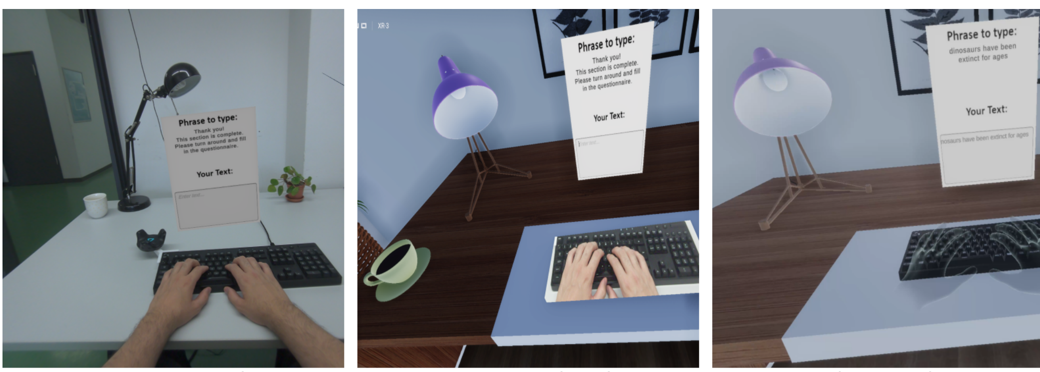
Evaluating Typing Performance in Different Mixed Reality Manifestations Using Physiological Features

FRANCESCO CHIOSSI, LMU Munich, Germany
YASSMINE EL KHAOUDI, LMU Munich, Germany
CHANGKUN OU, LMU Munich, Germany
LUDWIG SIDENMARK, University of Toronto, Canada
ABDELRAHMAN ZAKY, University of Konstanz, Germany
TIARE FEUCHTNER, University of Konstanz, Germany
SVEN MAYER, LMU Munich, Germany



Augmented Reality | Augmented Virtuality | Virtual Reality

Fig. 1. In a user study we investigate physiological correlates for engagement and workload states, performance during a typing task in three mixed reality environments, thereby varying the degree of immersion.

Mixed reality enables users to immerse themselves in high-workload interaction spaces like office work scenarios. We envision physiologically adaptive systems that can move users into different mixed reality manifestations, to improve their focus on the primary task. However, it is unclear which manifestation is most conducive for high productivity and engagement. In this work, we evaluate whether physiological indicators for engagement can be discriminated for different manifestations. For this, we engaged participants in a typing task in three different mixed reality manifestations (augmented reality, augmented virtuality, virtual reality) and monitored physiological correlates (EEG, ECG, and eye tracking) of users' engagement and workload. We found that users achieved best typing performances in augmented reality and augmented virtuality. At

Authors' Contact Information: Francesco Chiossi, LMU Munich, Munich, Germany, francesco.chiossi@um.ifi.lmu.de; Yassmine El Khaoudi, LMU Munich, Munich, Germany, y.khaoudi@campus.lmu.de; Changkun Ou, LMU Munich, Munich, Germany, research@changkun.de; Ludwig Sidenmark, University of Toronto, Toronto, Canada, lsidenmark@dgp.toronto.edu; Abdelrahman Zaky, University of Konstanz, Konstanz, Germany, abdelrahman.zaky@uni-konstanz.de; Tiare Feuchtner, University of Konstanz, Konstanz, Germany, tiare.feuchtner@uni-konstanz.de; Sven Mayer, LMU Munich, Munich, Germany, info@sven-mayer.com.

ACM 2573-0142/2024/12-ART542
https://doi.org/10.1145/3698142

the same time, physiological engagement peaked in augmented virtuality, while workload decreased. We conclude that augmented virtuality strikes a good balance between the different manifestations, as it facilitates displaying the physical keyboard for improved typing performance and, at the same time, allows one to block out the real world, removing many real-world distractors.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Mixed / augmented reality**; **Virtual reality**; *User studies*; *Text input*.

