

Handwriting for Text Input and the Impact of XR Displays, Surface Alignments, and Sentence Complexities

Florian Kern , Jonathan Tschanter , and Marc Erich Latoschik 

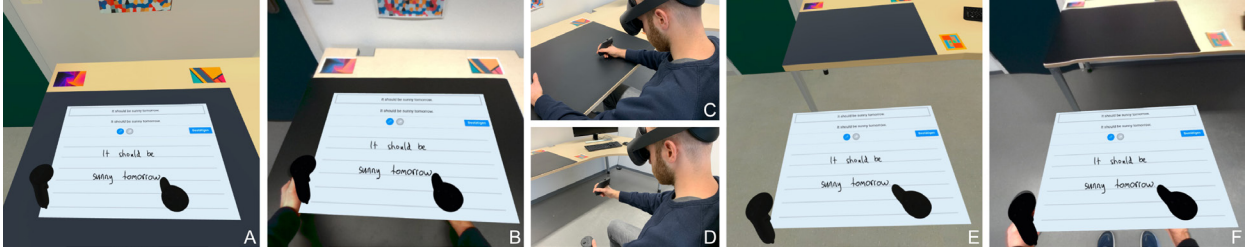


Fig. 1: In our study, participants used handwriting for text input in Virtual Reality (VR) (A, E) or Video See-Through Augmented Reality (VST AR) (B, F) on physically aligned (A, B, C) and mid-air (D, E, F) surfaces. For VR, we created a digital twin of the real room and aligned it precisely. The controller was held in a pen-like posture to mimic a stylus. To detect surface contact, we used the controller's pressure-sensitive tip on the physical surface and an orthogonal distance-based approach in mid-air. Participants wrote simple and complex sentences. The example shows the simple sentence, "It should be sunny tomorrow."

Abstract—Text input is desirable across various eXtended Reality (XR) use cases and is particularly crucial for knowledge and office work. This article compares handwriting text input between Virtual Reality (VR) and Video See-Through Augmented Reality (VST AR), facilitated by physically aligned and mid-air surfaces when writing simple and complex sentences. In a 2x2x2 experimental design, 72 participants performed two ten-minute handwriting sessions, each including ten simple and ten complex sentences representing text input in real-world scenarios. Our developed handwriting application supports different XR displays, surface alignments, and handwriting recognition based on digital ink. We evaluated usability, user experience, task load, text input performance, and handwriting style. Our results indicate high usability with a successful transfer of handwriting skills to the virtual domain. XR displays and surface alignments did not impact text input speed and error rate. However, sentence complexities did, with participants achieving higher input speeds and fewer errors for simple sentences (17.85 WPM, 0.51% MSD ER) than complex sentences (15.07 WPM, 1.74% MSD ER). Handwriting on physically aligned surfaces showed higher learnability and lower physical demand, making them more suitable for prolonged handwriting sessions. Handwriting on mid-air surfaces yielded higher novelty and stimulation ratings, which might diminish with more experience. Surface alignments and sentence complexities significantly affected handwriting style, leading to enlarged and more connected cursive writing in both mid-air and for simple sentences. The study also demonstrated the benefits of using XR controllers in a pen-like posture to mimic styluses and pressure-sensitive tips on physical surfaces for input detection. We additionally provide a phrase set of simple and complex sentences as a basis for future text input studies, which can be expanded and adapted.

Index Terms—VR, AR, XR, physically aligned, mid-air, text input, handwriting, digital ink, recognition, phrase set, digital twin

1 INTRODUCTION

Text input is desirable across various eXtended Reality (XR) use cases and is particularly crucial for knowledge and office work. XR overcomes the physical limitations of traditional seated and stationary workspaces, enabling flexible and portable work in Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), for short, XR. While XR workspaces can be employed everywhere [6,24,33,42,43,50], they also require versatile text input techniques. Despite virtual and physical keyboards, handwriting text input has great potential, offering unique advantages like natural intuitiveness, flexibility, and expressiveness. Handwriting in XR also benefits from prior knowledge acquired through the use of pen and paper or styluses on tablets. Furthermore, handwriting enhances cognitive engagement, improves understanding, and supports memory retention [51,53], making it valuable for education and work. For VR and Optical See-Through (OST) AR, previous

work demonstrated the potential of handwritten text input using handwriting recognition [18,19,21,23,65]. Nevertheless, handwriting text input in Video See-Through (VST) AR remains underexplored. Handwriting typically takes place on physical surfaces, providing passive haptic feedback that supports fine-grained movement, improves performance [15,67], and reduces arm fatigue [31,71,72]. Consequently, physical surfaces are also integrated in XR. However, text input on mid-air surfaces is essential for flexible and non-stationary XR use cases. While prior work evaluated handwriting text input on mid-air surfaces using controllers [18,21,23,65] or fingers [65], physical surfaces were partially investigated [23,58,65], leaving a fundamental research gap. Furthermore, text input studies often include transcription tasks, requiring participants to copy predefined sentences. While participants in previous handwriting studies transcribed simple text phrases by writing individual letters [18,19,58], words [21,65] or sentences [23], this does not necessarily represent real-world scenarios. Therefore, the impact of sentence complexities on handwriting text input in XR is unclear.

To address these gaps, we investigated the following research questions: (1) How do XR displays (VR and VST AR) affect usability, user experience, task load, text input performance, and handwriting style? (2) How do surface alignments (physically aligned and mid-air surfaces) affect usability, user experience, task load, text input performance, and handwriting style? (3) How do sentence complexities (simple and complex sentences) affect usability, user experience, task load, text input performance, and handwriting style?

- Florian Kern and Marc Erich Latoschik are with the Human-Computer Interaction (HCI) Group, Jonathan Tschanter is with the HCI group and the PIIS group from the University of Würzburg. E-mail: {florian.kern | jonathan.tschanter | marc.latoschik}@uni-wuerzburg.de.

Manuscript received 4 October 2023; revised 17 January 2024; accepted 24 January 2024. Date of publication 5 March 2024; date of current version 15 April 2024.
Digital Object Identifier no. 10.1109/TVCG.2024.3372124

This work provides novel insights into handwriting for text input in VR and VST AR, facilitated by physically aligned and mid-air surfaces when writing simple and complex sentences. We assessed handwriting performance by text input speed and error rate and identified characteristics of handwriting style using stroke-level analysis. We collected subjective user ratings through questionnaires to evaluate usability, user experience, and task load. From our results, we derived three key findings and discussed four design implications to support developers and researchers in using handwriting text input for future XR use cases. We also highlight the benefits of XR controllers as styluses and demonstrate the use of pressure-sensitive tips on physical surfaces for input detection. We additionally provide a phrase set of simple and complex sentences for text input studies, which can be expanded and adapted.

2 RELATED WORK

Advancements in VR, MR, AR, for short, XR, overcome physical limitations and enable flexible and portable workspaces that can be employed everywhere [6, 24, 33, 42, 43, 50]. While VR presents a purely virtual environment, OST AR enables users to perceive the real world with virtual content optically superimposed. In contrast, VST AR, or AR passthrough, shows the real world to the user captured by the front-facing cameras of the device and allows an opaque and integrated overlay of virtual content. MR is considered an advanced form, merging [69] or situating [62] virtual content in the real world to facilitate context-aware interaction and response to the physical surroundings. Consequently, the visual-sensory display features can affect the design of text input techniques in XR for both typing and handwriting. The most common text input solutions are physical [25, 38, 49, 55] and virtual keyboards [17, 33]. They are widely available, benefit from prior knowledge, and offer efficient text input performance. Besides the prevalence of keyboards, handwriting emerges as a promising candidate for text input in XR [18, 21, 65], offering unique advantages like natural intuitiveness, flexibility, and expressiveness.

2.1 Handwriting for Text Input

Handwriting, typically developed in childhood [22, 37], requires fine and precise movements, which activates neural processes and brain activation patterns associated with learning, creativity, and cognition [54, 63]. In digital spaces, handwriting is visualized by digital ink, offering a natural writing experience by mimicking the fluidity and responsiveness of pen and paper in the real world. Additionally, digital note-taking with a stylus and tablet showed similar benefits to pen and paper [51, 52], forming the basis for handwriting in XR. For activities like searching, organizing, and editing content, handwriting recognition is essential to transform digital ink into a textual representation. Common use cases are note-taking, document editing, entering emails, phone numbers, and passwords, or browsing the internet [14, 31, 64].

Handwriting recognition can be categorized into online and offline recognition [12, 57, 64]. Offline handwriting recognition, like Optical Character Recognition (OCR), interprets handwriting based on shapes and patterns in static data after writing is completed [12, 28]. Online handwriting recognition, or digital ink recognition, analyzes the handwritten trajectory during the writing process, considering dynamic and contextual information like position, direction, velocity, and acceleration [36]. Digital ink recognition also enables text composition and dynamic document editing via gestures. Examples are erasing by scratching out, separating or joining words using vertical lines, and formatting by underlining, italicizing, or bolding. As a result, digital ink recognition offers greater flexibility than offline handwriting recognition, making it a more suitable choice for text input in XR. Moreover, digital ink also forms the basis for stroke-level analysis in user studies.

2.2 Handwriting for Text Input in XR

In 1998, Poupyrev et al. [58] introduced the Virtual Notepad, a pioneering tool for handwritten notes in VR. Users navigated the notebook by writing individual letter commands using handwriting recognition. Although the handwriting recognition capabilities were limited, the authors acknowledged its future potential for handwritten text input.

González et al. [23] conducted a study using letter-based online handwriting recognition for text input in VR. Participants transcribed sentences using a pen and tablet placed on a physical surface in a horizontal orientation. Corrections were omitted to preserve consistent writing speeds. With a text input speed of about 2.3 words per minute (WPM) and an error rate (ER) of 23%, the study marks a successful implementation of handwriting recognition in VR. However, this work also showed the impact of recognition model constraints by requiring participants to write each letter individually and adapt their handwriting style to predefined shapes to minimize recognition errors.

Elmgren [18] investigated letter-based recognition for handwritten text input in VR. Participants transcribed a pangram using distance-based raycasting with controllers on a vertical surface in mid-air without error correction. Although the input speed of 4.16 WPM (36.86% ER) nearly doubled compared to prior work [23], it remained relatively slow. The authors explained this with letter-based handwriting and an artificial pause before recognition. Further, they noted that handwriting in mid-air via raycasting could be challenging due to hand tremors and lack of physical support, which may lead to increased error rates.

