high chair with fixed armrests to simulate the most common and natural sitting postures.

4.1.3 Study Design and Procedure. This within-subject study measures and compares the range of right ankle rotation between standing and sitting postures. Throughout the study, participants were instructed to sit/stand in a natural way with their feet naturally positioned in their most comfortable posture. To capture representative measurements for the range of rotation, we measured the angles at three positions on each trial: the middle-rest position, the far-left position, and the far-right position.

Participants began the study by filling out a pre-test questionnaire with demographic information. After a brief introduction of the purpose of the study, participants were required to wear our customized tracking shoe on the right foot and hold an HTC Vive controller. For each measurement, participants rotated their ankle to the instructed position and pressed the trigger to record the angle data. Each trial began at the middle-rest position, representing the most comfortable resting position, followed by gradually rotating to either the far-left or far-right position within the comfort range. To mitigate the order effect, the order of the left and right measurements is counter-balanced in each trial.

Each block consisted of 10 trials, and participants were required to complete two blocks per posture (standing vs. sitting), with one-minute breaks between blocks. The posture order was counterbalanced across participants. As a result, each participant contributed 2 (blocks) × 10 (trials) × 2 (postures) = 40 trials, with each trial consisting of three recorded angles from three positions.

4.1.4 Results. To facilitate the data analysis, we transferred the angle data into the left range (range from far-left to middle-rest) and the right range (range from middle-rest to far-right). This yielded two dependent variables (DV), with the posture (standing vs. sitting) as the independent variable (IV). Shapiro-Wilk tests indicated except for the left range of standing, other data are normally distributed ($p > .05$).

We applied Wilcoxon Signed-Rank test to compare the left range between standing and sitting. The results indicated that for the left range, standing ($M = 33.66$, $SD = 6.52$) was significantly larger than sitting ($M = 30.00$, $SD = 5.63$), $Z = -3.2$, $p < 0.005$. For the right range, paired-samples t-test revealed the right range was greater in standing ($M = 50.77$, $SD = 8.83$) than in sitting ($M = 44.58$, $SD = 6.91$), $t(21) = -4.57$, $p < .005$.

We further look into the proportion of the left and the right angle ranges in the overall angle span. Paired-samples t-test indicates that there were no significant differences ($t(21) = .31$, $p = .76$) in left proportion for standing ($M = .40$, $SD = .05$) and sitting ($M = .40$, $SD = .06$), and no significant differences ($t(17) = -.31$, $p = .76$) in right proportion for standing ($M = .60$, $SD = .05$) and sitting ($M = .60$, $SD = .06$).

4.1.5 Findings and Insights. The results reveal that ankle horizontal rotation range in the sitting posture is significantly smaller than that in the standing posture, while no significant differences were observed when considering their proportions (**RQ1**). One possible reason is that participants tended to involve knee movement during standing, resulting in a larger ankle rotation range compared to sitting. This gives us several key insights.

F1 Key Size vs. Key Number. These findings suggest the need to explore separate keyboard layouts for each posture. Specifically, the differences in absolute range inspire us that the key size should be distinct for each posture to adapt to the range difference. Meanwhile, the lack of significant differences in left and right proportions across postures indicates that users' relative movements are similar across postures, suggesting the number of keys for the left and right sections can be consistent across postures to maintain a similar user experience.

F2 Natural Movement. During the study, we also observed that participants preferred to rotate their ankle around the heel with their sole slightly lifted while keeping their heel as support, rather than rotating the entire foot flat on the ground. This observation inspired us to optimize our tracking algorithm to align with users' natural movement patterns.

4.2 Preliminary Study 2: User-defined Foot Gesture for Ankle-Based Typing Task

4.2.1 Participants. We recruited 18 participants (12 male, 6 female) back from the previous study, aged between 18-26 years ($M = 22.67, SD = 2.83$). Their shoe sizes (EU) ranged from 41 to 44 ($M = 42.08, SD = 1.00$) for male and from 37 to 39 ($M = 37.75, SD = 1.41$) for female. All participants were right-handed users. Eight participants reported prior VR experiences, and four of them had more than 0.5 years of experiences, across applications such as gaming, documenting, and movie watching.

4.2.2 Apparatus. We adopted a similar hardware setting to the previous study. To demonstrate the design task, we implemented a prototype radial interface with separated letter and word input areas as described in section 3.1. The letter area was equally divided into 26 keys, each corresponding to one letter, and we allowed the participants to control the cursor by rotating their ankle. The cursor could be switched between the letter area and the word area by pressing the space button on the keyboard.

4.2.3 Study Design and Procedure. The purpose of this study is to generate a series of user-preferred foot gestures for an ankle-based typing task. To this end, we adopted a gesture-elicitation study scheme [10], by providing the users a set of functions or action referents, and asked them to define his/her desired foot gesture using different input strategies (unipedal vs. bipedal) accordingly. We chose four referents, with their effect as shown in the Table 1.

At the beginning of the session, the experiment facilitator introduced the study purpose and asked participants to complete a pre-study questionnaire collecting anonymous demographic information. Participants were then instructed to wear the customized shoe and the HMD, through which the interface was displayed in a virtual scene. To ensure that the participants fully understand the effects of the referent, we allowed them to familiarize themselves with the interface before each rating task. For those referents involving menu switching, the facilitator demonstrated the effect by pressing the space button to switch the cursor. To reduce the legacy bias, we adapted the *Production* technique [45], by requiring each participant to design 3 foot-based gestures for each referent for both unipedal and bipedal input, and rate them according to their preference. Referents were presented in the same order to each participant. The study was conducted without posture constraints. This setting allowed participants to generate a broader range of gesture ideas, which we later evaluated and analyzed in the context of both sitting and standing use cases. After they completed each of the referents, we performed a semi-structured post-task interview to elicit feedback about their experience, including their design consideration and qualitative feedback on the design.

Referents	Effects
Confirm	Enter a letter or word selected by the cursor
Cancel	Delete a letter or word
Switch in	Move the cursor from the letter input area to the word input area
Switch out	Move the cursor from the word input area to the letter input area

Table 1: List of referents and their effect for the foot gesture elicitation task.

4.2.4 Results and Findings. We calculated a weighted score of each gesture based on participants' preference, where we assigned their first preference with score of 3, their second preference with score of 2, and their least preferred with score of 1. We then select the top 3 gesture candidates for each referent under both unipedal and bipedal strategies, as illustrated in Figure 3. These results and the post-task interview revealed several key insights for **RQ3**.

F3 Symmetricity. Participants consistently tended to produce symmetrical gestures for paired referents, regardless of whether the input was unipedal or bipedal. For example, for paired referents such as *Confirm* and *Cancel*, users tend to use symmetric gestures *forefoot single tap* and *rearfoot single tap* for both unipedal and bipedal input. For the other paired referents, *Switch in* and *Switch out*, the gestures with the highest score are *foot flat forward* and *foot flat backward* for both input strategies, which are also symmetric gestures. This insight also aligns with the results from other gesture elicitation studies that yield gestures with symmetricity [10, 73].

F4 Natural Mapping. We also observed that participants prefer gestures with intuitive spatial mappings. Since the word-input area is positioned above the letter-input area, participants have a consistent preference to push forward their feet to enter the word area and backward to return to letters, regardless of whether the input is unipedal or bipedal. This finding also aligns with Norman's design principle of natural mapping [49].

F5 Foot Role Balancing. With bipedal input, participants tended to use their dominant foot to perform precise input, such as cursor control, while assigning less precise input, such as control commands, to the non-dominant foot, thereby balancing interaction workload. While participants generally agree that bipedal input would improve input efficiency, the interview revealed that 11 participants prefer

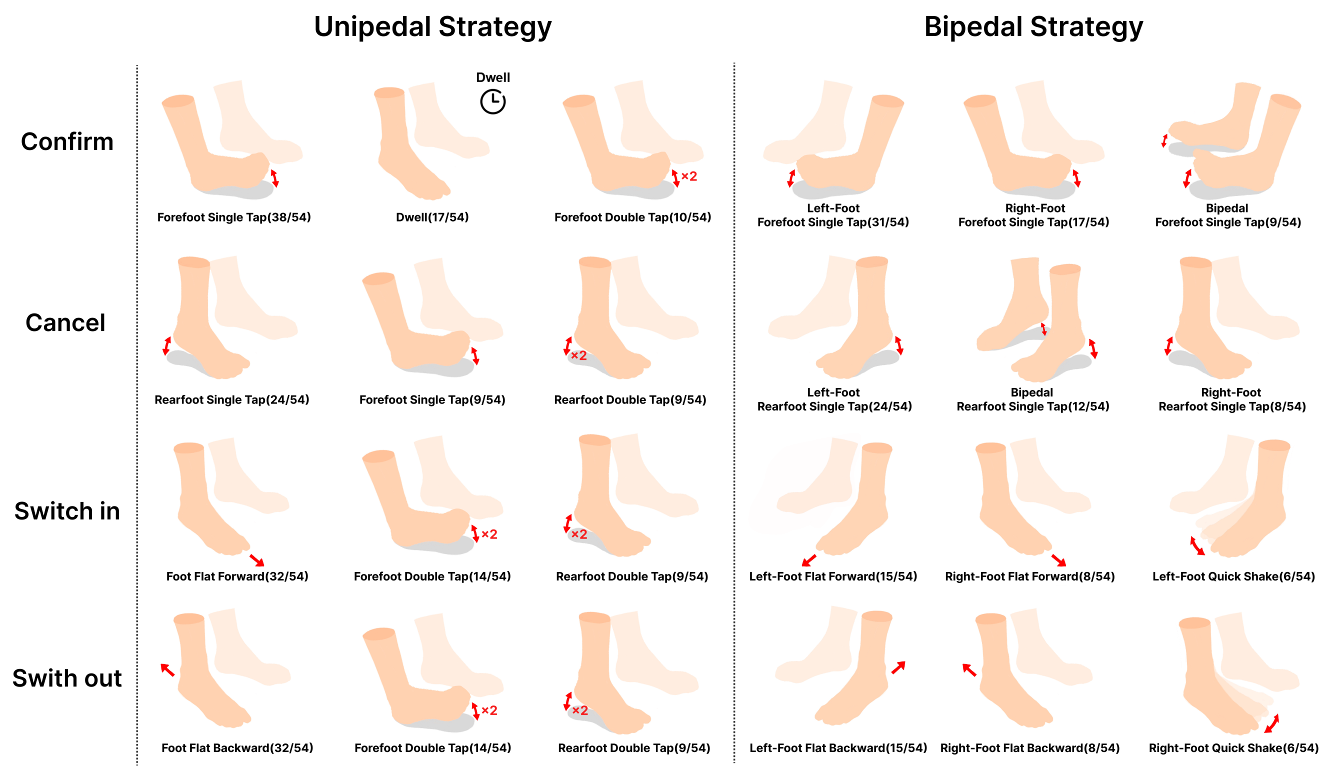


Figure 3: Foot gesture elicitation study results. This figure shows the top 3 gesture illustrations for each referent under both unipedal and bipedal strategies with their weighted preference score. Each gesture had a maximum score of 54.

unipedal input in standing conditions, as the other foot was needed for physical support, whereas 4 participants mentioned that they prefer bipedal input for both standing and sitting postures to maintain design consistency and reduce learning fatigue.