Additional Key Words and Phrases: Mixed Reality, Virtual Reality, Augmented Reality, Augmented Virtuality, Engagement, Electroencephalography, Eye Tracking, Physiological Computing

ACM Reference Format:
Francesco Chiossi, Yassmine El Khaoudi, Changkun Ou, Ludwig Sidenmark, Abdelrahman Zaky, Tiare Feuchtner, and Sven Mayer. 2024. Evaluating Typing Performance in Different Mixed Reality Manifestations Using Physiological Features. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 542 (December 2024), 30 pages. https://doi.org/10.1145/3698142

1 Introduction

Mixed reality (MR) technologies may revolutionize human-computer interaction (HCI) by seamlessly integrating digital and physical environments, offering a virtually infinite display space, situated information visualization, and affordances of physical reality. As MR systems evolve, their integration into our daily and productive lives becomes increasingly feasible, ranging from industry 4.0 to the medical domain to office spaces. A central challenge for productive work in MR that remains is text-based input, which can be efficiently achieved through traditional keyboards but is tedious to support for mid-air virtual keyboards (e.g., Dudley et al. [42], Speicher et al. [140]), as are commonly experienced in Virtual Reality (VR). Although alternative text entry methods have been proposed for MR (e.g., tapping surfaces [142], mid-air multi-finger gestures [141], or even voice input [116]), wider adoption of novel modalities may take a long time [102]. At the same time, immersing users in virtual worlds enhances their focus on the primary tasks without external distractions. With these opposing concepts, it is unclear how to incorporate real-world objects, such as the keyboard, while keeping the focus (i.e., immersion) on the primary task high.

Recent studies have focused on investigating various factors impacted by MR, such as spatial awareness [119, 127], object manipulations [49], and visual perception [153]. As MR technologies advance and become more accessible, typing and productivity tasks in these environments are expected to become increasingly common. However, the integration of digital and physical information faces challenges related to keyboard display and layout, far from fully supporting user experience, performance, and productivity in MR [56]. McGill et al. [102] conducted seminal work on the effects of visual feedback and keyboard size, showing that Augmented Virtuality (AV) allows for improved typing performance. More recently, Lu et al. [90] investigated typing performance with mid-air keyboards in Augmented Reality (AR), confirming this MR manifestation as a feasible alternative. However, it is yet unclear how the degree of virtuality of the environment and the resulting variations in visual detail, lighting, and possibility of distractions affect task performance and user experience. Traditionally, typing is understood solely from a performance perspective as one instance of high-workload tasks [78]. However, the users' cognitive overload and task-disengagement that may occur despite good performance, have not yet been considered. Moreover, for most daily tasks, performance cannot be measured easily. Thus, in line with prior work (e.g., [18, 31, 32]), an alternative view is to look at the physiological correlates of executing tasks in MR environments and investigate how these can be used to determine the user's cognitive state. A better understanding of physiological correlates in MR will contribute to the design of adaptive systems in MR [73, 87] that can dynamically adjust to the user's cognitive state. For example, through real-time monitoring of the user's engagement and attention levels, the system can respond appropriately, adapting the user's perceived environment by adding, removing, or replacing visual content, e.g., to reduce cognitive load or provide timely assistance.

We envision a new class of transitional interfaces [6, 101] driven by implicit physiological responses rather than explicit user commands. Our investigation into the physiological correlates of workload and engagement across different MR environments provides insights for developing adaptive interfaces that seamlessly transition users through the reality-virtuality continuum based on their cognitive state and support task performance. This study defines engagement as the degree of attention and cognitive resource allocation exhibited by participants when completing a task [111]. High engagement is beneficial, leading to better performance and a more immersive experience [111]. Engagement can be characterized by increased attention, positive affect, and sustained cognitive effort [100]. Conversely, cognitive load refers to the mental effort required to perform the task. Here, Paas et al. [112] found that high cognitive load can negatively impact performance, leading to increased errors and fatigue. While engagement and cognitive load are related, they have distinct implications for user experience. Ideally, high engagement should coincide with low cognitive load for optimal performance. However, high engagement can sometimes occur with high cognitive load, which, if sustained, may lead to cognitive fatigue and stress [20].