Venkatakrishnan et al. [65] evaluated the effects of input methods, canvas geometries, and inking triggers on word-based VR handwriting on vertical surfaces. They used online handwriting recognition and enabled corrections by deleting the last stroke. Handwriting with controllers in direct contact outperformed raycasting and finger use regarding input speed (10.33 WPM), accuracy (99.50%), and overall NASA TLX workload (32.33). Writing using fingers was significantly faster on physical surfaces (8.10 WPM) than on mid-air surfaces (7.23 WPM) with a trigger button. These results emphasize the impact of input methods and passive haptic feedback on handwriting performance and underline the importance of a thoughtful selection.

Fourrier et al. [21] examined handwriting text input in VR using online handwriting recognition. The study focused on vertical and slanted surfaces in mid-air and visual, auditive, and active haptic feedback. Participants wrote individual words from predefined sentences and triggered handwriting by direct contact with the mid-air surface. For all conditions, usability was considered good to excellent (80.05) using the SUS. Slanted surfaces yielded the highest text input speeds (14.15 WPM), the lowest physical demands (42.31), and the best overall NASA TLX workload scores (40.11). Multimodal feedback reduced error rates to 1.05%, although some participants felt overwhelmed by the sounds and vibrations. This suggests a potential benefit in reducing feedback cues. The authors proposed to improve text editing by reducing the erasing granularity from rewriting entire words to more fine-grained correction (e.g., strokes). Similar to previous work, this study underlines the great potential of handwriting text input in VR, with high usability, low task load, and reasonable error rates. Nevertheless, further flexibility in error correction and more adaptive and accurate handwriting recognition are desirable.

For OST AR, Fang et al. [19] developed handwriting velcro, a wearable system with touch sensors for fingers using letter-based handwriting recognition. Participants achieved 12.32 WPM (97.21% accuracy), highlighting the applicability of AR handwriting on physical surfaces.

However, handwriting in VR and OST AR was still notably slower than with pen and paper at 30.36 WPM [21] and 21.9 WPM [11] for transcription tasks, 31.1 WPM [11] with memorized sentences, and 21.5 WPM [39] with styluses on laptop touchscreens.

The evident room for improvement and the merging boundaries between VR and VST AR require a profound understanding of impact factors. Handwriting has been explored in VR, but remained underexplored in VST AR, particularly regarding visual incongruencies [33, 41, 44]. The impact of surface alignments also needs more investigation, with previous research preferring physical surfaces over mid-air surfaces for tasks requiring precision, fine-grained movements [5, 56], and low physical demand [71, 72]. Furthermore, prior work limited handwriting text input to transcribing simple text phrases by writing individual letters [18, 19, 58], words [21, 65] or sentences [23]. While writing individual entities or simple sentences represents fundamental parts of text input, it does not necessarily reflect the complexity of text input in real-world scenarios.

2.3 Visual Incongruencies in AR

Handwriting text input has been explored in VR [18, 21, 23, 65] and OST AR [19], but remained underexplored in VST AR. In VST AR, a significant challenge is the visual augmentations of the real-world environment, which can lead to visual incongruencies [33, 44], impacting performance and user experience [4, 33]. Common issues are depth distortion and object misplacement [41, 59, 68], forming the basis for many other visual incongruencies experienced by users. Adams et al. [1] found that depth perception is a fundamental aspect of AR and varies across displays. They explain differences by the diverse cues and characteristics specific to each AR device, which can affect distance judgment. Additionally, factors like misalignment in VST AR, magnification or minification effects through cameras, and even the weight of Head-Mounted Displays (HMDs) could significantly influence users' spatial understanding. Pham and Stuerzlinger [56] revealed performance discrepancies between VR and OST AR, with VR users significantly outperforming their AR counterparts in pointing tasks, likely due to fewer visual inconsistencies. Similarly, Kern et al. [33] observed faster text input in VR compared to VST AR when using virtual tap and swipe keyboards. The authors attributed this difference to visual mismatches inherent in VST AR that can affect depth perception and, consequently, performance and user experience. While these studies showed that visual incongruencies in AR can influence performance and user experience, their impact on handwriting for text input in XR needs to be investigated.

2.4 Physical and Mid-Air Surfaces

Physical surfaces are omnipresent in traditional knowledge and office work and will be leveraged in future XR workspaces. Physical surfaces support precise and fine-grained movements, which are recommended for tasks like sketching [3, 13], handwriting [30, 31], and interactions with physical and virtual keyboards [17, 38]. In contrast, mid-air surfaces offer more flexibility but lack the passive haptic feedback and stability of physical surfaces, linked to higher performance [15, 67] and reduced task load, particularly due to arm fatigue [31, 71, 72]. For handwritten text input in VR, previous work investigated the impact of surface alignments and orientations. Venkatakrishnan et al. [65] found that VR handwritten text input with fingers was significantly more performant on physical surfaces compared to mid-air surfaces using a trigger button. Employing controllers for handwriting on mid-air surfaces resulted in higher performance and lower overall workload than finger use. This suggests an influence of both input methods and physical and mid-air surfaces. Fourrier et al. [21] showed that surface orientation also significantly influenced mid-air handwriting. Slanted surfaces achieved better results than vertical surfaces regarding performance and physical demand using controllers. Horizontal surfaces were not examined due to usability issues when standing. These findings indicate that handwriting on physical surfaces outperforms mid-air surfaces and that controllers are superior to fingers. Additionally, surface orientations towards more horizontal inclinations seem preferable. Notably, these studies did not examine handwriting with controllers on physical surfaces and horizontal surface orientations as well as finger-based raycasting, leaving substantial research gaps.

2.5 Sentence Complexities

Evaluating handwriting for text input in XR often involved transcription tasks [40], in which participants had to write simple text phrases by individual letters [18, 19, 58], words [21, 65] or sentences [23]. Although writing letters or words is a fundamental part of text input, it does not necessarily reflect complex real-world scenarios. Phrase sets address this limitation by providing different levels of sentence complexity. The MacKenzie and Soukoreff phrase set [47] provides simple sentences, omitting capitalization and punctuation. In contrast, the EnronMobile phrase set created by Vertanen and Kristensson [66] contains simple and complex phrases with uppercase and lowercase letters, digits, punctuation marks, and other symbols. Curran et al. [16] investigated sentence complexities for text input on mobile phones. They used simple phrases with lowercase alphabetical characters. Complex phrases also included uppercase, digits, parentheses, punctuation marks, and other symbols.

The results showed that simple phrases led to higher typing speed and lower error rates than complex phrases. These findings are in line with results from Pham and Stuerzlinger [55] evaluating physical keyboards, where participants typed lowercase simple sentences significantly faster than complex sentences in letter case, digits, parentheses, punctuation marks, and other symbols. While keyboards require the use of modifier keys for uppercase or digits, handwriting faces interruptions due to single strokes for complex units (e.g., numbers, parentheses, or colons). Therefore, simple sentences might lead to more fluent and connected cursive handwriting than complex sentences, which can result in higher performance and adapted handwriting style regarding the number, length, and height of strokes. However, the impact of sentence complexity on XR handwriting text input was not evaluated.

2.6 XR Controllers as Styluses

Previous work successfully used controllers in a power grip for handwriting text input in VR [18, 21, 65], the traditional gripping technique of input devices [5, 31]. However, handwriting tools are usually held in a pen-like posture, the so-called precision grip, enabling fine-grained and intricate movements using wrist and fingers [31, 61]. While the power grip offers more stability, it restricts hand and wrist mobility, requires more arm strength, and involves larger movements than fine adjustments with fingers [5, 31]. Consequently, prior work recommend using XR styluses and controllers in precision grip for tasks requiring fine-grained movements and performance. Examples are pointing [5, 56], drawing [13], handwriting [31], and keyboard typing [33]. Additionally, input devices can support pressure-sensitive stylus tips, like the Logitech VR Ink or Meta Quest Touch Pro Controllers, which are highly beneficial for input detection on physical surfaces [8, 9]. Although the potential of XR controllers as styluses is clearly evident, their use for handwriting text input in XR remains underexplored.

3 METHODS

Our study evaluated the impact of XR displays, surface alignments, and sentence complexities on handwriting text input in a structured study using qualitative questionnaires and quantitative performance measures. An ethics application was submitted to the ethics committee of the Institute for Human-Machine-Media (MCM) at the University of Würzburg, which found the study to be ethically unobjectionable.

3.1 Hypotheses

For our first research question, addressing the impact of XR displays, we expect higher text input speed, usability, and user experience in VR than in VST AR.

H1.1: VR leads to higher text input speed (WPM) than VST AR.

H1.2: VR leads to higher usability (SUS) than VST AR.

H1.3: VR leads to higher user experience (UEQ) than VST AR.

With our second research question focusing on surface alignments, we expect higher text input speed, usability, and user experience ratings for physically aligned surfaces (PA) and increased physical demand for mid-air surfaces (MA). Furthermore, we expect an enlarged and more connected cursive handwriting style in mid-air.

H2.1: PA lead to higher text input speed (WPM) than MA.

H2.2: PA lead to higher usability (SUS) than MA.

H2.3: PA lead to higher user experience (UEQ) than MA.

H2.4: MA lead to higher physical demand (RTLX) than PA.

H2.5: PA lead to higher stroke counts than MA.

H2.6: MA lead to higher stroke lengths than PA.

H2.7: MA lead to higher stroke heights than PA.

For our third research question, targeting the impact of sentence complexities, we expect higher text input speed and fewer errors for simple sentences (SS) than for complex sentences (CS). We also expect an enlarged and more connected cursive handwriting style when writing simple sentences than complex sentences.

H3.1: SS lead to higher text input speed (WPM) than CS.

H3.2: CS lead to higher error rates (MSD ER) than SS.

H3.3: CS lead to higher stroke counts than SS.

H3.4: SS lead to higher stroke lengths than CS.

H3.5: SS lead to higher stroke heights than CS.