5 Study 1: Keyboard Layout Optimization

To design an optimal keyboard layout for AnkleType that balances the pointing efficiency and input efficiency under eyes-free conditions, we adopted a data collection study to understand users' natural expectation of ankle-rotation-based input through their spatial locatability on the keys of an ankle-based keyboard without visual feedback.

5.1 Participants

We recruited 12 participants (8 male, 4 female), aged between 19-25 years ($M = 20.92, SD = 2.31$), through social media. Their shoe sizes (EU) ranged from 41 to 44 ($M = 42.00, SD = 0.96$) for male and from 36 to 39 ($M = 37.72, SD = 1.15$) for female. All participants were right-handed users. Three of them reported prior VR experiences.

5.2 Apparatus

We implemented a customized shoe with an HTC Vive Tracker 2.0 that was mounted on the toe cap to enable rotation tracking. To detect the two foot gestures for inputting letter and deleting letter with the highest user preference, *forefoot single tap* and *rearfoot single tap*, we used two infrared photo-reflective sensors (SG-105) embodied at the back and front of the shoe's sole (Figure 4ⓐ), similar to Chan et al. [9]. An ESP-32 chip was used to process and transfer the sensor data to the desktop interface through the serial port. We built the experiment interface using Unity3D (2023.2.20f1c1) with SteamVR Unity plugin (version 2.8.0) on a desktop PC with an i9-14900K CPU, 64 GB RAM, and an Nvidia GeForce RTX 4060 GPU for the experiment. The implementation of the rotation tracking algorithm follows the inspiration of **F2**.

5.3 Initial Keyboard Layout Design

As suggested by prior research on data-driven keyboard layout design [78], we began our exploration with a full alphabet keyboard. For the initial layout, we distributed 27 keys (26 letters + space bar) evenly across the keyboard area for both standing and sitting postures (Figure 4ⓑ). **F1** revealed that the segmentation between the left and right sections of the keyboard followed the same proportion across postures. This would result in an equal number of keys on each side across postures. We assign the space bar to the middle-rest position as we consider that word input naturally ends

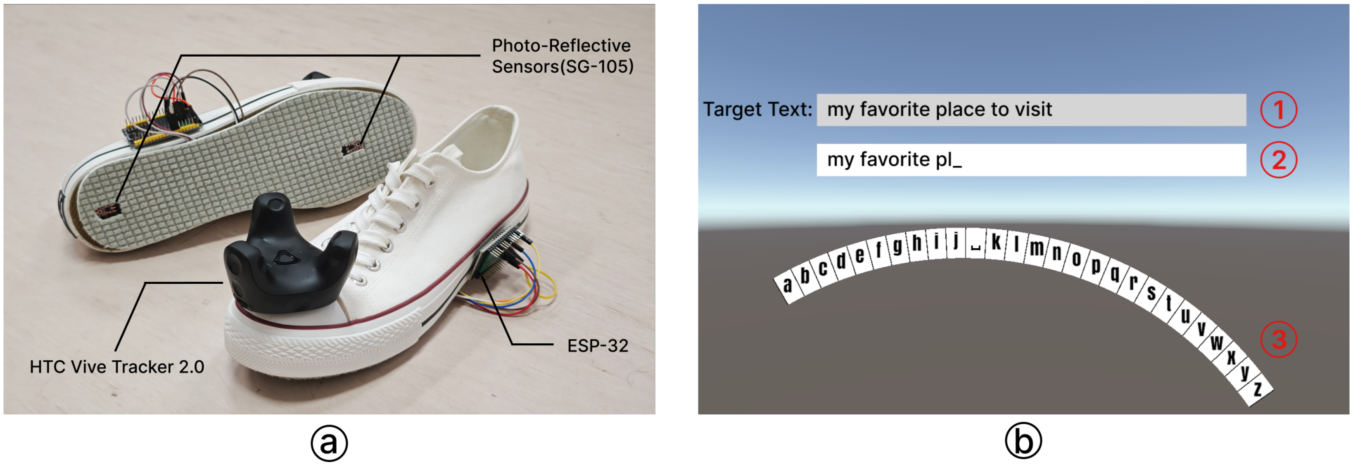


Figure 4: ⓐThe shoe prototype for AnkleType, including two photo-reflective sensors at the fore and rear of the shoe, an ESP-32 to process the sensor signal, and an HTC Vive Tracker 2.0 to enable ankle rotation tracking. This implementation was based on Chan et al.'s [9] solution.ⓑThe interface for the keyboard layout optimization study. ①The target text to be transcribed, ②The current input text, ③The initial keyboard layout, where we evenly distributed 27-key (26 letters + space bar) across the keyboard area.

with a space for segmentation, and the middle-rest is the most intuitive position for this input. This setting also ensures that each word entry begins from the middle-rest position, offering a consistent spatial reference for subsequent key inputs. Notably, considering the movement range difference across participants and postures, the absolute size of the keyboard was proportionally scaled to fit before each trial.

5.4 Task and Procedure

The purpose of this study is to examine users' spatial locatability on the keys of the ankle-based keyboard without visual feedback. To this end, we designed an eyes-free (i.e., no cursor was visible) within-subject text entry task using a Wizard of Oz keyboard [78, 86]. Participants were required to transcribe 2 blocks of 10 phrases in both standing and sitting postures respectively, for a total of 40 phrases. Of these, 32 phrases were picked randomly from MacKenzie's phrase set [42], while the rest 8 phrases, as suggested by previous works [20, 21, 77], were randomly selected pangrams[1] to ensure that every letter had a minimum presence of 15 times each. All selected phrases were case-insensitive and contained no numbers or symbols.The order of standing and sitting conditions was counterbalanced using a Latin Square to mitigate order effects. The requirements for standing and sitting postures are the same as in previous experiments(Section 4.1.3).

Prior to the study, participants completed a demographic questionnaire. After a brief introduction, they were required to wear our customized shoes and an HTC Vive HMD for calibration. During calibration for each trial, we recorded each participant's far-left, middle-rest, and far-right positions within their comfort range, then proportionally scaled the size of the keyboard to match their individual range. Participants then completed five phrases for training, during which they could see the keyboard layout and a cursor highlighting the currently selected key. Inspired by our exploration of **RQ3**, participants were instructed to navigate the cursor by rotating their ankle and confirmed the input with a *forefoot single tap* gesture. During training, incorrect letters were displayed in red color to inform the user, but we did not provide a delete function.

In the formal study, participants were provided a static reference keyboard layout without a visual cursor to indicate the currently selected key. They were asked to select the letter by rotating their ankle and perform a *forefoot single tap* gesture on an imaginary key location for selection based on their natural spatial awareness. The system always displayed the correct letters regardless of their actual tap location. Different from the training process, we did not highlight incorrect letters in the formal study. Participants were instructed to keep typing regardless of whether they felt the inputted key was correct or not. They were required to complete the whole letter sequence for every word, without word auto-completion. Therefore, we only displayed the letter area on the interface for this study. After each phrase, the system switched to the next phrase until all phrases were completed. Participants were allowed a 3-minute break between blocks. The study lasted approximately 80 minutes, and participants were compensated with 15 USD for their time.

5.5 Results

We collected 18,582 foot tapping positions from the study, which were distributed along a 1D space and varied across participants. To facilitate the comparison, we align the data points from different users by applying min-max normalization across users to form a general distribution for standing and sitting posture, respectively, as shown in Figure 5. For clarity, we present each key's distribution with a 95% confidence interval, and the adjacent keys are shown on different sides with distinct colors. The result shows that the tapping

[1]https://mseffie.com/assignments/calligraphy/Plethora%20of%20Pangrams.pdf

distributions vary across standing and sitting postures. The tapping locations between letters are noisy with significant overlaps among neighboring distributions. This finding also highlights the need to answer **RQ2**, as typing with 27 keys is not feasible, particularly in an eye-free condition.

Despite the noise, the distributions largely followed the alphabetical order, with some keys being indistinguishable from each other and some of these characteristics consistent between standing and sitting conditions. For instance, the distributions of "p" and "r" almost completely overlapped in each condition. This highlights the necessity of incorporating a language model to decompose the input into keys with letter ambiguity. Moreover, we particularly observed that apart from the space bar, the key with the smallest standard deviation in the standing posture was the letter "q" ($\sigma = .070$), while in the sitting posture it was the letter "e" ($\sigma = .075$). In contrast, the key with the largest standard deviation in the standing posture was the letter "r" ($\sigma = .119$), and in the sitting posture it was the letter "w" ($\sigma = .169$). These findings suggest the reachability varies at different positions, highlighting the need to consider the spatial reachability separately for each posture condition. We later integrated this spatial information with a language model to determine the final layout.

5.6 Determine AnkleType Keyboard Layout

There exists a trade-off between keyboard layout and typing efficiency. Layouts with larger keys facilitate pointing speed and accuracy, but reduce the total number of keys, thereby lowering alphabet separability and increasing language ambiguity, which eventually slows input speed. In contrast, layouts with smaller keys allow letters to be distributed more separately across the different keys. This will improve language disambiguation and typing efficiency but reduce key size, thereby reducing pointing speed and accuracy. Based on these insights, we explore an optimal keyboard layout for our task by jointly considering the spatial model in our ankle-based typing context and the language model.