We investigate a typing task on a physical keyboard at three different manifestations of the reality-virtuality continuum [107, 137], namely: *AR*, *AV*, *VR*. For this, we conducted a within-subjects lab study ($N = 18$), asking participants to type phrases for 6 minutes per condition. During the typing tasks, we recorded three physiological signals: electrocardiogram (ECG), electroencephalography (EEG), and eye tracking (ET), thereby analyzing physiological responses associated with engagement, attention allocation, and workload levels. We are chiefly interested in exploring distinctions of physiological states between different MR environments and identifying physiological patterns for individual cognitive states during the interaction.

Our findings indicate that AV balanced high engagement and moderate cognitive load, leading to typing performance comparable to AR. AR resulted in moderate engagement but a higher cognitive load due to real-world distractions. VR led to a high cognitive load with somewhat lower engagement, resulting in higher error rates and longer task completion times. Our results suggest that AV may strike the best balance among the different MR manifestations, shielding users from external distractions while displaying task-relevant real-world content to support typing performance and engagement.

Contribution Statement. The contribution of this paper is fourfold. (1) We present the first comprehensive study that integrates physiological measures (EEG, ECG, and eye tracking) to evaluate typing performance, workload and engagement in MR environments, thereby establishing physiological metrics of such constructs in an applied context. (2) We propose and evaluate design modifications to the MR continuum that can improve typing performance. Our findings indicate that AV environments provide the best balance of high engagement and moderate cognitive load, leading to superior typing performance. These insights are critical for optimizing MR environments for various productivity tasks. (3) Our findings build upon and extend the work of McGill et al. [102] by incorporating physiological measures to explore typing performance in MR environments. We provide detailed insights into the physiological and design factors influencing typing, thereby supporting the design and evaluation of both static and adaptive MR interfaces. Our study lays the groundwork for future adaptive transitional interfaces that employ physiological features as inputs for interaction. (4) we make our MR implementation and the collected dataset, which includes co-registered EEG, ECG, and eye-tracking features, available as open source to support future research. The novelty of our work lies in integrating comprehensive physiological measures to gain a deeper understanding of how different MR environments impact engagement, workload, and attention while executing a production task. We identify design modifications grounded on physiological measures and discuss applications for adaptive engagement and workload-aware MR environments.

2 Related Work

This section provides a background on knowledge work in MR and reviews the physiological correlates of task engagement and workload that motivated our choice of dependent variables.

2.1 Supporting Work in Mixed Reality

While head-mounted displays (HMDs) for MR have not yet become daily accessories comparable to the smartphone, immersive technologies are increasingly adopted in professional contexts, where the big price tag, specific infrastructure requirements, and lack of available content are less of an obstacle. For example, in the manufacturing industry, MR can provide novel customer experiences and support training [19, 150]. These are frequently immersive VR scenarios, where user experience is often enriched by including a user representation [136] that impacts the sense of presence, body ownership, spatial perception. AR has gained popularity for on-the-job support, like remote assistance [118, 127], situated information visualization for assembly instructions [143], and human-robot collaboration [2, 91, 97, 114], as it permits simultaneous access to virtual content and the real world. MR technologies are also showing promise in the medical domain for training and planning of surgeries, as well as live data exploration or remote assistance in the operating room [50, 51, 130]. Further, there is a long research tradition of MR technologies in the context of data analysis and immersive analytics [83, 106], and education [76, 134]. Lastly, while limitations of current technologies still hinder the adoption of VR-based work environments for prolonged use [16], MR harbors the potential for effectively transforming our typical office spaces, e.g., by expanding the available display space [68] and facilitating remote collaboration [47].