3.2 Participants

We recruited 72 participants via the university's participant management system. They received either 15 euros or student credit points. Inclusion criteria were fluency in German and an age of 18 years or older. We utilized the Simulator Sickness Questionnaire (SSQ) [29] to assess well-being. No participants were excluded due to a high difference before and after the experiment or consistently high scores. Participants' ages ranged from 18 to 48 years, with an average of 23.79 years and a standard deviation of 4.90 years. Regarding social gender, 40 identified as female, 32 as male, and none as non-binary. 69 were right-handed and three left-handed. The vast majority, 61 participants, were students. Four were employed, two were job-seekers, two were pupils, and one had other employment. 69 participants reported not being color-blind, while three had red-green blindness. None reported hearing impairments. 41 had no visual impairment, 21 used contact lenses, seven wore glasses, and three had uncorrected visual conditions. All participants reported fluency in the German language, the ability to think arithmetically, and no impairments in wrist movements. Two reported a limitation in literacy skills and one participant in linguistic abilities. Based on experiment observations, these participants were not excluded. All participants used handwriting with pen and paper before, with 71 having more than 20 hours of experience. For handwriting on tablets, 30 participants have more than 20 hours of experience, while four have no experience. 17 participants used tablets with handwriting recognition for more than 5 hours. 17 participants never used handwriting recognition, and four never tried handwriting on a tablet. Of 67 participants who used VR and 44 who used AR, only 26 spent more than 5 hours in VR and only four in AR. Conversely, five participants never tried VR before, and 28 had no experience with AR. Interestingly, nine participants used handwriting in VR and five in AR, although 65 in VR and 68 in AR had no experience with handwriting recognition.

3.3 Apparatus

The study was conducted in a 20m² room to ensure adequate distance between the experimenter and participants. Participants were seated to enhance safety (see Figure 1). In the physically aligned surface condition, participants were positioned directly at the physical table, while in the mid-air condition, they were located one meter away from the table. On the physical table, we used the black Logitech VR Ink Drawing Mat in DIN A1 (594mm x 841mm) as writing surface, providing low and uniform friction and increasing future applicability.

3.3.1 Hardware Environment

We used the Meta Quest Pro with two Meta Quest Touch Pro controllers. A unique feature of the controllers is the native pressure-sensitive stylus tip. This feature makes the Meta Quest Pro an outstanding device for applications that require precise tracking and reliable contact detection on physical surfaces. To ensure sufficient battery life and avoid interruptions, we charged the Meta Quest Pro and its controllers with the docking station between experiment sessions. Our technical setup included an XMG Pro 17 laptop with an Intel Core i7-10875H CPU, 16GB DDR4 RAM, an NVIDIA GeForce RTX 2070 Super GPU, and a 1TB Samsung 970 EVO Plus NVMe M.2 SSD. Questionnaires were answered on another desktop computer using mouse and keyboard. We used a self-hosted instance of LimeSurvey 4 to collect responses.

3.3.2 Software Environment

We developed two applications: an XR handwriting application optimized for the Meta Quest Pro running on Android and an experimenter application using Microsoft Windows 10, enabling monitoring and management. Experimental log files were recorded on the XR device.

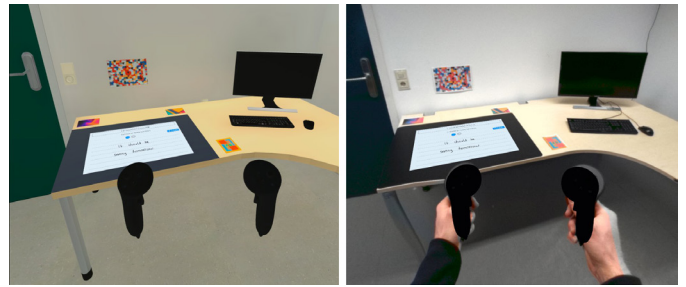


Fig. 2: In VR, we showed a precisely aligned digital twin, which we created with Blender (left). In VST AR, the real world was captured by the front-facing cameras of the Meta Quest Pro (right).

We chose Unity 2020.3 LTS for both applications. We imported the Oculus integration asset and based our implementation on the Reality Stack I/O framework [32]. For providing web browser functionality, we used the Vuplex 3D WebView asset for Android and iOS. We opted for MyScript, an online handwriting recognition solution based on digital ink, and developed a web interface. We used the Exit Games Photon PUN2 asset and a self-hosted server instance to create a network connection between the Meta Quest Pro and the experimenter application. A stable Wi-Fi 6 wireless network connection was provided by the mobile hotspot capability of the laptop.

A precise alignment of the virtual surface on the physical table is essential for handwriting with passive haptic feedback. Therefore, we decided to use the publicly available Circle Refinement Technique (CRT) introduced by Kern et al. [35], which enhances the original 3ViSuAI technique [31] by recording numerous measure points in a circular movement to improve precision. We also installed four colored visual markers to support tracking stability.

The handwriting controller was held in a pen-like posture, the precision grip, to mimic a stylus, allowing more precise movements than the regular power grip [5, 31, 56] used in prior research [18, 21, 65]. The other controller was held in power grip and had no interaction functionality but provided a visual reference point. Both virtual controller representations were visible in VR and VST AR.

Previous studies employed forward-direction raycasting to identify surface contact points for controllers and fingers [18, 65]. While this approach is typically used for distance-based interaction, it also introduces significant cursor movements with even a small controller/wrist displacement at the original position. Instead, we determined contact points by an orthogonal projection of the stylus tip on the virtual surface, preventing cursor displacements. We detected input intention for handwriting on mid-air surfaces by the orthogonal distance between the stylus tip and the virtual surface (i.e., less than 1cm in front and 5cm behind). For handwriting on the physical surface, we utilized the pressure-sensitive stylus tips of the Meta Quest Touch Pro controllers.

We provided visual feedback through digital ink on the virtual surface and added continuous controller vibrations when writing on mid-air surfaces. Passive haptic feedback was present when using physically aligned surfaces. Surface snapping of the controller was not implemented to prevent feeling disconnected when lifting the physical controller without an immediate response from the visual counterpart [8].

The virtual surface (DIN A2, 420mm x 594mm) consists of a handwritten area with horizontal lines at a distance of 3.5 cm. The control area shows the target sentence for the transcription and the currently recognized text below it. Two round buttons are used to switch between digital inking and erasing by scratching out the produced digital ink. The blue rectangular submit button confirms the transcribed text.

We created a realistic and correctly sized digital twin of the physical room to show a similar environment in VR and VST AR (see Figure 1 and Figure 2). We used Blender 3.5 and captured real-world textures with an Apple iPhone 13 Pro. In VR, participants saw the digital twin, which was implicitly aligned by the virtual surface. In VST AR, they viewed the real environment through the stereoscopic and perspective-corrected video stream of the Meta Quest Pro's front-facing cameras.

Table 1: Our phrase set includes twenty sentences. For the ten simple sentences, we took the German translation of Kern et al. 2023 [33] based on the MacKenzie and Soukoreff corpus [47]. For complex sentences, we self-developed ten phrases based on the EnronMobile corpus [66]. Simple sentences include lower and uppercase letters, commas, and terminal punctuation at the end of a sentence. Complex sentences also contain numbers, internal punctuation within a sentence, and other symbols. German and English translations can vary due to language differences.

Simple Sentences		Complex Sentences	
German	English	German	English
Der frühe Vogel fängt den Wurm.	The early bird gets the worm.	Ruf mich an unter: +49 800 1234567	Call me at: +49 800 1234567
Was du siehst, ist was du bekommst.	What you see is what you get.	Ich arbeite in Gebäude 6 (Raum 4.025).	I work in building 6 (Room 4.025).
Bist du sicher, dass du das willst?	Are you sure you want this?	Das Frühstück & Abendessen kostet \$30.	Breakfast & dinner costs \$30.
Das ist eine sehr gute Idee.	This is a very good idea.	Ich verspäte mich um ca. 8 Minuten.	I will be about 8 minutes late.
Morgen soll es sonnig werden.	It should be sunny tomorrow.	Auf den Artikel gibt es 20% Rabatt.	There is a 20% discount on the article.
Diese Kamera macht schöne Fotos!	This camera takes nice photographs!	Die Besprechung geht von 14 - 15 Uhr.	The meeting is from 2 - 3 pm.
Lerne laufen, bevor du rennst!	Learn to walk before you run!	Mein Flug landet um 17:38 Uhr in Berlin.	My flight lands at 17:38 in Berlin.
Bewegung ist gut für den Geist!	Exercise is good for the mind!	Am 25.06.2023 gehe ich auf ein Konzert.	On 06.25.2023 I am going to a concert.
Wo habe ich meine Brille gelassen?	Where did I leave my glasses?	Eine vegetarische Pizza kostet 12,50€.	A vegetarian pizza costs 12.50€.
Kaufst du gerne am Sonntag ein?	Do you like to shop on Sunday?	Die Geschwindigkeit liegt bei 230 km/h.	The speed is at 230 km/h.

3.4 Design

We applied a 2x2x2 study design with two levels for the between-subject factor XR displays (VR and VST AR), two levels for the within-subject factor surface alignments (physically aligned and mid-air), and two levels for the within-subject factor sentence complexities (simple and complex sentences). We followed a strict procedure to minimize the experimenter’s influence, with one person conducting the entire user study. We applied the applicable COVID-19 guidelines and believe these regulations did not affect the results of our study. After each participant, the room was ventilated, and all physical surfaces, including tables and electronic equipment, were disinfected. We informed participants that they could stop the experiment at any time if they felt uncomfortable. The experimenter also asked participants about their well-being before and after each exposure. We applied a systematic counterbalance, a type of quasi-randomization when assigning participants to the experimental conditions. This improves the internal validity of our analysis and reduces order and sequence effects. Additionally, we used social gender as a blocking variable to obtain similar sample sizes. We developed a phrase set for our user study, including simple and complex sentences (see Table 1). Simple sentences were taken from the MacKenzie and Soukoreff phrase set [47] and translated by Kern et al. [33] with correction of lower and uppercase, commas, and terminal punctuation at the end of the sentence (i.e., period, question mark, and exclamation mark). Complex sentences were self-developed based on the EnronMobile corpus [66] and additionally contain numbers, internal punctuation within the sentence (i.e., hyphens, colons, and parentheses), and other symbols. In total, we collected $72 \times 22 \times 2 = 3168$ phrases.