Before deriving the two models, we first determined the acceptance range of the key number in our context. As mentioned above, the most precise key has a minimum standard deviation of around 0.07 for both standing and sitting positions. We defined the minimum spatial resolution as the condition where the overlap between two adjacent keys does not exceed half of their size. Based on this criterion, we set the maximum number of keys to 14. Moreover, as suggested by previous works [78], those keyboards with letter key numbers lower than 5 would suffer from high input ambiguity. Therefore, we set our minimum acceptance letter key number to 5. Note that in our case, the space bar takes up one key position. As a result, we set our acceptance range of key numbers from 6 to 14 in our context.

5.6.1 Word Disambiguation Score. We sequentially mapped the 26 letters onto keyboards with the number of keys ranging from 5 to 13 (reserve one key for the space bar), resulting in 16774590 possible letter layout combinations. However, it is impossible to run a user study to test through these options. Instead, we adapted a simulation-based approach, as suggested by previous works [53, 78], using a computer-simulated typing test to compare the theoretical performance across all different layouts. Specifically, we select the top 10,000 words from the American National Corpus [3] and simulate the key sequence required to input each of them on every candidate letter layout. For each sequence, the system generated a list of candidate words (due to input ambiguation) that matched the input and were ordered by word frequency. We then recorded whether the intended word appeared within the top three entries of this list. The proportion of successful occurrences was computed as the word disambiguation score L^k for each letter layout, where k is the number of keys. Finally, we select the top 100 L^k in each k as letter layout candidates, 900 in total, denoted as J.

5.6.2 Integrate with Spatial Information. The exploration of spatial information starts by grouping the tapping positions whose spatial distributions are concentrated. To this end, we applied an unsupervised Gaussian Mixture Model (GMM) clustering approach, and iteratively set the number of clusters from 6 to 14 for both standing and sitting postures. This process yielded 18 cluster layout candidates, with 9 for each posture, denoted as I.

For each cluster layout candidate in each posture, we first determine the cluster with the largest number of points as the space key, and we exclude the space key for the following calculation. Then, we sequentially mapped the letters from all letter layout candidates with the same key number setting on it. This process resulted in 1,800 combinations in total, with 900 for each posture. For each cluster-layout and letter-layout pair $(i, j | i \in I, j \in J)$, we computed a spatial matching score $S(i, j)$ that integrates tapping accuracy with corpus frequency. Specifically, for each letter l in $(i, j | i \in I, j \in J)$, we calculated the weighted letter score $s_l(i, j)$ as:

$$s_l(i, j) = \frac{n_l(i, j)}{N_l} \times f_l \tag{1}$$

where n_l is the number of tapping samples for letter l that fall within the assigned key cluster, N_l is the total number of tapping samples for l, and f_l is the frequency of l in the corpus[2]. We then summarize the score across all letters to compute the cluster-layout and letter-layout pair$(i, j | i \in I, j \in J)$ spatial matching score $S(i, j)$:

$$S(i, j) = \sum_{l \in A} s_l(i, j) \tag{2}$$

where A denotes the set of all 26 letters.

With the inspiration from **F1** and the consideration of following design consistency, we decided to keep the number of keys and letter layout consistent for standing and sitting posture. Therefore, we jointly consider $S(i, j)$ across postures by accumulating those that have the same number of key k, denoted as $S^k_{joint}(i, j)$.

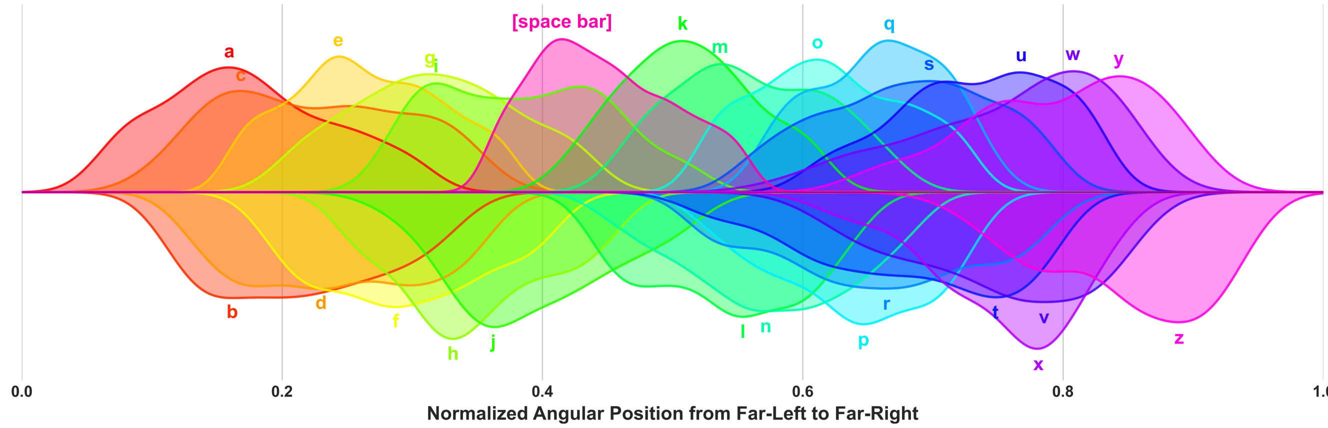
$$S^k_{joint}(i, j) = S^k_{stand}(i, j) + S^k_{sit}(i, j) \tag{3}$$

For each $(i, j | i \in I, j \in J)$ under each k, we calculate their sum of its word disambiguation score $L^k(j)$ and $S^k_{joint}(i, j)$ as the final score:

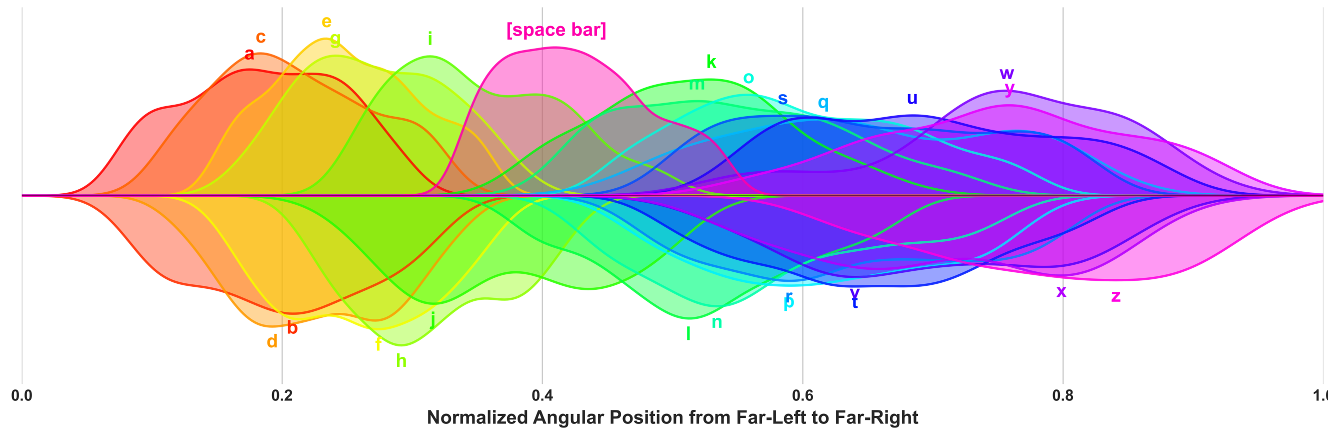
$$F^k(i, j) = L^k(j) + S^k_{joint}(i, j) \tag{4}$$

Then for each k, we average the top 10 $F^k(i, j)$ and plot in Figure 6. As shown in the plot, the word disambiguation score increases

[2] https://pi.math.cornell.edu/ mec/2003-2004/cryptography/subs/frequencies.html



(a) Distribution of foot tapping positions for standing condition.



(b) Distribution of foot tapping positions for sitting condition.

Figure 5: Distribution of foot tapping positions with 95% confidence interval in a 26-key Alphabetical-order keyboard. (a) Distribution of foot tapping positions for standing condition. (b) Distribution of foot tapping positions for sitting condition.

with the number of keys, whereas the spatial score decreases. Notably, we observed a dramatic change in the spatial score at the key number of 9, where the growing trend of the word disambiguation score starts to slow down. Based on this trade-off, we selected 9 as the final key number. We then looked into the top 10 letter layouts for the key number of 9, and we were pleasantly surprised to find that one of them had a top 1 score in L^9 (rank 6 in F^9). As a result, we chose this cluster-layout and letter-layout pair as the final layout (**RQ2**) (see Supplementary Materials for other pair candidates). Please refer to Figure 7 for the detailed keyboard layout.

6 Study 2: Exploring Ankle-Based Text Entry Technique when Standing and Sitting

With the optimal keyboard layout, we conducted a within-subject comparison study to evaluate the performance and user experience of different input strategies (unipedal vs. bipedal) under standing and sitting postures.

6.1 User-elicited Ankle-Based Text Entry Technique

Based on the user-defined gesture set (Figure 3), we first designed two ankle-based text entry techniques for both unipedal and bipedal using the most popular gestures. The same techniques were used across postures, yielding four ankle-based text entry techniques for this comparison study.

Unipedal Stand/Sit (UPStand/UPSit). Users navigate the cursor by rotating their right ankle. A right-foot *forefoot single tap* gesture for entering a letter or word, while a right-foot *rearfoot single tap* gesture for deleting. Users perform a *foot flat forward* gesture with their right foot to enter the word input interface, and a right-foot *foot flat backward* gesture to return to the letter input interface.

Bipedal Stand/Sit (BPStand/BPSit). Users navigate the cursor by rotating their right ankle. A left-foot *forefoot single tap* gesture for entering a letter or word, while a left-foot *rearfoot single tap* gesture for deleting. Users perform a *foot flat forward* gesture with

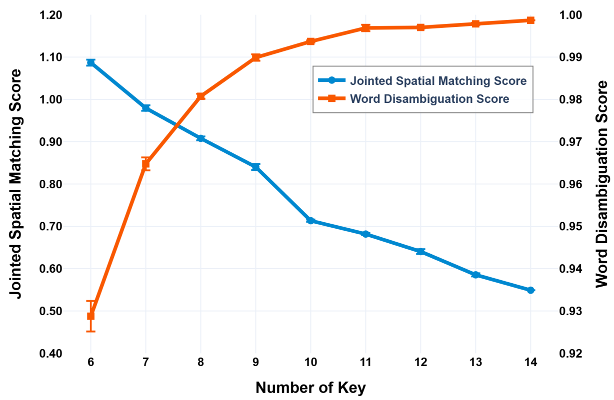


Figure 6: Word disambiguation score L^k and jointed spatial matching score S^k_{joint} of each key number. Note that we did not include the space key when calculating these metrics. Therefore, the actual number of keys for calculating these metrics should be minus one.

their left foot to enter the word input interface, and a left-foot *foot flat backward* gesture to return to the letter input interface.