Across all of these domains, the input of textual information is often required and would traditionally be supported by (analog or digital) note-taking, such as text input on a computer. While this task is often neglected and rarely effectively supported by MR technologies, a variety of approaches have been proposed, ranging from mid-air virtual keyboards (e.g., Dudley et al. [42], Speicher et al. [140]), to multi-finger gestures [141], or finger tapping [142]. The latter was the most efficient of these techniques [142], with researchers reporting 25 WPM for proficient typists (and 19 WPM for non-experts). This is followed closely by freehand typing interactions using mid-air gestures [141] achieving 22 WPM, while with controllers, users merely achieved 15.4 WPM [42]. Lu et al. [90] investigated typing performance using eye gaze. With an average of 13.76 WPM, they demonstrate AR as a feasible alternative setting. With recent advances in speech recognition, voice input presents a further alternative for a text input. Investigating text entry in a CAVE virtual environment using speech recognition with correction [63], Pick et al. [116] report an average of 23.6 WPM, while Hoste and Signer [66] using SpeeG2 interface achieved 21 WPM. This modality is limited by the requirement for low noise level environments and potential issues with privacy concerns [140], and is arguably not suited for common note-taking scenarios (e.g., writing in a shared office space, or making a meeting protocol).

With regards to task performance in typing on a physical keyboard, McGill et al. [102] conducted a study comparing typing in 4 conditions: (1) physical reality, (2) VR, (3) AV with partial blending, and (4) AV with full blending (i.e., video feed of reality). Compared to typing with no keyboard view (VR: 23.6 WPM), the authors report improved typing rates in both AV conditions (partial blending: 38.5 WPM, full blending: 36.6 WPM). Further, in AV the error rate was significantly reduced, approaching the baseline rate of the reality condition. While this illustrates the benefit of using physical keyboards (or indeed any keyboard visualization) for typing in an MR environment, the baseline typing rate (Reality: 58.9 WPM) could not be achieved even when seeing the entire physical world in the full blending AV condition. This may hint at some impact of the video-see-through display technique. Supporting typing on a physical keyboard while presenting a virtual keyboard replica in VR, Knierim et al. [78] have shown that experienced typists can achieve a similar typing performance (above 60 WPM) as in physical (non-augmented) reality. Even non-expert typists, who in the real world setup achieved approx. 45 WPM, were only slightly slower in VR (39.80 WPM). This underlines the need to visualize the keyboard faithfully. However, investigating the need for visualizing users' hands, the authors report that proficient typists could also type "blindly" (i.e., not seeing any indication of where their hands were), while for inexperienced typists, the presentation of virtual hands animated with their own finger movements led to significant improvements in typing performance. Further, the workload was reported to be higher in VR for all typists, and particularly so when no virtual hands were shown. Next, when investigating keyboard display, Pham and Stuerzlinger [115] introduced a portable keyboard for VR that allows users to achieve high-speed text entry while standing, similar to desktop performance. The study evaluated different visualizations of the keyboard and found that the video-based visualization, which shows the keyboard only when the user looks down, offers the best performance, with text entry rates comparable to traditional setups (77.7 WPM for experienced typists). Their findings reinforce the necessity of faithful keyboard visualization to achieve optimal typing performance in VR environments. The presented research suggests the superiority of supporting 10-finger typing on a physical keyboard in MR environments, compared to mid-air keyboards that frequently support just a single touch point per hand (e.g., individual finger input controller [42, 140]) and alternative typing mappings without keyboards (e.g., Sridhar et al. [141], Streli et al. [142]).

2.2 Physiological Correlates of Engagement and Workload in Mixed Reality

Pohl and Murray-Smith [120] proposed a focused-casual continuum defining interaction techniques based on the users' flexibility to adjust their attention and engagement during an interaction, depending on their present circumstances. In other words, it allows users to adapt their level of engagement [40, 120]. Their approach to defining engagement, based on a control-theoretic perspective, suggests that users tend to sample feedback more frequently and accurately as they become more engaged. We envision that transitions in MR are linked to user engagement, as they impact how users interact, perform, and experience the MR space [102]. Specifically, engagement states have been identified using different physiological features extracted from ECG, gaze behavior, and EEG data, allowing for implicit interaction [135]. While the control-theoretic perspective emphasizes the adaptability of user engagement during interactions, the physiological computing perspective complements it by providing objective physiological measures of users' engagement states. Researchers have explored various physiological correlates to understand and measure user engagement. This section reviews the physiological correlates of users' engagement and how they have been investigated in MR environments, i.e., gaze behavior, peripheral physiology (e.g., ECG), and electrophysiology (e.g., EEG).