3.5 Study Procedure

Figure 3 shows the study procedure. The duration of each study session was approximately 60 minutes. The experimenter assigned participants to either the VR or VST AR condition. In the beginning, participants had to read and give consent to the study information. They answered questions about demographics and prior experience, limitation of abilities, and completed the SSQ. Participants watched a video showing handwriting input and correction, and were instructed on using the Meta Quest Pro and the Meta Quest Touch Pro controllers. We used Meta’s fitting procedure with lens spacing adjustment. The controller chosen by the participants for handwriting, typically the dominant hand, was held upside down in a pen-like posture, the precision grip. The other controller was held in the regular power grip. Participants were advised to enter text phrases as fast and error-free as possible. By touching the round buttons at the top center of the virtual surface, they could switch between digital inking and erasing. Before each handwriting session, the experimenter aligned the virtual surface using CRT [35] and confirmed the alignment using the experimenter application on the laptop. In the training phase, participants were asked to familiarize themselves with the system by writing two training sentences, “Ein

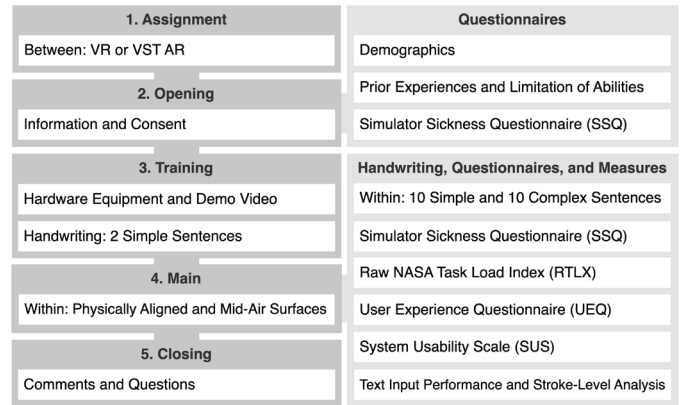


Fig. 3: Overview of the study procedure, divided into five phases.

Fuchs ist ein sehr schlaues Tier.” (“A fox is a very smart animal.”) and “Spielen wir eine Runde Karten?” (“Shall we play a round of cards?”). In the main phase, participants wrote ten simple sentences followed by ten complex sentences on each surface alignment (physically aligned and mid-air). Within each complexity, the order of sentences was randomized. The current sentence remains visible during handwriting. One handwriting session took approximately ten minutes. After each alignment, participants were asked to put down the XR device and to complete questionnaires (SSQ, RTLX, UEQ, and SUS). They also had the option to add comments or ask questions. Text input performance measures and stroke-level analysis were computed during the exposure or afterward using the experimental log files. The procedure was then repeated for the second surface alignment condition.

3.6 Measures

Text input performance was measured with Words Per Minute (WPM) and Minimum String Distance Error Rate (MSD ER), based on formulas provided by Arif and Stuerzlinger [2]. As recommended, we omitted the first character subtraction in the calculation of WPM because, in our study, the input of the first character is timed. An analysis on the stroke level provides information about the individual’s handwriting style and legibility, indicating whether the text was written in print or more connected cursive. For this, we included stroke height in millimeters (mm), representing vertical handwriting size, as well as stroke (path) length in millimeters (mm) and stroke count (also called pen lifts), indicating connected cursive handwriting. We also collected participants’ subjective feedback using a series of well-established questionnaires. We used the Simulator Sickness Questionnaire (SSQ) [29] as a control measure for discomfort or unwanted symptoms that participants might experience during VR and VST AR exposures. We chose the User Experience

Questionnaire (UEQ) [45, 60] to evaluate the overall user experience in terms of attractiveness, perspicuity, efficiency, dependability, novelty, and stimulation. To assess perceived workload, we included the NASA Raw Task Load Index (RTLX) [26, 27], assessing several dimensions of task load for mental, physical, and temporal demand, as well as performance, effort, and frustration. Finally, we measured usability with the System Usability Scale (SUS) [10] and an extension separating the SUS into a usable and learnable subscale [7, 46].

4 RESULTS

Each model was evaluated against an alpha level of .05 (*). For a more detailed view, we also consider p -values $< .01$ (**) and $p < .001$ (***). We found violations of assumptions essential to parametric statistical analysis, namely related to outliers, normality, and homoscedasticity. As a result, we adopted a robust statistical approach using the R package *WRS2* [48]. We used the *bwtrim* function of the package for robust two-way between-within (mixed) ANOVA and the *rmanova* function for robust one-way repeated measures ANOVA. Robust Cohen's d effect sizes were computed and can be categorized as 0.2 (small effect), 0.5 (medium effect), and 0.8 (large effect). Following Mair and Wilcox [48], we chose a trimmed mean of 20% because it has nearly equivalent power to the mean of a normally distributed sample and can lead to a substantially smaller standard error when outliers exist. Information about robust methods as an alternative to classic inferential analysis can be found in Field and Wilcox [20] and Wilcox [70].

We used Visual Studio Code with Jupyter notebooks for data analysis. Data pre-processing and aggregation were performed using Python (v3.11.5) and the pandas package (v2.0.3). Descriptive results and visualizations were generated in Python using matplotlib (v3.7.2), seaborn (v0.12.2), and numpy (v1.25.2). Robust statistical analyses were performed with R (v4.3.1) using *WRS2* (v1.1-4), *rstatix* (v0.7.2), *chemometrics* (v1.4.2), and *descTools* (v0.99.49). For visualizing users' handwriting in Figure 5, we used Python with *cairosvg* (v2.7.1) and *pillow* (v10.0.0). Descriptive results are shown in Table 2 and Figure 4.

4.1 System Usability Scale (SUS)

The robust mixed ANOVA models for the SUS subscales overall and usable revealed no significant interaction effects and no significant main effects for XR displays and surface alignments. The subscale learnable revealed no significant interaction effect but a significant main effect, with higher scores on physically aligned surfaces than in mid-air, $F(1, 41.64) = 6.06$, $p = .018$, robust Cohen's $d = -0.31$. For XR displays, contrary to H1.2, the results suggest no differences in usability between VR and VST AR. For surface alignments, partially confirming H2.2, the results indicate comparable usability but easier learnability on physically aligned than mid-air surfaces.

4.2 User Experience Questionnaire (UEQ)

The robust mixed ANOVA models for all UEQ subscales revealed no significant interaction effects. We found no significant main effects for attractiveness, perspicuity, efficiency, and dependability, for XR displays and surface alignments. The subscales novelty and stimulation revealed significant main effects for surface alignments. Novelty ratings were significantly higher in mid-air than on physically aligned surfaces, $F(1, 39.64) = 19.64$, $p < .001$, robust Cohen's $d = 0.55$. Stimulation ratings were also significantly higher in mid-air than on physically aligned surfaces, $F(1, 41.97) = 6.44$, $p = .015$, robust Cohen's $d = 0.25$. For XR displays, contrary to H1.3, the results suggest similar user experiences in VR and VST AR. For surface alignments, contrary to H2.3, the results show similar user experience, except for higher novelty and stimulation for handwriting on mid-air surfaces than physically aligned, indicating participants' less prior knowledge.

4.3 NASA Raw Task Load Index (RTLX)

The robust mixed ANOVA models revealed no significant interaction effects and no significant main effects for the RTLX subscales mental demand, temporal demand, performance, effort, frustration, and overall, for XR displays and surface alignments. The subscale physical demand revealed no significant interaction effect but a significant main effect.

Table 2: Descriptive results of subjective user ratings, text input performance, and stroke-level analysis ($n = 72$). Reported as Mean (SD). Square brackets after the name indicate the values' range or unit.

Subscale	VR		VST AR	
	Mid-Air	Phys. Aligned	Mid-Air	Phys. Aligned
SUS [0 - 100], [0 - 80], [0 - 20]				
Overall	78.07 (09.22)	81.88 (06.48)	75.83 (08.71)	82.70 (08.54)
Usable	59.89 (07.81)	62.92 (05.30)	59.90 (07.64)	63.54 (07.62)
Learnable	18.47 (01.67)	19.36 (01.11)	18.75 (01.57)	19.25 (01.16)
UEQ [-3 - 3]				
Attractiven.	01.02 (00.77)	01.04 (00.47)	01.20 (00.57)	00.99 (00.62)
Perspicuity	01.93 (00.82)	02.12 (00.35)	02.12 (00.40)	02.16 (00.55)
Efficiency	01.27 (00.74)	01.13 (00.53)	01.22 (00.50)	01.11 (00.62)
Dependab.	00.98 (00.75)	00.98 (00.46)	01.06 (00.57)	01.29 (00.56)
Novelty	01.86 (00.54)	01.21 (00.59)	01.27 (00.53)	00.73 (00.77)
Stimulation	01.35 (00.56)	00.92 (00.60)	01.33 (00.34)	01.15 (00.48)
RTLX [0 - 100]				
Ment. Dem.	29.36 (15.73)	28.33 (15.27)	28.14 (08.96)	22.92 (08.78)
Phys. Dem.	36.46 (16.29)	23.84 (13.26)	29.32 (12.77)	22.88 (09.18)
Temp. Dem.	43.02 (18.18)	42.81 (19.21)	28.83 (08.38)	31.22 (13.19)
Perform.	30.16 (13.17)	37.60 (12.33)	29.82 (06.74)	27.04 (12.50)
Effort	36.35 (15.98)	34.90 (14.99)	38.18 (12.88)	32.22 (13.77)
Frustration	25.64 (17.02)	33.16 (18.86)	20.69 (09.08)	28.70 (17.43)
Overall	33.60 (08.59)	33.95 (10.71)	30.21 (04.94)	30.09 (11.72)
SSQ [0 - 235.64]				
Before	09.27 (04.35)	05.42 (05.15)	08.82 (04.10)	09.13 (04.19)
After	07.12 (06.78)	06.61 (06.78)	08.59 (03.99)	11.22 (04.07)
WPM				
Comb.	16.11 (01.28)	15.72 (01.61)	16.72 (01.31)	16.69 (01.72)
Simple	17.40 (01.47)	16.95 (01.87)	18.57 (01.57)	18.46 (01.78)
Complex	15.38 (01.44)	14.48 (01.79)	15.50 (01.33)	14.92 (01.56)
MSD ER [in %]				
Comb.	01.06 (00.31)	01.19 (00.52)	01.50 (00.48)	01.20 (00.39)
Simple	00.29 (00.27)	00.49 (00.54)	00.61 (00.62)	00.65 (00.54)
Complex	01.81 (00.51)	01.66 (00.58)	01.99 (00.71)	01.50 (00.50)
Stroke Count				
Comb.	34.81 (02.69)	43.76 (03.11)	33.82 (02.62)	42.53 (02.72)
Simple	29.66 (03.12)	37.79 (02.75)	27.65 (02.78)	36.40 (02.20)
Complex	40.10 (02.34)	49.52 (03.08)	39.48 (03.02)	48.52 (03.28)
Stroke Length [in mm]				
Comb.	33.37 (06.62)	22.41 (04.61)	33.05 (06.71)	19.63 (02.62)
Simple	37.30 (06.84)	23.76 (05.62)	36.66 (08.96)	21.08 (03.03)
Complex	29.26 (05.95)	21.11 (03.92)	29.33 (04.14)	18.28 (02.62)
Stroke Height [in mm]				
Comb.	12.17 (01.51)	08.99 (01.40)	12.11 (01.47)	08.02 (00.84)
Simple	12.75 (01.62)	09.18 (01.54)	12.58 (01.72)	08.32 (00.98)
Complex	11.69 (01.50)	08.85 (01.26)	11.74 (01.24)	07.73 (00.96)

Physical demands were significantly higher in mid-air than on physically aligned surfaces, $F(1, 38.87) = 9.4$, $p = .004$, robust Cohen's $d = 0.36$. For surface alignments, confirming H2.4, the results imply that handwriting is less fatiguing on physically aligned surfaces.