6.2 Participants

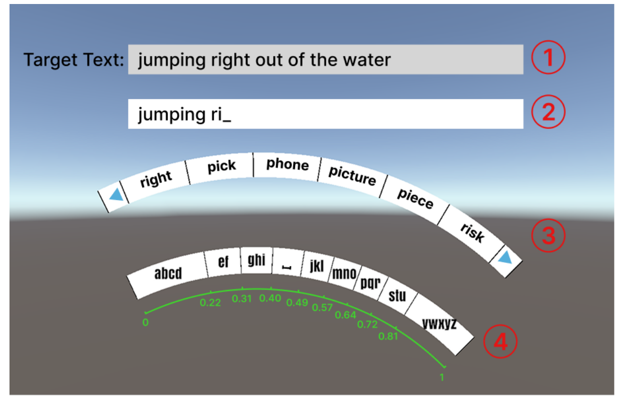
We recruited 16 participants (10 male, 6 female) through word of mouth, aged from 19-25 years (M = 22.13, SD = 2.23). Their shoe sizes (EU) ranged from 40 to 43 (M = 41.70, SD = 1.16) for male and from 37 to 39 (M = 37.83, SD = 2.23) for female. Their mean self-assessed typing proficiency is 6.5/10. All participants were right-handed users. One of them explicitly mentioned being familiar with VR text entry using controller ray casting. The other participants had no experience with VR text entry.

6.3 Apparatus

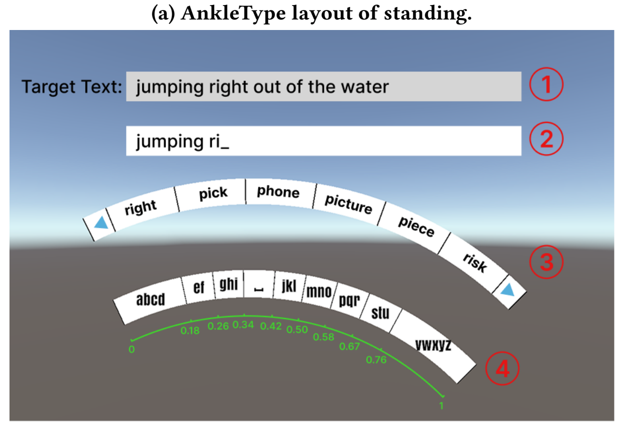
We adopted similar hardware and software settings as our previous study (Section 5.2). To simulate a real-world typing experience, we implemented an interface for this study with a language statistical decoding algorithm, based on a Bayesian maximum likelihood estimation [22], similar to [68]. The lexicon we used was the top 10k words from the American National Corpus [3]. With the decoding algorithm, the word-input area displays the predicted words ordered from left to right by their probabilities. The GUIs for this study are shown in Figure 7.

6.4 Study Design and Procedure

We conducted a within-subjects study with posture (standing vs. sitting) and strategies (unipedal vs. bipedal) as two independent variables, yielding a 2×2 design with 4 conditions in total. The requirements for standing and sitting postures are the same as in previous experiments(Section 4.1.3).The order of these conditions was counterbalanced using a Latin-Square design to mitigate the order effects. In each condition, participants were asked to transcribe 10 phrases randomly selected from MacKenzie’s phrase set [42]. All



(a) AnkleType layout of standing.



(b) AnkleType layout of sitting.

Figure 7: Illustration of the user study interface with the optimized keyboard layout for both standing (a) and sitting (b) conditions. ① The target text to be transcribed, ② The current input text, ③ The word input area. Two side buttons on the left and right are for page navigation, ④The optimized letter input area. The green scale below indicates the normalized angular position from far-left to far-right.

selected phrases were case-insensitive and contained no numbers or symbols. Since the goal of this study was to explore input strategies across postures rather than measuring the eye-free performance, participants were provided with visual feedback of the keyboard layout during the task.

Prior to the study, participants completed a demographic questionnaire. After a brief introduction, they were required to wear our customized shoes and an HTC Vive HMD for calibration. During the calibration, we measured each participant’s right-ankle rotation range in both sitting and standing conditions. This calibration data was used to proportionally scale the keyboard layout to fit participants’ individual movement range for the corresponding condition. Before each condition, participants completed a short training session consisting of 5 practice phrases to familiarize themselves with the input method.

During the main study, participants were instructed to enter phrases as quickly and accurately as possible. After each condition, they were asked to complete the post-study questionnaires, including a NASA-TLS questionnaire and a System Usability Scale (SUS) questionnaire. A two-minute break was provided between conditions. After finishing four conditions, we conducted a semi-structured interview to collect their subjective feedback. The study lasted approximately 75 minutes, and participants were compensated with 15 USD for their time.

6.5 Results

We measure the entry speed in words per minute (WPM), the error rate in total error rate (TER) and not corrected error rate (NCER), the workload in NASA-TLX score, and usability in SUS score. A Shapiro-Wilk test was performed before the analysis, indicating the WPM, TER, and SUS score were normally distributed (p > .05), while the NCER was not normally distributed (p < .05). The results are shown in Figure 8.

6.5.1 Entry Speed. We first applied an RM-ANOVA test on the WPM to see the effect of posture and strategies on the entry speed. The results show that posture does not significantly affect the entry speed ($F(1, 15) = 4.44, p = .052, \eta_p^2 = .23$), while strategies have a significant effect on WPM ($F(1, 15) = 8.86, p < .05, \eta_p^2 = .37$). A post-hoc pairwise comparison indicated that the entry speed of unipedal ($M = 9.28, SD = .41$) is significantly higher ($p < .05$) than that of bipedal ($M = 8.40, SD = .33$).

We further conducted a comparison across strategies for each posture. Paired-samples t-test indicate the WPM of unipedal ($M = 9.13, SD = 1.50$) is significantly higher that that of bipedal ($M = 7.80, SD = 1.88$) for standing posture, $t(15) = -3.20, p < .05$. However, there is no significant ($t(15) = -1.32, p = .205$) between unipedal ($M = 9.44, SD = 2.08$) and bipedal ($M = 8.99, SD = 1.26$) for sitting posture on WPM.

6.5.2 Error Rate. RM-ANOVA tests revealed the fact that both posture ($F(1, 15) = 12.73, p < .005, \eta_p^2 = .46$) and strategies ($F(1, 15) = 8.20, p < .05, \eta_p^2 = .36$) have significant effect on TER. Post-hoc pairwise comparison showed that the TER of sitting ($M = 3.99, SD = .31$) posture is significantly lower ($p < .005$) than that of a standing posture ($M = 4.88, SD = .33$). In terms of strategies, the TER of bipedal ($M = 3.95, SD = .26$) is significantly lower ($p < .05$) than that of unipedal ($M = 4.93, SD = .41$).

Paired-samples t-test indicates no significant effect ($t(15) = -.29, p = .78$) of TER between unipedal ($M = 4.97, SD = 1.71$) and bipedal ($M = 4.79, SD = 1.90$) while standing, but for the sitting posture, the TER of unipedal ($M = 4.89, SD = 1.78$) is significantly higher than that of bipedal ($M = 3.10, SD = 1.12$), $t(15) = 4.48, p < .001$. We then applied Wilcoxon Signed-Rank test on NCER. The results did not show any significant effect between unipedal and bipedal for either standing posture (unipedal: $M = .98, SD = .75$, bipedal: $M = 1.39, SD = 1.50, Z = -1.16, p = .245$) or sitting posture (unipedal: $M = 1.06, SD = 1.07$, bipedal: $M = .86, SD = .75$, $Z = -.79, p = .433$).

6.5.3 Usability and Workload. Paired-sample t-test on overall SUS score shows that for standing posture, the score of unipedal ($M = 72.34, SD = 9.55$) is significantly higher than that of bipedal ($M = 64.69, SD = 9.61$), $t(15) = -3.36, p < .005$. However, there is no significant difference between unipedal ($M = 75.47, SD = 12.85$) and bipedal ($M = 72.97, SD = 13.45$) for sitting posture, $t(15) = -1.06, p = .304$.

We then applied Wilcoxon Signed-Rank test to the NASA-TLX questionnaire results. For the standing posture, participants reported significantly lower mental demand and physical demand when typing with unipedal compared to the bipedal ($p < .005$). They also reported that bipedal input required more effort ($p < .05$), was more frustrating ($p < .05$), and had less perceived performance ($p < .05$) than unipedal input. In contrast, there were no significant differences found between unipedal and bipedal in the sitting posture.

6.6 Discussion

The results revealed meaningful insights to answer **RQ4**. The standing posture requires a higher physical demand which leads to a significantly higher error rate compared with the sitting posture. However, participants consistently preferred unipedal as it requires less mental and physical demand. Participants commented that when typing using bipedal while standing, they needed to pay much attention to keeping balance, which made them feel tired and tedious. This finding also aligns with **F5**. Moreover, typing with unipedal is faster than with bipedal. Therefore, we consider unipedal is an optimal input strategy for standing posture (**UPStand**), see Figure 1 ⓐ-ⓓ.

For sitting posture, there was no significant difference in entry speed between the two strategies, and participants reported no clear preference. However, bipedal typing was significantly more accurate than unipedal one. As bipedal input separates the navigation and selection operation, it also reduces the users' physical demand (See Figure 8ⓔ). Based on these considerations, we selected bipedal as the final input strategy for the sitting posture (**BPSit**), see Figure 1 ⓔ-ⓗ.

7 Study 3: Longitudinal Evaluation

Finally, we conducted a 7-day longitudinal user study to examine the progressive learning effect of AnkleType in both visual and eyes-free conditions. In the literature, eyes-free text entry refers to entering text without having to pay attention to the keyboard [7, 41, 87].