2.2.1 Electrocardiogram. ECG is a physiological measure of cardiac activity, recording changes in electrical potentials generated by cardiac muscles [54] and indexing sympathetic activity at the level of the autonomic nervous system (ANS). In the context of HCI and human factors, ECG has been used to monitor physiological stress, mostly sensible to affective and physical stressors [4], and in applied contexts such as office work [25] or automated scenarios [146]. However, popular ECG measures in the time domain, such as Heart Rate (HR) or Heart Rate Variability (HRV), did not always discriminate workload states consistently [110, 149], different interpretations have arisen, considering that, for example, systolic blood pressure and heart rate have been shown to response to motivational incentives [45], mental effort [44] and overall physiological activation in stressful situations [77]. Interestingly, ECG has found applications in automated adaptive systems [122], such as in Byrne and Parasuraman [21], where automation modes were activated/deactivated based on stress levels inferred from HR.

2.2.2 Electroencephalogram. EEG measures have been employed to investigate the relationship between brain activation states and the level of mental effort invested in a task [40]. Several experiments have shown that increased mental effort, particularly in response to working memory load (i.e., the number of items to be retained in memory), is associated with the synchronization of theta activity (4–7 Hz) in central frontal sites and the suppression of alpha activity in occipital areas in computer-based work [52], multitasking environments [123], and VR interaction [29]. An alternative hypothesis linked alpha to fluctuations in attentional load [48]. This rhythm serves as a sensory gating mechanism, amplifying the processing of relevant sensory information while enhancing task-relevant information [48, 71]. The existing literature has explored the potential of posterior alpha as a reliable index for discerning internal and external attention dynamics, with external attention being characterized by a decrease in alpha power and internal attention predominantly marked by an increase in alpha power [13, 35]. Interestingly, alpha classification successfully discriminated between the two internal and external states in an AR paradigm, and it has been employed to perform a person-dependent classification for external and internal allocation in an AR paradigm [147] and applied to AR smart home control [125] and VR settings [89].

Following a reverse direction, theta increased power has been linked with associated with higher task engagement [3, 81], while lower theta power indicates low task engagement [133]. Moreover, theta activity potentially mediates external and internal attention, highlighting their competitive interaction [35, 95], also in VR environments [94, 131]. Given this dual relationship, researchers developed and validated metrics of workload based on alpha and theta, such as their respective ratio for cognitive demands [128].

Moreover, EEG measures have been applied to infer engagement without interruptions. In these studies, predictive models leveraging EEG and motion-tracking data achieved a remarkable 67% accuracy in classifying engagement across different subjects and a promising 85% accuracy within the same subject [18]. With a similar objective, De Massari et al. [36] classified different cognitive states and their underlying task, i.e., spatial navigation, calculation, and reading in an MR environment. Their findings revealed an increase in alpha band activity in the bilateral parietal areas, consistent with previous studies that reported alpha decrease in these brain regions linked to mental workload, task engagement, or attention [69, 79, 121]. These findings highlight the potential of using EEG data to quantify engagement levels and workload and inform the design of adaptive systems that can enhance user experiences in various MR applications.

2.2.3 Eye Tracking. Studies have demonstrated a correlation between EEG theta power during cognitive engagement and changes in pupil diameter [34, 39]. The pupil diameter is considered a valuable metric for monitoring mental effort [26, 43, 145] and task engagement [53]. Two distinct aspects of pupil diameter dynamics can be distinguished: phasic pupillary responses, which are evoked by task demands [30, 129, 152], and tonic pupil activity, representing the pupil diameter during resting states [33]. This differentiation arises from the fact that the norepinephrine (NE) system plays a role in modulating the pupil diameter [65, 108] in both phasic and tonic activity features [72]. According to the adaptive gain theory [5], stronger phasic NE activity and larger phasic pupil diameter are associated with higher task engagement [53, 65, 72]. Consequently, tracking task-related (phasic) pupil diameter modulations provides a suitable means to assess task engagement and subjective effort - task performance tradeoff [85]. In the context of MR adaptive