4.4 Simulator Sickness Questionnaire (SSQ)

For well-being, the robust mixed ANOVA model for the SSQ revealed no significant interaction effects and no significant main effects for XR displays and surface alignments, indicating participants' well-being.

4.5 Words Per Minute (WPM)

For all sentences (simple and complex), the robust mixed ANOVA models for WPM revealed no significant interaction effect and no significant main effects for XR displays and surface alignments. Analyzing simple and complex sentences separately, we also found no significant interaction effect and no significant main effects. Comparing sentences by complexity (regardless of XR displays and surface alignments), the robust repeated measures ANOVA model revealed a significant main effect with higher WPM for simple than complex

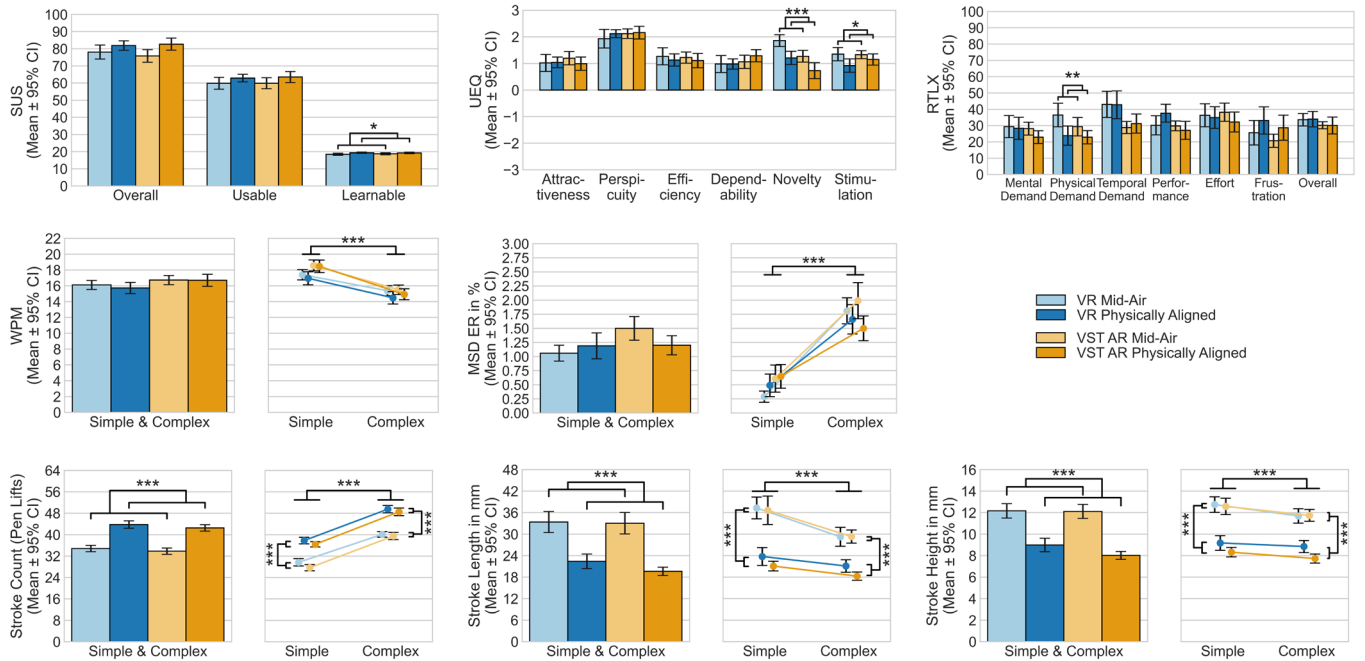


Fig. 4: The results of the System Usability Scale (SUS), User Experience Questionnaire (UEQ), and NASA Raw Task Load Index (RTLX). The SUS subscales range from 0 (worst imaginable) to 100 (best imaginable). The UEQ subscales range from -3 (bad) to 3 (excellent). The RTLX Performance subscale ranges from 0 (good) to 100 (poor). Other RTLX subscales range between 0 (low) and 100 (high). Additionally, the results of text input performance for Words Per Minute (WPM) and Minimum String Distance Error Rate (MSD ER), and stroke-level analysis for stroke count, stroke length, and stroke height. Asterisks indicates significance levels as $*p < .05$, $**p < .01$, and $***p < .001$.

sentences, $F(1, 43) = 89.51$, $p < .001$, robust Cohen's $d = 0.56$. For XR displays and surface alignments, contrary to H1.1 and H2.1, the results revealed similar text input speeds in VR and VST AR, and on physically aligned and virtual surfaces. For sentence complexities, confirming H3.1, the results highlight that handwriting text input is faster with simple sentences than with complex sentences.

4.6 Minimum String Distance Error Rate (MSD ER)

For all sentences, the robust mixed ANOVA models for MSD ER revealed no significant interaction effect and no significant main effects for XR displays and surface alignments. Analyzing sentences separately, we also found no significant interaction effect and no significant main effects for XR displays and surface alignments. Comparing sentences by complexity, the robust repeated measures ANOVA model revealed a significant main effect with higher MSD ER for complex than simple sentences, $F(1, 43) = 154.62$, $p < .001$, robust Cohen's $d = 0.75$. For sentence complexities, confirming H3.2, the results suggest that more errors remain in the transcribed text when writing complex sentences than simple sentences. This also indicates higher cognitive and motor demands.

4.7 Stroke Count

For all sentences, the robust mixed ANOVA models for stroke counts revealed no significant interaction effect but a significant main effect for surface alignments, with higher stroke counts for physically aligned than mid-air surfaces, $F(1, 41.93) = 144.37$, $p < .001$, robust Cohen's $d = -1.48$. Analyzing sentences separately, we found no significant interaction effect but significant main effects for surface alignments with higher stroke counts on physically aligned than mid-air surfaces, for both simple sentences $F(1, 41.06) = 133.54$, $p < .001$, robust Cohen's $d = -1.51$, and complex sentences, $F(1, 42.0) = 132.76$, $p < .001$, robust Cohen's $d = -1.42$. Comparing sentences by complexity, the robust repeated measures ANOVA model revealed a significant main effect with higher stroke counts for complex sentences than simple sentences, $F(1, 43.0) = 776.49$, $p < .001$, robust Cohen's $d = 0.91$. For surface alignments, confirming H2.5, the results imply that handwriting

is more connected on mid-air surfaces than on physically aligned surfaces. For sentence complexities, confirming H3.3, the results indicate that complex sentences require more strokes than simple sentences.

4.8 Stroke Length

For all sentences, the robust mixed ANOVA for stroke length in mm revealed no significant interaction effect but a significant main effect for surface alignments, with higher stroke lengths for mid-air than physically aligned surfaces, $F(1, 41.99) = 87.42$, $p < .001$, robust Cohen's $d = 1.13$. Analyzing sentences separately, we found no significant interaction effect but significant main effects for surface alignments with higher stroke lengths for mid-air than physically aligned surfaces, for both simple sentences $F(1, 39.54) = 74.0$, $p < .001$, robust Cohen's $d = 1.16$, and complex sentences, $F(1, 34.29) = 94.29$, $p < .001$, robust Cohen's $d = 1.08$. Comparing sentences by complexity, the robust repeated measures ANOVA model revealed a significant main effect with higher stroke lengths for simple sentences than complex sentences, $F(1, 43.0) = 68.56$, $p < .001$, robust Cohen's $d = 0.36$. For surface alignments, confirming H2.6, the results indicate that handwriting on mid-air surfaces is more connected than on physically aligned surfaces. For sentence complexities, confirming H3.4, the results also suggest that simple sentences enable longer strokes than complex sentences.

4.9 Stroke Height

For all sentences, the robust mixed ANOVA for stroke height in mm revealed no significant interaction effect but a significant main effect for surface alignments, with higher stroke heights for mid-air than physically aligned surfaces, $F(1, 40.99) = 140.4$, $p < .001$, robust Cohen's $d = 1.3$. Analyzing sentences separately, we found no significant interaction effect but significant main effects for surface alignments with higher stroke heights for mid-air than physically aligned surfaces, for both simple sentences $F(1, 41.54) = 146.54$, $p < .001$, robust Cohen's $d = 1.33$, and complex sentences, $F(1, 39.76) = 131.56$, $p < .001$, robust Cohen's $d = 1.3$. Comparing sentences by complexity, the robust repeated measures ANOVA model revealed a significant main effect with higher stroke heights for simple sentences than complex sentences,

	Simple sentence Kaufst du gerne am Sonntag ein? (Do you like to shop on Sunday?)		Complex sentence Die Geschwindigkeit liegt bei 230 km/h. (The velocity is 230 km/h.)	
	Physically Aligned	Mid-Air	Physically Aligned	Mid-Air
VR				
VR				
VST AR				
VST AR				

Fig. 5: Handwriting samples from eight participants in our user study for simple and complex sentences. Participants 1 - 4 participated in VR, and participants 5 - 8 used VST AR. They wrote each sentence on a physically aligned and mid-air surface. Colored dots indicate measure points acquired during the writing process. Fewer colors within a sentence represent lower stroke counts and higher stroke lengths, indicating more connected cursive handwriting.

$F(1, 43.0) = 29.56, p < .001$, robust Cohen's $d = 0.19$. For surface alignments, confirming H2.7, the results show larger handwriting on mid-air surfaces compared to physically aligned. For sentence complexities, confirming H3.5, the results also indicate larger handwriting with simple sentences than complex sentences.