7.1 Participants

We recruited 12 participants (8 male, 4 female) through the university's internal social media platform, aged between 20-26 years ($M = 22.25, SD = 2.30$). The shoe sizes (EU) of male participants ranged from 40 to 43 ($M = 41.50, SD = 1.20$), and the shoe sizes (EU) of female participants ranged from 37 to 39 ($M = 37.75, SD = .96$). Five participants had prior VR experience, but none reported experiences with VR typing. To ensure a fair comparison, none of the participants had taken part in any of our previous studies. The 12 participants were randomly divided into two groups, with 4 male and 2 female in each group, for two experimental conditions.

7.2 Apparatus

We adopted similar hardware and software settings to our previous study (Section 5.2). In addition, to support the learnability study

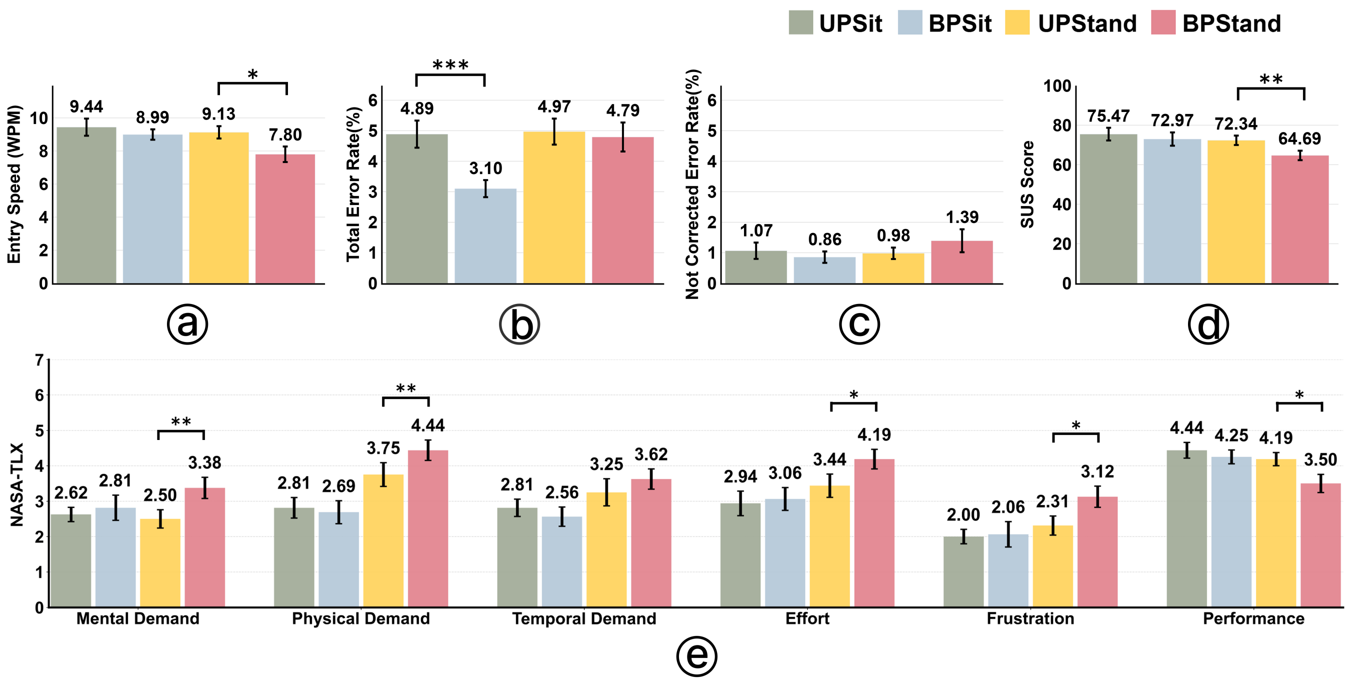


Figure 8: The means of ⓐ Entry speed, ⓑ TER, ⓒ NCER, ⓓ SUS Score and ⓔ NASA-TLX scores. *, **, and * represent a .001, .01, and .05 significance level, respectively.**

under eye-free conditions, we implemented an interface that allows the experiment to be conducted in both *Visual* and *Blind* modes.

7.3 Study Design and Procedure

We conducted a between-subject study under two typing conditions: 1) *Visual.* Participants could see the keyboard layout along with cursor position indicators; 2) *Blind.* Participants could see neither the keyboard layout nor the cursor position indicators [7, 41, 87]. For the study under the *Blind* condition, as suggested by previous works [74, 79], we provided a cheat sheet mechanism by allowing the users to have a glance at the keyboard layout if they required to. The keyboard layout cue would appear for 10 seconds each time, and we recorded the number of requests each participant made each day.

Each day, participants were required to transcribe one block of 10 phrases using both **UPStand** and **BPSit** , with the order counterbalanced across participants. The requirements for standing and sitting postures are the same as in previous experiments (Section 4.1.3), and the phrases we used every day were picked randomly from MacKenzie's phrase set [42]. We ensured that all selected phrases were case-insensitive and contained no numbers or symbols.

Before the experiment of each day, participants were allowed to practice in either visual or blind condition for as long as they wanted. Throughout the study, participants were instructed to type as quickly and accurately as possible, with a mandatory 5-minute break between blocks to reduce fatigue. Each daily session lasts approximately 45 minutes, and participants were compensated with 40 USD for their time after finishing all sessions.

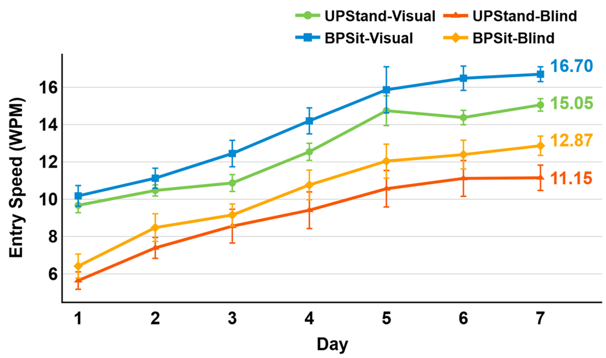
7.4 Results

During the study, we measured the entry speed in words per minute (WPM), total error rate (TER), and not corrected error rate (NCER). A Shapiro-Wilk test was performed before the analysis, indicating the WPM and TER were normally distributed (p > .05), while the NCER was not normally distributed (p < .05).

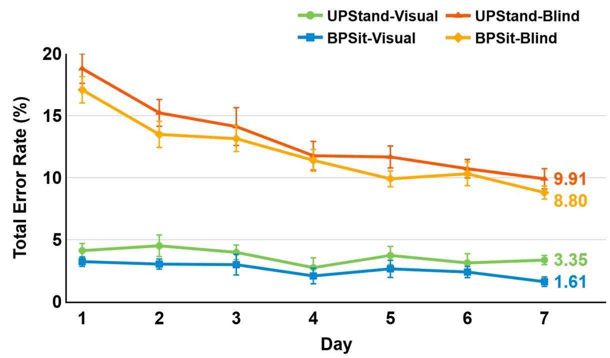
7.4.1 Entry Speed. As shown in Figure 9a, all conditions show a clear increasing trend in WPM over time. RM-ANOVA test revealed that day significant effect on WPM for all cases: **UPStand**-*Visual* ($F(6, 30) = 40.30, p < .005, \eta_p^2 = .89$), **BPSit**-*Visual* ($F(6, 30) = 23.01, p < .005, \eta_p^2 = .82$), **UPStand**-*Blind* ($F(6, 30) = 18.04, p < .005, \eta_p^2 = .78$), and **BPSit**-*Blind* ($F(6, 30) = 41.32, p < .005, \eta_p^2 = .89$).

For the *Visual* condition, the WPM increased from 9.67 to 15.05 (+55.63%), and from 10.18 to 16.70 for **BPSit** (+56.07%), with **BPSit** slightly outperforming **UPStand** throughout the period. We further run post-hoc test on each pair of consecutive days. For **UPStand**, significant improvements were observed between Day 3 and Day 4 ($p < .001$), Day 4 and Day 5 ($p < .05$). For **BPSit**, post-hoc test revoloves significant improvements between Day 1 and Day 2 ($p < .05$), Day 2 and Day 3 ($p < .05$), and Day 3 and Day 4 ($p < .005$).

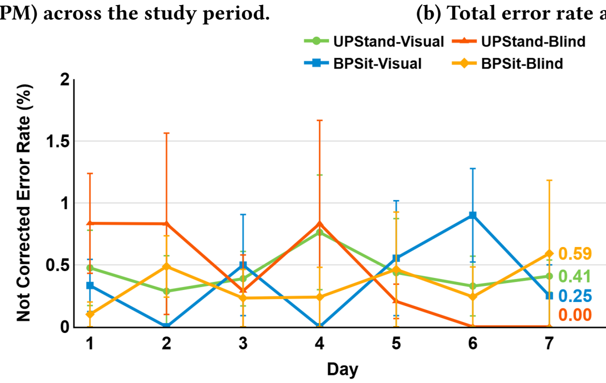
For the *Blind* conditions, the WPM increased from 5.64 to 11.15 (+97.70%) for **UPStand**, and 6.42 to 12.87 (+100.48%) for **BPSit**. The post-hoc test shows a significant improvement between the first two days for both **UPStand** ($p < .001$) and **BPSit** ($p < .005$), and between Day 4 and Day 5 for **BPSit** ($p < .05$). These consistently



(a) Text entry speed (in WPM) across the study period.



(b) Total error rate across the study period.



(c) Not corrected error rate across the study period.

Figure 9: Results of the longitudinal user study. (a) Mean of text entry speed (in WPM) across 7 days. (b) Mean of total error rate (TER) across 7 days. (c) Mean of not corrected error rate (NCER) across 7 days.

increasing trends of WPM across conditions highlight that participants were able to adapt to AnkleType quickly. The results also demonstrate the potential of AnkleType to support efficient text entry, even under challenging eye-free conditions.

7.4.2 Error Rate. As shown in Figure 9b, the average TER of *Blind* cases was obviously higher than that of a visual case, but the TER showed a decreasing trend for both cases. For the *Visual* condition, RM-ANOVA did not indicate any significant effect of day. However, TER consistently decreased across the period: from 4.12 to 3.35 for **UPStand** (-18.68%), and from 3.23 to 1.61 for **BPSit** (-50.15%). In particular, Post-hoc test shows a significant effect ($p < .005$) for **BPSit** between Day 1 and Day 7.