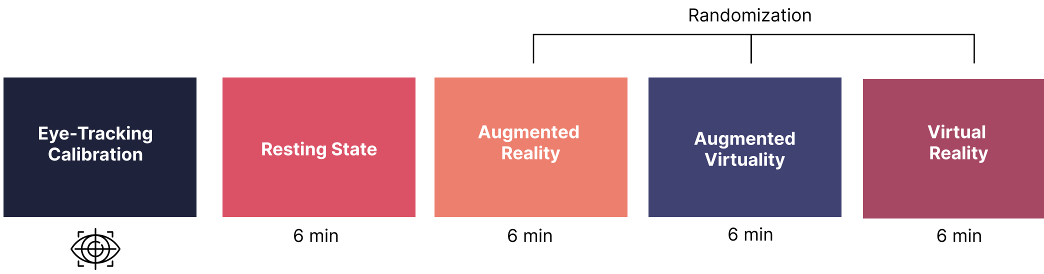


Fig. 2. *Experiment Procedure.* The experiment encompassed five blocks. In the first block, participants performed a calibrated eye tracker and, in the Resting state block, participants relaxed in the neutral VR environment without distracting elements. Finally, the experimental blocks started, manipulating Manifestation in a randomized order. In between blocks, participants filled in NASA-TLX reporting their perceived workload.

systems, Lindlbauer et al. [88] employed the Index of Pupillary Activity (IPA) to design context-aware MR interfaces and adjust application display, location, timing, and amount of information. Other work employed saccade frequency as a model for the workload to perform covert scene changes when the user is occupied [98].

3 User Study

The primary objective of our study is to systematically examine how different manifestations of the reality-virtuality continuum affect typing performance and if we can observe variation in workload and engagement states as shown by different physiological signals (ECG, EEG, eye tracking). Building on this, we aim to understand how the manifestations influence the physiological markers associated with workload and task engagement. Thus, we conducted a within-subjects experimental design with one independent variable Manifestation on the reality-virtuality continuum (three levels: *AR*, *AV*, and *VR*), see Figure 1.

In the *AR* condition, participants typed on a physical keyboard with a view of their surrounding physical environment with typing content displayed over a digital layout. In the *AV* condition, participants typed on a view of the keyboard, and their hands were blended into the virtual environment. And finally, in the VR condition, participants typed on a high-fidelity haptic VR keyboard with virtual hands. To enhance the ecological validity of the study, we conducted the experiment in a real office. We seated participants at a standard office desk facing the door – a typical setup in our institution. As such, the participants will be confronted with real-world visual distractions in the *AR* condition, e.g., bypasses Figure 1 left. On the other hand, auditory instructions from the environment are noticeable in all conditions. We excluded a fully physical condition, i.e. physical reality without HMD, as we anticipate that participants would perform best and experience the least workload in a fully physical condition due to their greater familiarity with physical keyboards. Second, the AR condition, besides requiring the headset and displaying a virtual textbox, includes elements of a physical condition, such as using a physical keyboard and the surrounding physical environment.

3.1 Apparatus

We designed the virtual environment for the study in Unity (Version 2021.3.19f1) and presented it via a Varjo XR-3 MR headset with a resolution of $1920 \times 1920 px$ per eye for the focus area and $2880 \times 2720 px$ per eye for the peripheral area. As per the manufacturer's recommendation, we used

Real Office Virtual Office

Fig. 3. During the study, participants were seated at a desk in a simple and clean office space (left). In the VR and AV conditions, a corresponding virtual office was presented (right).

three HTC Vive lighthouses 2.0 for environment tracking. For hand-tracking, we employed the Ultraleap SDK[1] for Varjo with a tracking frequency of 200Hz.