5 DISCUSSION

In this work, we compared handwriting text input between VR and VST AR (XR Displays), facilitated by physically aligned and mid-air surfaces (Surface Alignments) when writing simple and complex sentences (Sentence Complexities). Our results emphasize that handwriting is a promising text input technique, offering high usability, similar to previous work [21]. Participants reported good to excellent user experience and perceived less physical demand and overall workload than in prior studies [21, 65]. Moreover, high perspicuity showed that it was easy to learn and familiarize with handwriting text input in XR. The averaged text input performance of 15.92 WPM and 1.13% MSD ER in VR, and 16.71 WPM with 1.35% MSD ER in VST AR, clearly surpassed earlier results in VR and OST AR [18, 19, 21, 23, 65], but also shows great potential for improvement compared to handwriting with pen and paper [11, 21] and styluses on laptop screens [39].

We summarize our key findings as follows: (1) XR displays did not affect text input performance and subjective user ratings. (2) Surface alignments yielded similar text input performance but significantly impacted physical demand, learnability, novelty, and stimulation, and led to adaptations in handwriting style. (3) Sentence complexities significantly influenced text input performance and handwriting style.

In the following, we present four design implications to support developers and researchers. Our considerations and suggestions aim to make handwriting text input more reliable and reusable in future XR use cases, which is particularly crucial for knowledge and office work.

5.1 Visual Incongruencies in VST AR

According to prior research [33, 41, 44, 68], we expected that potential visual incongruencies in VST AR would adversely impact text input speed (H1.1), usability (H1.2), and user experience (H1.3). However, this was not the case in our study. We attribute this to using the Meta Quest Pro, a recent XR device that produces subjectively less depth distortion and camera magnification than the Varjo XR3 employed in previous work [33]. This likely contributed to mitigating the negative effects of VST AR on handwriting text input. On the downside, the Meta Quest Pro compromises color fidelity by merging gray and color camera images, potentially affecting color perception. Based on our results, we assume that depth distortion and camera magnification have a greater impact on text input in VST AR than color fidelity. It also suggests that text input techniques perform differently across XR devices due to the varying visual fidelity of VST AR displays [33, 41], which can significantly impact text input performance and user experience.

Consequently, we advise evaluating text input techniques for each relevant display modality and, if possible, also for target XR devices, as findings from one setting may not necessarily apply to another.

5.2 Desirable but not Indispensable Physical Surfaces

Our findings indicate that it was easy for participants to learn and familiarize themselves with XR handwriting text input on physically aligned and mid-air surfaces, reflecting a successful transfer of handwriting skills to the virtual domain. However, notably higher learnability ratings for physically aligned surfaces (H2.2) suggest that handwriting on a physical table more closely resembles pen and paper or tablet stylus work. This assumption is supported by significantly higher novelty and stimulation ratings for mid-air surfaces (H2.3), indicating less prior experience, which we expect to diminish with more practice. While text input speed (H2.1) and error rate were similar on physically aligned (16.21 WPM, 1.20% MSD ER) and mid-air surfaces (16.42 WPM, 1.28% MSD ER), participants adapted their handwriting style to enlarged (H2.7) and more connected cursive (H2.5, H2.6) handwriting on mid-air surfaces. We attribute this to the lack of passive haptic feedback or resistance, which reduces control over precise movements. Although we simulated passive haptic feedback by controller vibrations for mid-air surfaces, the tactile sensation of depth and distance is not replicated, making it difficult to perceive virtual surface contact accurately. Nevertheless, participants in our study still produced legible handwriting, minimizing the need for user corrections, which in turn can reduce error rates and increase input speeds. Figure 5 visualizes handwriting legibility and adaptations of handwriting style. The lower physical demand for handwriting on physically aligned than mid-air surfaces (H2.4) suggests that both surfaces are suitable for short handwriting sessions but also indicates that physical surfaces are preferable for prolonged handwriting. This is in line with prior work, recommending physical surfaces for tasks requiring precise and fine-grained movements [3, 13, 31] or reducing arm and shoulder fatigue [15, 71, 72]. However, mid-air surfaces offer more flexibility, which can be particularly useful for non-stationary XR use cases [24, 33, 42]. Therefore, physical surfaces for handwriting are desirable for their familiarity and low physical demand but are not indispensable for all XR use cases. We propose a balanced approach, using surface alignment techniques [31, 35] to support physically aligned and mid-air surfaces in arbitrary orientations, accommodating user preferences and situations, as well as making handwriting text input reusable across XR devices and different input methods (e.g., controllers and fingers).

5.3 Representative Handwriting by Sentence Complexities

Sentence complexities significantly influenced text input performance, with higher input speeds (H3.1) and fewer errors (H3.2) for simple sentences (17.85 WPM, 0.51% MSD ER) than complex sentences (15.07 WPM, 1.74% MSD ER). We also found differences in hand-

writing style, which we explain by the transcribed phrases. Simple sentences usually consist of consecutive letters [16, 47, 55]. Complex sentences also include numbers, punctuation marks, and other symbols [16, 55, 66], that have to be represented individually, each requiring different shapes and separated strokes [28, 57]. This suggests that the cognitive and motor demands of writing complex sentences are notably more challenging, requiring thought and careful movement to reflect the content accurately. In combination with ensuring legibility for handwriting recognition, this likely impacts input speeds and error rates. As a result, complex sentences caused higher stroke counts (H3.3), lower stroke lengths (H3.4), and lower stroke heights (H3.5), leading to enlarged and more connected cursive handwriting of simple sentences, as shown in Figure 5. These findings highlight the participants' adaptability to different task demands and underscore the importance of including varying sentence complexities in future text input studies.

5.4 Improving Precision with XR Controller as Styluses

The gripping technique of controllers can highly influence performance and user comfort [5, 31, 56]. With our study, we showed that handwriting text input benefits from XR controllers in precision grip, mimicking styluses in pen-like postures. In contrast, previous research used controllers in power grip for VR handwriting [18, 21, 65], offering more stability, but limiting finger and wrist mobility, requiring more strength, and involving larger arm movements [5, 31, 61]. We assume that the precision grip supported precise and intricate movements [5, 31, 56], reflected in small stroke heights on physically aligned (8.50 mm) and mid-air surfaces (12.14 mm), while maintaining handwriting legibility, visualized in Figure 5. The size, shape, weight, and weight distribution of the XR controller can also influence handwriting text input. Uneven weight distribution makes input devices more uncomfortable, and heavier weight could limit fine-grained movements [8, 31, 56], leading to increased physical demand and potentially affecting handwriting over time. For example, the HTC Vive Pro controller, weighing 202 grams [31], was found to be unbalanced and impractical for finger movements [56]. In contrast, the Meta Quest Touch Pro Controller is lighter and more compact at 164 grams, contributing to a better weight distribution. Even more lightweight input devices like the 68 gram Logitech VR Ink stylus seem preferable but are often expensive or unavailable [31]. Therefore, we agree with previous work and recommend using XR controllers as styluses for tasks requiring precision and fine-grained movements, such as handwriting. Handwriting text input on physical surfaces also benefits from pressure-sensitive tips (e.g., Meta Quest Touch Pro controller and Logitech VR Ink), enabling fast and accurate input detection. For mid-air surfaces where physical contact is impossible, we suggest calculating orthogonal distances instead of forward raycasting [18, 65] to minimize the influence of wrist movements. Distance-based approaches can also be used for controllers without pressure-sensitive tips on physical surfaces [31, 34].

6 LIMITATIONS

Our research offers new insights into XR handwriting, yet there are limitations. We expect visual incongruencies like color and depth distortions can affect handwriting performance, but our research did not specifically investigate XR display characteristics. Further, our findings indicate that handwriting text input on mid-air surfaces is more physically demanding and less learnable than on physically aligned surfaces, despite similar text input speed, error rates, and other subjective ratings. This suggests the need for longer handwriting sessions with more sentences to understand the impact of physical demand on performance and user experience. While our study revealed higher text input speeds than in prior work, there is a noticeable gap to handwriting with pen and paper or tablet styluses, indicating evident potential for optimization. Another limitation is that we focused on horizontal surfaces. However, vertical and non-stationary surfaces such as whiteboards or tablets could yield different results. Given the potential impact of a controller's shape and weight on task performance, additional studies are essential to clarify how different input devices affect handwriting in XR. Lastly, we used virtual controller models in VR and VST AR, but the influence of externalized embodiment on handwriting is unclear.

7 CONCLUSION AND FUTURE WORK

Handwriting text input, with its high usability, natural intuitiveness, flexibility, and expressiveness, is a promising text input technique for many XR use cases, including office and knowledge work. We found that handwriting performance and handwriting style are consistent across XR displays, though visual incongruencies may impacted performance and user experience. Surface alignments did not affect performance but significantly influenced handwriting style, which we attribute to the lack of passive haptic feedback on mid-air surfaces. The higher novelty and stimulation ratings for mid-air surfaces likely decrease with more experience. Therefore, physical surfaces are desirable for lower physical demand and higher learnability but are not indispensable for XR handwriting. Sentence complexities significantly influenced text input performance and handwriting style, indicating higher cognitive and motor demands for complex sentences. Consequently, it is reasonable to include different sentence complexities in text input studies, reflecting real-world scenarios. XR controllers are promising styluses, especially when enhanced with pressure-sensitive tips for input detection on physical surfaces. We additionally provide a phrase set of simple and complex sentences with lower and uppercase letters, numbers, punctuation marks, and other symbols, that serves as a basis for text input studies and can be expanded and adapted.

Future work should explore how visual incongruencies in XR display characteristics, such as color and depth fidelity, affect handwriting. Furthermore, investigating the effects of prolonged handwriting on both physically aligned and mid-air surfaces is important. Vertical and non-stationary surfaces are currently underexplored and also require more studies. Although this study successfully used XR controllers in precision grip, the potential of dedicated XR styluses is evident.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF) in the projects ViLeArn More (Grant 16DHB2214) and HiAvA (Grant 16SV8781), and by the Bavarian State Ministry For Digital Affairs in the project XR Hub (Grant A5-3822-2-16).