In contrast, the *Blind* conditions show stronger learning effects. For **UPStand**, RM-ANOVA showed a significant effect of time ($F(6, 30) = 27.02, p < .005, \eta_p^2 = .84$). The TER decreases from 18.80 at Day 1 to 9.91 at Day 7 (-47.29%). Post-hoc analysis shows significant differences between Day 1 and Day 2 ($p < .001$) as well as Day 3 and Day 4 ($p < .05$). For **BPSit**, RM-ANOVA showed a significant effect of time ($F(6, 30) = 11.73, p < .005, \eta_p^2 = .70$), with TER decreasing from 17.12 at Day 1 to 8.80 at Day 7 (-48.60%). Post-hoc test also reveals significant effect for **BPSit** between Day 1 and Day 2 ($p < .001$) as well as Day 4 and Day 5 ($p < .05$).

For NCER, Friedman test did not show significant effects across days ($p > .05$). As shown in Figure 9c, NCER values fluctuated slightly over days but without clear trends. The overall averages were 0.44 for **UPStand**-*Visual*, 0.36 for **BPSit**-*Visual*, 0.43 for **UPStand**-*Blind*, and 0.34 for **BPSit**-*Blind*.

7.4.3 Number of Layout Cue Requirement. The average number of keyboard layout cue requirements in the *Blind* condition ranged from 4 to 12 ($M = 7.83, SD = 2.79$) on the first day, with means of 4.67 and 3.83 for **UPStand** and **BPSit**, respectively. The average number of requirements on the second day dropped sharply to the range between 0 to 4 ($M = 1.83, SD = 1.72$), with an average of 0.83 times for **UPStand** and 1 time for **BPSit**. From the third day onwards, none of the participants required the visual cues.

7.5 Discussion

The results demonstrate users' potential toward a quicker and more accurate ankle spatial orientation intuition through short-term self-training, both under visual and blind conditions. For the text entry speed, both **UPStand** and **BPSit** under both visual and blind conditions follow similar growth trajectories: the performance improved rapidly in the first 5 days and then gradually stabilized. This indicates that users can achieve more precise and faster ankle movement control through short-term practice. It also reveals AnkleType's high learning potential under both fully visual and fully blind conditions. Besides, we observed a significant increase in learning effectiveness between day 3 and day 5 under the blind condition.

The number of keyboard layout cue requirement also highlight a reduction in cognitive effort as users became familiar with AnkleType. On the first day, participants frequently consulted the keyboard layout, and this number dropped sharply on the second day. From the third day onwards, none of the participants required a keyboard cue at all. This provides an intuition that users are able to quickly memorize the AnkleType keyboard layout through practice.

8 Application and Discussion

8.1 Application Scenarios

AnkleType was motivated and designed to support situations where users' hands are occupied and less available for text entry. In immersive environments, users' hands may be either engaged with virtual context, such as holding virtual objects or performing direct manipulation gestures with virtual objects, or holding physical objects. In these situations, users' hands become less available to perform conventional text entry. Note that AnkleType does not target complex dual-task scenarios involving multiple cognitively demanding tasks simultaneously. Instead, we focus on developing an alternative hand- and eye-free text entry approach to enhance the user experiences in virtual environments. The concept of AnkleType can also be extended to MR/AR scenarios, where users are likely to pay more attention to or interact with a physical artefact. We envision that AnkleType could be applied to various scenarios. Below, we propose three potential application scenarios across posture and context, as illustrated in Figure 10.

While gaze-based text entry is intuitive and effective, it occupies visual attention [55, 65] and is prone to causing motion sickness [44, 57, 69]. As AnkleType achieves a competitive performance with other gaze-based techniques (see Table 2) without the potential over-use of visual/gaze attention, which is important for many daily tasks. We believe that AnkleType serves as an alternative input method to gaze in such scenarios.

8.2 Comparison with Prior Works

We compared the performance of AnkleType after longitudinal training with prior works on hand-free VR/MR text entry. To the best of our knowledge, no existing research has directly focused on hand- and eye-free text entry in VR/MR. On the other hand, iText [39] presented a gaze-based typing method on an invisible keyboard in VR, which shared a similar design concept with our work. As shown in Table 2, **BPSit** in *Visual* condition outperforms other methods in terms of text entry speed and typing accuracy, reaching the highest entry speed at 16.70 WPM with the lowest TER at 2.48% (Section 7.4). More importantly, under the more challenging eye-free condition, both **UPStand** and **BPSit** achieve competitive performance compared with prior hand-free approaches. In addition, unlike prior works that primarily focus on a single posture, AnkleType supports both sitting and standing conditions, with consistent performance across these conditions, enabling a flexible use of AnkleType across diverse VR scenarios. This comparison shows that AnkleType not only achieves state-of-the-art performance in the hand-free VR text entry task, but also demonstrates a competitive performance in eye-free scenarios.

8.3 Foot-based Typing Ergonomics

AnkleType follows a design foundation of heel-pivot rotation with toes as the pointing direction. From an embodied cognition perspective, this design guideline aligns with the Source–Path–Goal image schema [33, 43]. Humans naturally treat the front direction of the body or limb as the intended direction for moving or targeting. A straight index finger pointing gesture is an intuition of this cognition concept [32]. Therefore, heel-pivot rotation and point with toes also align with the body image schema to move from the heel (Source) follow the sole of the foot (Path) to the toes (Goal). This intuition has also been widely adopted in foot-based interaction research [47, 56, 64]. The user-defined foot gesture yielded from our gesture elicitation study also aligned with this concept: foot flat forward/backward movement naturally maps the toe orientation as the target direction.

AnkleType is highly motivated by recent foot-based text entry research. Compared with prior foot-based text entry systems, our design further minimizes low-limb joint involvement to reduce physical effort. Wan et al. [68] engaged nearly all lower-limb joints (i.e., hip, knee, and ankle) in their typing mechanism and explicitly suggested reducing leg movement in future designs. Building on this recommendation, our design not only mitigates the movement of the hip and knee but also avoids lifting or repositioning of the entire foot. AnkleType mainly relies on ankle movements to perform the main input action, with a small amount of knee movement for menu switching. It does not involve hip movements. Throughout the typing process, AnkleType does not require leaving the entire foot off the ground, which reduces the muscle tension and thus reduces physical demand and fatigue. Involving more joint movements requires more precise control of muscles, which increases fatigue and error-proneness. Moreover, compared with Wan et al. [68], our design also adapts to standing posture, which further highlights the importance of reducing the number of participating joints when designing foot-based input techniques.

8.4 The Speed and Accuracy Trade-Off

Our study further highlights a speed-accuracy trade-off in foot-based text entry. Wan et al. [68] explicitly suggested for alternative use of feet to reduce fatigue. Our study extends this perspective and decomposes the typing input into two sub-tasks, namely navigation and control. In the unipedal schema, one foot was used to carry both sub-tasks. This well-integrated interaction task in one foot leads to efficient execution, but it would cause higher fatigue and result in higher error rates. In contrast, the bipedal schema distributes the two sub-tasks across both feet. This subdivision reduces the task load on each foot and thus alleviates user fatigue. But alternative use between the two feet leads to higher task-switching costs, which reduces the typing efficiency but improves the reliability.

Moreover, our elicitation study reveals that user prefer to assign actions that require higher precision control, such as navigation, to their dominant foot, and leave simple control actions to their non-dominant foot. This asymmetry preference suggests that bipedal design should not simply balance workload across feet, but rather consider natural control capabilities between the dominant and non-dominant foot.

Method	Year	Modality	Affordance	WPM	TER	NCER	Performance after Long-term Train
RingText [76]	2019	head motions	hand-free	12.27	3.10%	2.25%	Yes
Blinktype [40]	2020	gaze	hand-free	13.47	10.44%	0.90%	Yes
iText [39]	2021	gaze	hand-free, invisible keyboard	13.77	less than 3% (WER)	/	Yes
FeetSymType [68]	2024	foot-based, sitting	hand-free	11.12	3.59%	0.46%	No
SkiMR [26]	2024	gaze	hand-free	12.05	/	/	Yes
AnkleType.UPStand	2025	foot-based, standing	hand-free	15.05	3.71%	0.44%	Yes
AnkleType.BPSit	2025	foot-based, sitting	hand-free	**16.70**	**2.48%**	0.36%	Yes
AnkleType.UPStand	2025	foot-based, standing	hand- and eye-free	11.15	9.91%	0.43%	Yes
AnkleType.BPSit	2025	foot-based, sitting	hand- and eye-free	12.87	8.80%	**0.34%**	Yes

Table 2: Summary of prior works focusing on hand-free text entry in VR environment. We also show AnkleType's performance for pair-wise comparison.

Figure 10: Potential application scenarios of AnkleType. ⓐ A user is browsing and managing content in a VR workspace. He grasps and holds two documents using controllers with his left and right hands by pulling the triggers. He can input the virtual sticky notes via AnkleType without putting the document down. ⓑ A user is sitting and watching a VR movie with one hand holding a bag of potato chips while the other hand is engaged in eating. He can input text via AnkleType to respond to notifications without clearing his hands. ⓒ A user wearing an MR HMD is walking in a library with books and coffee in both hands. He can quickly respond to short messages using AnkleType without freeing his hand and leaving his eye off the surroundings.

9 Limitations and Future Works

In this section, we highlight the limitations of our research and discuss their potential for future improvements.

9.1 Footedness Issue

The current design of AnkleType did not account for differences between left- and right-footed users. All participants in our study were right-foot dominant. We draw two potential directions for future work to bridge this gap: 1) explicitly design specific unipedal and bipedal strategies for left-footed users and evaluate, 2) seek adaptive mapping strategies of AnkleType to support a broader user population. Future work should also investigate how footedness affects performance, user preference, and learning effect.

9.2 Split-Keyboard Design Space for Bipedal

The design space of bipedal input strategies remains underexplored. In this work, to maintain design consistency and enable fair pair-wise comparison between unipedal and bipedal strategies, we did not consider involving the left foot for the navigation task. This remains a wider design space in the bipedal ankle-based typing context to provide richer and efficient interaction. One promising future exploration direction is adapting a split-keyboard design, where letters are distributed across both feet. Such a design strategy could better leverage the natural capabilities of bipedal input and expand the keyboard space. With this, we could assign letters more separately, thereby reducing the word ambiguity. At the same time, splitting the keyboard may shorten letter navigation distance, thereby increasing the overall input speed. Thus, we envision great potential in the split-keyboard design, especially for sitting conditions.