In line with prior work such as McGill et al. [103], we used a Logitech G810 Orion with a UK English layout as a physical keyboard. In VR, we rendered an identical VR copy of the keyboard using a model provided by Logitech. We used the Logitech G BridgeSDK[2] to render the key movements (up and down) in VR. Moreover, we constantly tracked the physical keyboard with a VIVE tracker to align and move both virtual and physical keyboards, see Figure 5 left. Finally, we recorded environmental noise from the experimental room via a professional microphone (NT-USB+, RØDE, Australia, 48kHz). We recorded the environmental noise to control it in our statistical modeling, see Section 4. Overall, in the AR condition, the average sound level was 64.6 (SD = 5.16) dB on average, for the AV condition, the average sound level measured 69.1 (SD = 8.43) dB and, in the VR condition, the average sound level recorded was 60.1 dB (SD = 2.86).

We acquired four physiological measurements: ECG via PolarH10 chest strap (Polar, Finland, 130 Hz), and EEG signal (LiveAmp Amplifier, BrainProducts, Germany, 500 Hz), and for eye tracking, we employed the built-in system in the Varjo XR-3 (Varjo, Finland, 200 Hz) employing the Tobii XR SDK [3]. Physiological data were streamed within the Unity VR environment within the Lab Streaming Layer (LSL) framework[4] to the acquisition PC (Windows 10, HP Z1 Entry Tower G6, i9, 3,8 GHz, 32GB RAM).

We designed the VR and AV environments for the experiment based on a virtual office; see Figure 4. The virtual office was a room in an apartment. The desk in the room corresponded to the desk in the real world, with biophilic and realistic elements, i.e., plants and office stationery. On the left side of the table was a floor-to-ceiling window. The virtual office design prioritized a realistic office rendering while not evoking negative cognitive and psychological responses [14, 151] and controlling for saliency and its effect on attention allocation and task performance [132]. We maintained the luminance consistent across conditions, following eye-tracking best practices [22, 99] with 180 nits as a comfortable overall fit.

[1] https://docs.ultraleap.com/varjo/
[2] https://github.com/Logitech/logi_bridge_sdk/tree/master
[3] https://developer.tobii.com/xr/
[4] https://github.com/labstreaminglayer/

Fig. 4. Frontal and side view of the office model used for the VR and AV conditions.

3.2 Dependent Variables

We evaluated four aspects of the MR-typing interaction: (i) typing performance based on previous work [102, 139], (ii) accuracy and time to first key press to gauge how effectively participants could locate the keyboard [102], (iii) perceived workload (raw NASA-TLX [60]), (iv) physiological correlates of task-engagement extracted from ECG (HR), EEG (alpha and theta powers), and eye tracking (IPA and saccade frequency). For EEG, we selected one more EEG feature related to workload. We chose the workload index, defined as the ratio between alpha (Pz, P3, P4) [13, 94], and theta, for which we chose frontal channels (F3, F4,Fz, Cz, F7, F8) [81, 94]. Research indicates that task load manipulations are followed by an increase of theta band activity in frontal brain regions, followed by a decrease in alpha power in the parietal areas [109, 128, 133].

3.3 Physiological Recording and Processing

In this section, we report the physiological measures collected and the preprocessing stages that allowed us to compute physiological correlates of workload and engagement. We acquired four physiological measures, i.e., EEG, eye tracking, ECG, and EDA. However, due to strong motion artifacts and missing data, the EDA signal delivered by BITalino biomedical toolkit (BITalino, 1000 Hz) [12] was unusable. Thus, we do not report them in this work. For EEG Alpha and Theta and heart rate we used the Resting state condition for normalization [70].

3.3.1 EEG Recording and Preprocessing. EEG data were recorded from 32 Ag-AgCl pin-type passive electrodes mounted over a water-based EEG cap (R-Net, BrainProducts GmbH, Germany) at the following electrode locations: Fp1, Fz, F3, F7, F9, FC5, FC1, C3, T7, CP5, CP1, Pz, P3, P7, P9, O1, Oz, O2, P10, P8, P4, CP2, CP6, T8, C4, Cz, FC2, FC6, F10, F8, F4, Fp2, AF7, AF3, AFz, F1, F5, FT7, FC3, C1, C5, TP7, CP3, P1, P5, PO7, PO3, Iz, POz, PO4, PO8, P6, P2, CPz, CP4, TP8, C6, C2, FC4, FT8, F6, F2, AF4, AF8 according to the 10–20 system. One LiveAmp amplifier acquired EEG signals with a sampling rate of 500 Hz. All electrode impedances were kept below ≤ 20 kΩ. We used FCz as an online reference and Fpz as ground. For offline preprocessing, we used MNE Python [57]. We first notch-filtered at 50 Hz to remove power line interference, followed by a band-pass filter between 1-50 Hz to eliminate noise at high and low frequencies. Next, we re-referenced the signal to the common average reference (CAR) and applied the Infomax extended algorithm for Independent Component Analysis (ICA) [86]. We utilized the "ICLabel" MNE plugin [117] for automatic artifact detection and correction. We considered each epoch as a typing trial. Epochs that showed blinks, eye movement, muscle, or single-channel artifacts in any of the electrodes were rejected.