REFERENCES

- [1] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth Perception in Augmented Reality: The Effects of Display, Shadow, and Position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801. IEEE, Christchurch, New Zealand, Mar. 2022. doi: 10.1109/VR51125.2022.00101 3
- [2] A. S. Arif and W. Stuerzlinger. Analysis of text entry performance metrics. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pp. 100–105. IEEE, Toronto, ON, Canada, Sept. 2009. doi: 10.1109/TIC-STH.2009.5444535 5
- [3] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. Fitzmaurice. Experimental Evaluation of Sketching on Surfaces in VR. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5643–5654. ACM, Denver, Colorado USA, May 2017. doi: 10.1145/3025453.3025474 3, 8
- [4] A. U. Batmaz, A. K. Mutasim, M. Malekmakan, E. Sadr, and W. Stuerzlinger. Touch the Wall: Comparison of Virtual and Augmented Reality with Conventional 2D Screen Eye-Hand Coordination Training Systems. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 184–193. IEEE, Atlanta, GA, USA, Mar. 2020. doi: 10.1109/VR46266.2020.00037 3
- [5] A. U. Batmaz, A. K. Mutasim, and W. Stuerzlinger. Precision vs. Power Grip: A Comparison of Pen Grip Styles for Selection in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 23–28. IEEE, Atlanta, GA, USA, Mar. 2020. doi: 10.1109/VRW50115.2020.00012 2, 3, 4, 9
- [6] V. Biener, E. Ofek, M. Pahud, P. O. Kristensson, and J. Grubert. Extended Reality for Knowledge Work in Everyday Environments. In A. Simeone, B. Weyers, S. Bialkova, and R. W. Lindeman, eds., *Everyday Virtual and Augmented Reality*, pp. 21–56. Springer International Publishing, Cham, 2023. doi: 10.1007/978-3-031-05804-2_2 1, 2
- [7] S. Borsci, S. Federici, and M. Lauriola. On the dimensionality of the System Usability Scale: a test of alternative measurement models. *Cognitive Processing*, 10(3):193–197, Aug. 2009. doi: 10.1007/s10339-009-0268-9 6

- [8] B. Bowers, A. Rukangu, C. Ball, A. Tuttle, A. Franzluebbens, and K. Johnsen. Comparing Virtual Constraints and a Physical Stylus for Planar Writing and Drawing in Virtual Reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, p. 6. IEEE, Lisbon, Portugal, 2021. doi: [10.1109/VRW52623.2021.00048](https://doi.org/10.1109/VRW52623.2021.00048) 3, 4, 9
- [9] B. Bowers, A. Rukangu, and K. Johnsen. Making it Simple: Expanding Access and Lowering Barriers to Novel Interaction Devices for Virtual and Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 1–6. IEEE, Atlanta, GA, USA, Mar. 2020. doi: [10.1109/VRW50115.2020.00289](https://doi.org/10.1109/VRW50115.2020.00289) 3
- [10] J. Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996. Publisher: London–. 6
- [11] C. M. L. Brown. Comparison of Typing and Handwriting in Two-Finger Typists. *Proceedings of the Human Factors Society Annual Meeting*, 32(5):381–385, Oct. 1988. doi: [10.1177/154193128803200553](https://doi.org/10.1177/154193128803200553) 2, 8
- [12] H. Bunke, M. Roth, and E. Schukat-Talamazzini. Off-line cursive handwriting recognition using hidden markov models. *Pattern Recognition*, 28(9):1399–1413, Sept. 1995. doi: [10.1016/0031-3203\(95\)00013-P](https://doi.org/10.1016/0031-3203(95)00013-P) 2
- [13] C. Chen, M. Yarmand, Z. Xu, V. Singh, Y. Zhang, and N. Weibel. Investigating Input Modality and Task Geometry on Precision-first 3D Drawing in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 384–393. IEEE, Singapore, Singapore, Oct. 2022. doi: [10.1109/ISMAR55827.2022.00054](https://doi.org/10.1109/ISMAR55827.2022.00054) 3, 8
- [14] Y.-T. Chen, C.-H. Hsu, C.-H. Chung, Y.-S. Wang, and S. V. Babu. iVRNote: Design, Creation and Evaluation of an Interactive Note-Taking Interface for Study and Reflection in VR Learning Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 172–180. IEEE, Osaka, Japan, Mar. 2019. doi: [10.1109/VR.2019.8798338](https://doi.org/10.1109/VR.2019.8798338) 2
- [15] Y. F. Cheng, T. Luong, A. R. Fender, P. Strelly, and C. Holz. Comfortable User Interfaces: Surfaces Reduce Input Error, Time, and Exertion for Tabletop and Mid-air User Interfaces. pp. 150–159. IEEE, Singapore, Singapore, Oct. 2022. doi: [10.1109/ISMAR55827.2022.00029](https://doi.org/10.1109/ISMAR55827.2022.00029) 1, 3, 8
- [16] K. Curran, D. Woods, and B. O. Riordan. Investigating text input methods for mobile phones. *Telematics and Informatics*, 23(1):1–21, Feb. 2006. doi: [10.1016/j.tele.2004.12.001](https://doi.org/10.1016/j.tele.2004.12.001) 3, 9
- [17] J. Dudley, H. Benko, D. Wigdor, and P. O. Kristensson. Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 289–300. IEEE, Beijing, China, Oct. 2019. doi: [10.1109/ISMAR.2019.00027](https://doi.org/10.1109/ISMAR.2019.00027) 2, 3
- [18] R. Elmgren. *Handwriting in VR as a Text Input Method*. PhD thesis, School of Computer Science and Communication, Stockholm, Sweden, July 2017. 1, 2, 3, 4, 8, 9
- [19] F. Fang, H. Zhang, L. Zhan, S. Guo, M. Zhang, J. Lin, Y. Qin, and H. Fu. Handwriting Velcro: Endowing AR Glasses with Personalized and Posture-adaptive Text Input Using Flexible Touch Sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(4):1–31, 2023. doi: [10.1145/3569461](https://doi.org/10.1145/3569461) 1, 2, 3, 8
- [20] A. P. Field and R. R. Wilcox. Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, 98:19–38, Nov. 2017. doi: [10.1016/j.brat.2017.05.013](https://doi.org/10.1016/j.brat.2017.05.013) 6
- [21] N. Fourrier, G. Moreau, M. Benaouicha, and J.-M. Normand. Handwriting for efficient text entry in industrial VR applications: influence of board orientation and sensory feedback on performance. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2023. doi: [10.1109/TVCG.2023.3320215](https://doi.org/10.1109/TVCG.2023.3320215) 1, 2, 3, 4, 8, 9
- [22] S. Gerth, A. Klassert, T. Dolk, M. Fliesser, M. H. Fischer, G. Nottbusch, and J. Festman. Is Handwriting Performance Affected by the Writing Surface? Comparing Preschoolers', Second Graders', and Adults' Writing Performance on a Tablet vs. Paper. *Frontiers in Psychology*, 7, Sept. 2016. doi: [10.3389/fpsyg.2016.01308](https://doi.org/10.3389/fpsyg.2016.01308) 2
- [23] G. González, J. P. Molina, A. S. García, D. Martínez, and P. González. Evaluation of Text Input Techniques in Immersive Virtual Environments. In J. A. Macías, A. Granollers Saltiveri, and P. M. Latorre, eds., *New Trends on Human-Computer Interaction: Research, Development, New Tools and Methods*, pp. 109–118. Springer, London, 2009. doi: [10.1007/978-1-84882-352-5_11](https://doi.org/10.1007/978-1-84882-352-5_11) 1, 2, 3, 8
- [24] J. Grubert, E. Ofek, M. Pahud, and P. O. Kristensson. The Office of the Future: Virtual, Portable, and Global. *IEEE Computer Graphics and Applications*, 38(6):125–133, Nov. 2018. doi: [10.1109/MCG.2018.2875609](https://doi.org/10.1109/MCG.2018.2875609) 1, 2, 8
- [25] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Effects of Hand Representations for Typing in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 151–158. IEEE, Reutlingen, Mar. 2018. doi: [10.1109/VR.2018.8446250](https://doi.org/10.1109/VR.2018.8446250) 2
- [26] S. G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, Oct. 2006. doi: [10.1177/154193120605000909](https://doi.org/10.1177/154193120605000909) 6
- [27] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, vol. 52, pp. 139–183. Elsevier, 1988. doi: [10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9) 6
- [28] M. Karthi, R. Priscilla, and K. S. Jafer. A Novel Content Detection Approach for Handwritten English letters. *Procedia Computer Science*, 172:1016–1025, 2020. doi: [10.1016/j.procs.2020.05.149](https://doi.org/10.1016/j.procs.2020.05.149) 2, 9
- [29] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, July 1993. doi: [10.1207/s15327108ijap0303_3_4_5](https://doi.org/10.1207/s15327108ijap0303_3_4_5)
- [30] F. Kern, T. Keser, F. Niebling, and M. E. Latoschik. Using Hand Tracking and Voice Commands to Physically Align Virtual Surfaces in AR for Handwriting and Sketching with HoloLens 2. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, VRST '21*. Association for Computing Machinery, New York, NY, USA, 2021. event-place: Osaka, Japan. doi: [10.1145/3489849.3489940](https://doi.org/10.1145/3489849.3489940) 3
- [31] F. Kern, P. Kullmann, E. Ganal, K. Korwisi, R. Stingl, F. Niebling, and M. E. Latoschik. Off-The-Shelf Stylus: Using XR Devices for Handwriting and Sketching on Physically Aligned Virtual Surfaces. *Frontiers in Virtual Reality*, 2:684498, June 2021. doi: [10.3389/frvir.2021.684498](https://doi.org/10.3389/frvir.2021.684498) 1, 2, 3, 4, 8, 9
- [32] F. Kern and M. E. Latoschik. Reality Stack I/O: A Versatile and Modular Framework for Simplifying and Unifying XR Applications and Research. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 74–76. IEEE, Sydney, Australia, Oct. 2023. doi: [10.1109/ISMAR-Adjunct60411.2023.00023](https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00023) 4
- [33] F. Kern, F. Niebling, and M. E. Latoschik. Text Input for Non-Stationary XR Workspaces: Investigating Tap and Word-Gesture Keyboards in Virtual and Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2658–2669, May 2023. doi: [10.1109/TVCG.2023.3247098](https://doi.org/10.1109/TVCG.2023.3247098) 1, 2, 3, 5, 8
- [34] F. Kern, M. Popp, P. Kullmann, E. Ganal, and M. E. Latoschik. 3D Printing an Accessory Dock for XR Controllers and Its Exemplary Use as XR Stylus. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, VRST '21*. Association for Computing Machinery, New York, NY, USA, 2021. event-place: Osaka, Japan. doi: [10.1145/3489849.3489949](https://doi.org/10.1145/3489849.3489949) 9
- [35] F. Kern, J. Tschanter, and M. E. Latoschik. Virtual-to-Physical Surface Alignment and Refinement Techniques for Handwriting, Sketching, and Selection in XR. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 502–506. IEEE, Shanghai, China, Mar. 2023. doi: [10.1109/VRW58643.2023.00109](https://doi.org/10.1109/VRW58643.2023.00109) 4, 5, 8
- [36] D. Keysers, T. Deselaers, H. A. Rowley, L.-L. Wang, and V. Carbune. Multi-Language Online Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1180–1194, June 2017. doi: [10.1109/TPAMI.2016.2572693](https://doi.org/10.1109/TPAMI.2016.2572693) 2
- [37] M. Kiefer, S. Schuler, C. Mayer, N. M. Trumpp, K. Hille, and S. Sachse. Handwriting or Typewriting? The Influence of Penor Keyboard-Based Writing Training on Reading and Writing Performance in Preschool Children. *Advances in Cognitive Psychology*, 11:136–146, Dec. 2015. doi: [10.5709/acp-0178-7](https://doi.org/10.5709/acp-0178-7) 2
- [38] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, pp. 1–9. Association for Computing Machinery, New York, NY, USA, Apr. 2018. doi: [10.1145/3173574.3173919](https://doi.org/10.1145/3173574.3173919) 2, 3
- [39] P. O. Kristensson and L. C. Denby. Text entry performance of state of the art unconstrained handwriting recognition: a longitudinal user study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 567–570. ACM, Boston MA USA, Apr. 2009. doi: [10.1145/1518701.1518788](https://doi.org/10.1145/1518701.1518788) 2, 8
- [40] P. O. Kristensson and K. Vertanen. Performance comparisons of phrase sets and presentation styles for text entry evaluations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces - IUI '12*, p. 29. ACM Press, Lisbon, Portugal, 2012. doi: [10.1145/2166966](https://doi.org/10.1145/2166966)