9.3 Evaluation in Realistic Dual-Task Scenario

AnkleType is motivated by scenarios in which users' hands are physically occupied, such as carrying or holding objects. In our experiments, the participants were not allowed to perform any action with their hands, similar to the constraint of hands cannot move when holding items. We proposed three low–cognitive-load application scenarios (Figure 10) as proof-of-concept demonstrations. However, we did not explicitly evaluate how the primary task in these scenarios affects typing performance, which points to a limitation of our study. Moreover, in more complex VR scenarios, such as gaming, users' visual or cognitive attention are mostly occupied by the primarily tasks, which may affect the performance of secondary tasks such as typing. In our longitudinal study, we found that users' typing proficiency and accuracy increased over time, while the number of keyboard layout cue requirements decreased over time as well. This trend suggests that users may develop motor automaticity through long-term practice. Prior researches shows that users' motor automaticity frees up cognitive resources for other processes [5, 52], which shows potential to support multiple-task performance [72]. We believe that evaluating AnkleType's performance in realistic applications and investigating its potential to support multi-tasking for VR text entry represent promising directions for our future work.

9.4 Evaluation in MR context

We implemented a proof-of-concept system using customized shoes with tracker and embedded sensors to evaluate AnkleType in VR environment. Beyond VR, the concept of AnkleType has potential and technically feasible to extend to other virtual environment such as AR/MR. For example, modern AR/MR HMD such as Apple Vision Pro[3] and Pico 4 Ultra[4] consist of build-in downward-facing camera array which can also be used to detect foot or ankle movements. These sensing capabilities show potentials to increase the flexibility and mobility of AnkleType to enrich the application scenario. We have not yet conducted experiments in either AR or MR environments, which will be our future work.

10 Conclusion

With AnkleType, we primarily contributed a new design space to support foot-based text entry in VR, with a series of user studies and evaluations in this new context. AnkleType is a novel hand- and eye-free text entry technique for VR that leverages ankle-based gestures to support both standing and sitting postures. Through two preliminary studies, we identified users' ankle movement ranges and elicited user-preferred gestures, which guide the design of our input strategies. We further optimized the keyboard layout by combining a user study on users' natural ankle spatial awareness with a computer-simulated language model to reduce word ambiguity. Through a pair-wise comparison user study across 4 user-defined input strategies candidates, we yielded two optimal strategies across postures: **UPStand** and **BPSit**, with consideration of both input performance and user preference. Finally, a 7-day longitudinal study demonstrated that participants could achieve promising eye-free typing performance, at 11.15 WPM for **UPStand** and 12.87 WPM for **BPSit**, with error rates decreasing over time, achieving a minimum TER of 9.91% for **UPStand** and 8.80% for **BPSit**. Through an offline comparison with previous works, we showed that AnkleType not only achieves state-of-the-art performance in hand-free VR text entry tasks, but also offers competitive performance in eye-free VR text entry scenarios.

[3]https://support.apple.com/en-us/117810
[4]https://www.picoxr.com/global/products/pico4-ultra/specs

Acknowledgments

This research is partially supported by the National Natural Science Foundation of China, Young Scientists Fund (Project No. 62402301), Natural Science Foundation of Guangdong Province, General Research Fund (Project No. 2025A1515010236), and the STU Scientific Research Initiation Grant (SRIG, Project No. NTF23024). This research is also partially supported by the Centre for Applied Computing and Interactive Media (ACIM) of School of Creative Media, City University of Hong Kong.

References

[1] Jiban Adhikary and Keith Vertanen. 2021. Text entry in virtual environments using speech and a midair keyboard. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2648–2658.
[2] Mehmet Akhoroz and Caglar Yildirim. 2024. Poke Typing: Effects of Hand-Tracking Input and Key Representation on Mid-Air Text Entry Performance in Virtual Reality. In *Proceedings of the 26th International Conference on Multimodal Interaction*. ACM, New York, NY, USA, 293–301.
[3] American National Corpus. 2025. https://anc.org/.
[4] Christopher R Austin, Barrett Ens, Kadek Ananta Satriadi, and Bernhard Jenny. 2020. Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality. *Cartography and Geographic Information Science* 47, 3 (2020), 214–228.
[5] Sian L Beilock and Thomas H Carr. 2001. On the fragility of skilled performance: What governs choking under pressure? *Journal of experimental psychology: General* 130, 4 (2001), 701.
[6] Doug A Bowman, Christopher J Rhoton, and Marcio S Pinho. 2002. Text input techniques for immersive virtual environments: An empirical comparison. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 46. SAGE Publications, Los Angeles, CA, USA, 2154–2158.
[7] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 473–480.
[8] Stuart K. Card, Thomas P. Moran, and Allen Newell. 1983. *The Psychology of Human-Computer Interaction*. CRC Press, Boca Raton, FL, USA.
[9] Liwei Chan, Tzu-Wei Mi, Zhung Hao Hsueh, Yi-Ci Huang, and Ming Yun Hsu. 2024. Seated-WIP: Enabling walking-in-place locomotion for stationary chairs in confined spaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery (ACM), New York, NY, USA, 1–13.
[10] Taizhou Chen, Lantian Xu, Xianshan Xu, and Kening Zhu. 2021. Gestonhmd: Enabling gesture-based interaction on low-cost vr head-mounted display. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2597–2607.
[11] Sebastian Cmentowski, Sukran Karaosmanoglu, Lennart E Nacke, Frank Steinicke, and Jens Harald Krüger. 2023. Never skip leg day again: Training the lower body with vertical jumps in a virtual reality exergame. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
[12] James J Cummings, Alexis Shore Ingber, and Yihan Danny Jia. 2025. Self-Disclosure in Social Virtual Reality: The Influence of Information Management

Dynamics, Social Presence, and Privacy Concerns. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

[13] Rajkumar Darbar, Xuning Hu, Xinan Yan, Yushi Wei, Hai-Ning Liang, Wenge Xu, and Sayan Sarcar. 2024. OnArmQWERTY: An Empirical Evaluation of On-Arm Tap Typing for AR HMDs. In *Proceedings of the 2024 ACM Symposium on Spatial User Interaction*. 1–12.

[14] Tafadzwa Joseph Dube, Kevin Johnson, and Ahmed Sabbir Arif. 2022. Shapeshifter: Gesture Typing in Virtual Reality with a Force-based Digital Thimble. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–9.

[15] John Dudley, Hrvoje Benko, Daniel Wigdor, and Per Ola Kristensson. 2019. Performance envelopes of virtual keyboard text input strategies in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 289–300.

[16] Saba Fallah and Scott Mackenzie. 2023. H4VR: One-handed gesture-based text entry in virtual reality using a four-key keyboard. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[17] Yasmin Felberbaum and Joel Lanir. 2018. Better understanding of foot gestures: An elicitation study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[18] Hyunjae Gil, Ashish Pratap, Iniyan Joseph, and Jin Ryong Kim. 2025. PropType: Everyday Props as Typing Surfaces in Augmented Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[19] Hyunjae Gil, Yonghwan Shin, Hyungki Son, Inwook Hwang, Ian Oakley, and Jin Ryong Kim. 2020. Characterizing In-Air Eyes-Free Typing Movements in VR. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*. 1–10.

[20] Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2687–2696.

[21] Mayank Goel, Alex Jansen, Travis Mandel, Shwetak N Patel, and Jacob O Wobbrock. 2013. ContextType: using hand posture information to improve mobile touch screen text entry. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2795–2798.

[22] Joshua Goodman, Gina Venolia, Keith Steury, and Chauncey Parker. 2002. Language modeling for soft keyboards. In *Proceedings of the 7th international conference on Intelligent user interfaces*. 194–195.

[23] Jens Emil Sloth Grønbæk, Juan Sánchez Esquivel, Germán Leiva, Eduardo Velloso, Hans Gellersen, and Ken Pfeuffer. 2024. Blended whiteboard: Physicality and reconfigurability in remote mixed reality collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.

[24] Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Text entry in immersive head-mounted display-based virtual reality using standard keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 159–166.

[25] Misato Hide, Yuji Hatada, Hideaki Kuzuoka, and Takuji Narumi. 2025. " Closer than Real": How Social VR Platform Features Influence Friendship Dynamics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.

[26] Jinghui Hu, John J Dudley, and Per Ola Kristensson. 2024. Skimr: Dwell-free eye typing in mixed reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 439–449.

[27] Haiyan Jiang and Dongdong Weng. 2020. Hipad: Text entry for head-mounted displays using circular touchpad. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 692–703.

[28] Snehanjali Kalamkar, Verena Biener, Daniel Pauls, Leon Lindlein, Morteza Izadifar, Per Ola Kristensson, and Jens Grubert. 2024. Accented Character Entry Using Physical Keyboards in Virtual Reality. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 662–670.

[29] Minji Kim, Kyungjin Lee, Rajesh Balan, and Youngki Lee. 2023. Bubbleu: Exploring augmented reality game design with uncertain ai-based interaction. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–18.

[30] Taeyong Kim, Hao Ju, and Jeremy R Cooperstock. 2018. Pressure or movement? Usability of multi-functional foot-based interfaces. In *Proceedings of the 2018 designing interactive systems conference*. 1219–1227.

[31] Taejun Kim, Amy Karlson, Aakar Gupta, Tovi Grossman, Jason Wu, Parastoo Abtahi, Christopher Collins, Michael Glueck, and Hemant Bhaskar Surale. 2023. Star: Smartphone-analogous typing in augmented reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[32] Lisa-Marie Krause and Oliver Herbort. 2024. Perception of pointing gestures in 3D space. *Scientific Reports* 14, 1 (2024), 27595.

[33] George Lakoff. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press, Chicago, IL.

[34] Aurélien Léchappé, Ross Johnstone, Aurélien Milliat, John H Williamson, Mathieu Chollet, and Julie R Williamson. 2025. Understanding Social Interactions in Reality Versus Virtuality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[35] Lik Hang Lee, Kit Yung Lam, Tong Li, Tristan Braud, Xiang Su, and Pan Hui. 2019. Quadmetric optimized thumb-to-finger interaction for force assisted one-handed text entry on mobile headsets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–27.