- 2166972 3
- [41] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12. IEEE, Seoul, Korea (South), Oct. 2010. doi: [10.1109/ISMAR.2010.5643530](https://doi.org/10.1109/ISMAR.2010.5643530) 2, 3, 8
- [42] W. S. Lages and D. A. Bowman. Walking with adaptive augmented reality workspaces: design and usage patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 356–366. ACM, Marina del Rey California, Mar. 2019. doi: [10.1145/3301275.3302278](https://doi.org/10.1145/3301275.3302278) 1, 2, 8
- [43] M. E. Latoschik, F. Kern, J.-P. Stauffert, A. Bartl, M. Botsch, and J.-L. Lugin. Not Alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2134–2144, May 2019. doi: [10.1109/TVCG.2019.2899250](https://doi.org/10.1109/TVCG.2019.2899250) 1, 2
- [44] M. E. Latoschik and C. Wienrich. Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. *Frontiers in Virtual Reality*, 3:694433, June 2022. doi: [10.3389/frvir.2022.694433](https://doi.org/10.3389/frvir.2022.694433) 2, 3, 8
- [45] B. Laugwitz, T. Held, and M. Schrepp. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, vol. 5298, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. 6
- [46] J. R. Lewis. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, Jan. 1995. doi: [10.1080/10447319509526110](https://doi.org/10.1080/10447319509526110) 6
- [47] I. S. MacKenzie and R. W. Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI '03 extended abstracts on Human factors in computing systems - CHI '03*, p. 754. ACM Press, Ft. Lauderdale, Florida, USA, 2003. doi: [10.1145/765891.765971](https://doi.org/10.1145/765891.765971) 3, 5, 9
- [48] P. Mair and R. Wilcox. Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52(2):464–488, Apr. 2020. doi: [10.3758/s13428-019-01246-w](https://doi.org/10.3758/s13428-019-01246-w) 6
- [49] M. McGill, S. Brewster, D. P. De Sa Medeiros, S. Bovet, M. Gutierrez, and A. Kehoe. Creating and Augmenting Keyboards for Extended Reality with the K keyboard A augmentation Toolkit. *ACM Transactions on Computer-Human Interaction*, 29(2):1–39, Apr. 2022. doi: [10.1145/3490495](https://doi.org/10.1145/3490495) 2
- [50] M. McGill, A. Kehoe, E. Freeman, and S. Brewster. Expanding the Bounds of Seated Virtual Workspaces. *ACM Transactions on Computer-Human Interaction*, 27(3):1–40, June 2020. doi: [10.1145/3380959](https://doi.org/10.1145/3380959) 1, 2
- [51] K. Morehead, J. Dunlosky, and K. A. Rawson. How Much Mightier Is the Pen than the Keyboard for Note-Taking? A Replication and Extension of Mueller and Oppenheimer (2014). *Educational Psychology Review*, 31(3):753–780, Sept. 2019. doi: [10.1007/s10648-019-09468-2](https://doi.org/10.1007/s10648-019-09468-2) 1, 2
- [52] K. Morehead, J. Dunlosky, K. A. Rawson, R. Blasiman, and R. B. Hollis. Note-taking habits of 21st Century college students: implications for student learning, memory, and achievement. *Memory*, 27(6):807–819, July 2019. doi: [10.1080/09658211.2019.1569694](https://doi.org/10.1080/09658211.2019.1569694) 2
- [53] P. A. Mueller and D. M. Oppenheimer. The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking. *Psychological Science*, 25(6):1159–1168, June 2014. doi: [10.1177/0956797614524581](https://doi.org/10.1177/0956797614524581) 1
- [54] E. Ose Askvik, F. R. R. van der Weel, and A. L. H. van der Meer. The Importance of Cursive Handwriting Over Typewriting for Learning in the Classroom: A High-Density EEG Study of 12-Year-Old Children and Young Adults. *Frontiers in Psychology*, 11:1810, July 2020. doi: [10.3389/fpsyg.2020.01810](https://doi.org/10.3389/fpsyg.2020.01810) 2
- [55] D.-M. Pham and W. Stuerzlinger. HawKEY: Efficient and Versatile Text Entry for Virtual Reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11. ACM, Parramatta, NSW, Australia, Nov. 2019. doi: [10.1145/3359996.3364265](https://doi.org/10.1145/3359996.3364265) 2, 3, 9
- [56] D.-M. Pham and W. Stuerzlinger. Is the Pen Mightier than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11. ACM, Parramatta, NSW, Australia, Nov. 2019. doi: [10.1145/3359996.3364264](https://doi.org/10.1145/3359996.3364264) 2, 3, 4, 9
- [57] R. Plamondon and S. Srihari. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):63–84, Jan. 2000. doi: [10.1109/34.824821](https://doi.org/10.1109/34.824821) 2, 9
- [58] I. Popyrev, N. Tomokazu, and S. Weghorst. Virtual Notepad: handwriting in immersive VR. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*, pp. 126–132. IEEE Comput. Soc, Atlanta, GA, USA, 1998. doi: [10.1109/VRAIS.1998.658467](https://doi.org/10.1109/VRAIS.1998.658467) 1, 2, 3
- [59] D. Schmalstieg and T. Höllerer. *Augmented reality: principles and practice*. Addison-Wesley usability and HCI series. Addison-Wesley, Boston, 2016. OCLC: ocn802326915. 3
- [60] D. M. Schrepp. User Experience Questionnaire Handbook, 2019. 6
- [61] A.-S. Selin. Pencil grip : a descriptive model and four empirical studies. Åbo Akademi University, Turku, Finland, 2003. 3, 9
- [62] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2:647997, Mar. 2021. doi: [10.3389/frvir.2021.647997](https://doi.org/10.3389/frvir.2021.647997) 2
- [63] A. L. H. van der Meer and F. R. R. van der Weel. Only Three Fingers Write, but the Whole Brain Works!: A High-Density EEG Study Showing Advantages of Drawing Over Typing for Learning. *Frontiers in Psychology*, 8:706, May 2017. doi: [10.3389/fpsyg.2017.00706](https://doi.org/10.3389/fpsyg.2017.00706) 2
- [64] P. C. Vashist, A. Pandey, and A. Tripathi. A Comparative Study of Handwriting Recognition Techniques. In *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 456–461. IEEE, Dubai, United Arab Emirates, Jan. 2020. doi: [10.1109/ICCAKM46823.2020.9051464](https://doi.org/10.1109/ICCAKM46823.2020.9051464) 2
- [65] R. Venkatakrishnan, R. Venkatakrishnan, C.-H. Chung, Y.-S. Wang, and S. Babu. Investigating a Combination of Input Modalities, Canvas Geometries, and Inking Triggers on On-Air Handwriting in Virtual Reality. *ACM Transactions on Applied Perception*, 19(4):1–19, Oct. 2022. doi: [10.1145/3560817](https://doi.org/10.1145/3560817) 1, 2, 3, 4, 8, 9
- [66] K. Vertanen and P. O. Kristensson. A versatile dataset for text entry evaluations based on genuine mobile emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, p. 295. ACM Press, Stockholm, Sweden, 2011. doi: [10.1145/2037373.2037418](https://doi.org/10.1145/2037373.2037418) 3, 5, 9
- [67] R. Viciano-Abad, A. R. Lecuona, and M. Poyade. The Influence of Passive Haptic Feedback and Difference Interaction Metaphors on Presence and Task Performance. *Presence: Teleoperators and Virtual Environments*, 19(3):197–212, June 2010. doi: [10.1162/pres.19.3.197](https://doi.org/10.1162/pres.19.3.197) 1, 3
- [68] F. Westermeier, L. Brübach, C. Wienrich, and M. E. Latoschik. A Virtualized Augmented Reality Simulation for Exploring Perceptual Incongruencies. In *29th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2. ACM, Christchurch New Zealand, Oct. 2023. doi: [10.1145/3611659.3617227](https://doi.org/10.1145/3611659.3617227) 3, 8
- [69] C. Wienrich, P. Komma, S. Vogt, and M. E. Latoschik. Spatial Presence in Mixed Realities—Considerations About the Concept, Measures, Design, and Experiments. *Frontiers in Virtual Reality*, 2:694315, Oct. 2021. doi: [10.3389/frvir.2021.694315](https://doi.org/10.3389/frvir.2021.694315) 2
- [70] R. R. Wilcox. *Introduction to robust estimation and hypothesis testing*. Elsevier, Waltham, MA, 4th edition ed., 2017. 6
- [71] D. Zielasko. *DeskVR: Seamless Integration of Virtual Reality into Desk-Based Data Analysis Workflows*. PhD Thesis, RWTH Aachen University, 2020. 1, 2, 3, 8
- [72] D. Zielasko, M. Kruger, B. Weyers, and T. W. Kuhlen. Passive Haptic Menus for Desk-Based and HMD-Projected Virtual Reality. In *2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–6. IEEE, Osaka, Japan, Mar. 2019. doi: [10.1109/WEVR.2019.8809589](https://doi.org/10.1109/WEVR.2019.8809589) 1, 2, 3, 8