[36] Jiaye Leng, Lili Wang, Xiaolong Liu, Xuehuai Shi, and Miao Wang. 2022. Efficient flower text entry in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3662–3672.

[37] Zhuojun Li, Chun Yu, Yizheng Gu, and Yuanchun Shi. 2023. ResType: Invisible and Adaptive Tablet Keyboard Leveraging Resting Fingers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[38] Daniel Lopes, Filipe Relvas, Soraia Paulo, Yosra Rekik, Laurent Grisoni, and Joaquim Jorge. 2019. Feetiche: Feet input for contactless hand gesture interaction. In *Proceedings of the 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*. 1–10.

[39] Xueshi Lu, Difeng Yu, Hai-Ning Liang, and Jorge Goncalves. 2021. itext: Hands-free text entry on an imaginary keyboard for augmented reality systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 815–825.

[40] Xueshi Lu, Difeng Yu, Hai-Ning Liang, Wenge Xu, Yuzheng Chen, Xiang Li, and Khalad Hasan. 2020. Exploration of hands-free text entry techniques for virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 344–349.

[41] Yiqin Lu, Chun Yu, Xin Yi, Yuanchun Shi, and Shengdong Zhao. 2017. Blindtype: Eyes-free text entry on handheld touchpad by leveraging thumb's muscle memory. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–24.

[42] I Scott MacKenzie and R William Soukoreff. 2003. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*. 754–755.

[43] Johnson Mark. 1987. The body in the mind: The bodily basis of meaning, imagination, and reason.

[44] Omar Merhi, Elise Faugloire, Moira Flanagan, and Thomas A Stoffregen. 2007. Motion sickness, console video games, and head-mounted displays. *Human factors* 49, 5 (2007), 920–934.

[45] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, MC Schraefel, and Jacob O Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3 (2014), 40–45.

[46] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the tap: Assessing foot-taps for interacting with head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[47] Florian Müller, Daniel Schmitt, Andrii Matviienko, Dominik Schön, Sebastian Günther, Thomas Kosch, and Martin Schmitz. 2023. Tictactoes: Assessing toe movements as an input modality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[48] Ruowen Niu, Ruishen Zheng, Chen Liang, and Minghui Liu. 2025. Exploring Joint Effects of Locomotion Continuity and Wayfinding Assistance in Non-Embodied VR Game Navigation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[49] Donald A. Norman. 2002. *The Design of Everyday Things*. Basic Books, Inc., USA.

[50] Nels Numan, Gabriel Brostow, Suhyun Park, Simon Julier, Anthony Steed, and Jessica Van Brummelen. 2025. CoCreatAR: Enhancing authoring of outdoor augmented reality experiences through asymmetric collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.

[51] Duc-Minh Pham and Wolfgang Stuerzlinger. 2019. Hawkey: Efficient and versatile text entry for virtual reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. 1–11.

[52] Russell A Poldrack, Fred W Sabb, Karin Foerde, Sabrina M Tom, Robert F Asarnow, Susan Y Bookheimer, and Barbara J Knowlton. 2005. The neural correlates of motor skill automaticity. *Journal of Neuroscience* 25, 22 (2005), 5356–5364.

[53] Ryan Qin, Suwen Zhu, Yu-Hao Lin, Yu-Jung Ko, and Xiaojun Bi. 2018. Optimal-t9: An optimized t9-like keyboard for small touchscreen devices. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. 137–146.

[54] Vijay Rajanna, Murat Russel, Jeffrey Zhao, and Tracy Hammond. 2022. PressTapFlick: Exploring a gaze and foot-based multimodal approach to gaze typing. *International Journal of Human-Computer Studies* 161 (2022), 102787.

[55] Lovis Schwenderling, Maximilian Schotte, Fabian Joeres, Florian Heinrich, Laura Isabel Hanke, Florentine Huettl, Tobias Huber, and Christian Hansen. 2025. Teach Me Where to Look: Dual-task Attention Training in Augmented Reality. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[56] Jeremy Scott, David Dearman, Koji Yatani, and Khai N Truong. 2010. Sensing foot gestures from the pocket. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 199–208.

[57] Rongkai Shi, Hai-Ning Liang, Yu Wu, Difeng Yu, and Wenge Xu. 2021. Virtual reality sickness mitigation methods: A comparative study in a racing game. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 4, 1 (2021), 1–16.

[58] Meng Ting Shih, Chun-Jui Chou, Tzu-Wei Mi, and Liwei Chan. 2025. SeeThroughBody: Mitigating Occlusion through Body Transparency to Enhance Foot-Floor Touch Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.

[59] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[60] Paul Streli, Mark Richardson, Fadi Botros, Shugao Ma, Robert Wang, and Christian Holz. 2024. TouchInsight: Uncertainty-aware Rapid Touch and Text Input for Mixed Reality from Egocentric Vision. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[61] S Sulaiman et al. 2008. Tangisoft: tangibles, software and fun. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*. ACM, New York, NY, USA, 263–266.

[62] Yanbo Tao, Tin Lun Lam, Huihuan Qian, and Yangsheng Xu. 2012. A real-time intelligent shoe-keyboard for computer input. In *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1488–1493.

[63] Hussain Tinwala and I Scott MacKenzie. 2009. Eyes-free text entry on a touchscreen phone. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*. IEEE, 83–88.

[64] Vincent van Rheden, Sasindu Abewickrema, Florian 'Floyd' Mueller, and Don Samitha Elvitigala. 2025. GestureSock: Exploring toe gestures as alternative input method. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[65] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In *Human-Computer Interaction INTERACT'97: IFIP TC13 International Conference on Human-Computer Interaction, 14th–18th July 1997, Sydney, Australia*. Springer, 509–516.

[66] Eduardo Velloso, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Interactions under the desk: A characterisation of foot movements for input in a seated position. In *IFIP Conference on Human-Computer Interaction*. Springer, 384–401.

[67] Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 1–35.

[68] Tingjie Wan, Liangyuting Zhang, Hongyu Yang, Pourang Irani, Lingyun Yu, and Hai-Ning Liang. 2024. Exploration of foot-based text entry techniques for virtual reality environments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[69] Jialin Wang, Hai-Ning Liang, Diego Monteiro, Wenge Xu, and Jimin Xiao. 2022. Real-time prediction of simulator sickness in virtual reality games. *IEEE Transactions on Games* 15, 2 (2022), 252–261.

[70] Kieran Watson, Robin Bretin, Mohamed Khamis, and Florian Mathis. 2022. The feet in human-centred security: Investigating foot-based user authentication for public displays. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–9.

[71] Qianjie Wei, Xiaoying Wei, Yiqi Liang, Fan Lin, Nuonan Si, and Mingming Fan. 2025. RemoteChess: Enhancing Older Adults' Social Connectedness via Designing a Virtual Reality Chinese Chess (Xiangqi) Community. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

[72] Christopher D Wickens. 2020. Processing resources and attention. In *Multiple task performance*. CRC Press, 3–34.

[73] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1083–1092.

[74] Pui Chung Wong, Kening Zhu, and Hongbo Fu. 2018. Fingert9: Leveraging thumb-to-finger interaction for same-side-hand text entry on smartwatches. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[75] Jian Wu, Ziteng Wang, Lili Wang, Yuhan Duan, and Jiaheng Li. 2024. FanPad: a fan layout touchpad keyboard for text entry in VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 222–232.

[76] Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Tianyu Zhang, Difeng Yu, and Diego Monteiro. 2019. Ringtext: Dwell-free and hands-free text entry for mobile head-mounted displays using head motions. *IEEE transactions on visualization and computer graphics* 25, 5 (2019), 1991–2001.

[77] Zheer Xu, Weihao Chen, Dongyang Zhao, Jiehui Luo, Te-Yen Wu, Jun Gong, Sicheng Yin, Jialun Zhai, and Xing-Dong Yang. 2020. Bitiptext: Bimanual eyes-free text entry on a fingertip keyboard. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[78] Zheer Xu, Pui Chung Wong, Jun Gong, Te-Yen Wu, Aditya Shekhar Nittala, Xiaojun Bi, Jürgen Steimle, Hongbo Fu, Kening Zhu, and Xing-Dong Yang. 2019. Tiptext: Eyes-free text entry on a fingertip keyboard. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 883–899.

[79] Linghui Ye, Frode Eika Sandnes, and I Scott MacKenzie. 2020. QB-Gest: qwerty bimanual gestural input for eyes-free smartphone text input. In *International Conference on Human-Computer Interaction*. Springer, 223–242.

[80] Xin Yi, Chun Yu, Mingrui Zhang, Sida Gao, Ke Sun, and Yuanchun Shi. 2015. Atk: Enabling ten-finger freehand typing in air based on 3d hand tracking data. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 539–548.

[81] Chun Yu, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. Tap, dwell or gesture? exploring head-based text entry techniques for hmds. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4479–4488.

[82] Difeng Yu, Kaixuan Fan, Heng Zhang, Diego Monteiro, Wenge Xu, and Hai-Ning Liang. 2018. PizzaText: Text entry for virtual reality systems using dual thumbsticks. *IEEE transactions on visualization and computer graphics* 24, 11 (2018), 2927–2935.

[83] Lian Yue, Lu Zongxing, Dong Hui, Jia Chao, Liu Ziqiang, and Liu Zhoujie. 2023. How to achieve human–machine interaction by foot gesture recognition: A review. *IEEE Sensors Journal* 23, 15 (2023), 16515–16528.

[84] Mingrui Ray Zhang, Shumin Zhai, and Jacob O Wobbrock. 2022. TypeAnywhere: A QWERTY-based text entry solution for ubiquitous computing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[85] Kang Zhong, Feng Tian, and Hongan Wang. 2011. Foot menu: Using heel rotation information for menu selection. In *2011 15th Annual International Symposium on Wearable Computers*. IEEE, 115–116.

[86] Suwen Zhu, Tianyao Luo, Xiaojun Bi, and Shumin Zhai. 2018. Typing on an invisible keyboard. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[87] Suwen Zhu, Jingjie Zheng, Shumin Zhai, and Xiaojun Bi. 2019. i'sFree: Eyes-free gesture typing via a touch-enabled remote control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.