Physical Keyboard AR Condition | **Augmented Reality Keyboard** AV Condition | **Virtual Reality Keyboard** VR Condition

Fig. 5. Top view of the keyboards as seen by study participants. We used a UK keyboard layout.

3.3.2 *Eye Tracking Data Preprocessing.* Before analysis, we performed outlier removal and filtering of all the collected gaze data. Our data cleaning consisted of several steps: 1) data points with invalid data flags from the eye tracker for either eye were discarded, 2) any data point with an eye velocity over $1000^{\circ}/s$ or acceleration over $100,000^{\circ}/s^2$ for either eye was disregarded as the movement is considered physically impossible [64], 3) blinks and accompanying artifacts were removed. Blinks were defined as missing data from the eye tracker, which have durations between 75ms and 500ms. Data 200ms before and after the blinks were removed [9]. 4) the disregarded data was linearly interpolated, 5) data was lastly smoothed with a 6th order Butterworth filter whose cutoff frequency was set at .15 Hz [148].

3.3.3 *ECG Recording and Preprocessing.* We collected ECG data at a sampling rate of 130 Hz using a Polar H10 chest strap (Polar, Finland). Before data collection, the ECG electrodes were moistened with lukewarm water and positioned over the xiphoid process of the sternum, just below the chest muscles. Our analysis of the ECG data focused on HR. For processing the ECG data, we utilized the Neurokit Python Toolbox [96]. Initially, we applied a Finite Impulse Response (FIR) band-pass filter with a range of 3 to 45 Hz and a 3rd order to preprocess the ECG signal. Subsequently, we employed Hamilton's method [59] to segment the signal and identify the QRS complexes.

3.4 Procedure

Upon the participants' arrival, we explained the study procedure, answered any open questions, and asked participants to give informed consent. The experimenter proceeded to configure the EEG and ECG recording setup. Then, we asked participants to wear the Varjo XR-3 headset and completed a one-point eye tracking calibration. Next, the experimenter calibrated the hand tracking by aligning the actual hands of the participant with the tracked rendered hands in the AR scene, ensuring that the fingers were aligned with the keyboard. Then, we started the experimental procedure. First, the participants observed a 6-minute resting state, where they sat comfortably in a neutral VR, i.e., Unity skybox, without distracting elements, keeping their hands on their thighs without moving. The resting state is a basal condition for normalizing the experimental conditions. After the Resting State, participants moved to the experimental phase, which consisted of three randomized experimental conditions lasting six minutes each.

The typing task was inspired by previously established work with physical [38, 84] and virtual keyboards in different MR environments [58, 102]. Participants were instructed to type phrases for six minutes during each MR condition, presented in randomized order, see Figure 2. At the beginning of every condition, one training phrase was provided to familiarize with the keyboard layout, as in McGill et al. [102]. Drawing from previous work [84, 102], we employed the MacKenzie 500 phrase set. Phrases were randomly chosen and displayed above the text entry field, see Figure 1. Participants were instructed to type as accurately as possible. Thus, they had the option but were not obligated to correct any errors by using the backspace key, aligning with error correction practices outlined in prior work [102, 140]. Once participants completed typing a single sentence, they pressed the "Enter" key to move to the next sentence. In between conditions, participants filled out the raw NASA TLX [60] to evaluate the perceived workload. Overall, the experiment lasted 45 minutes.

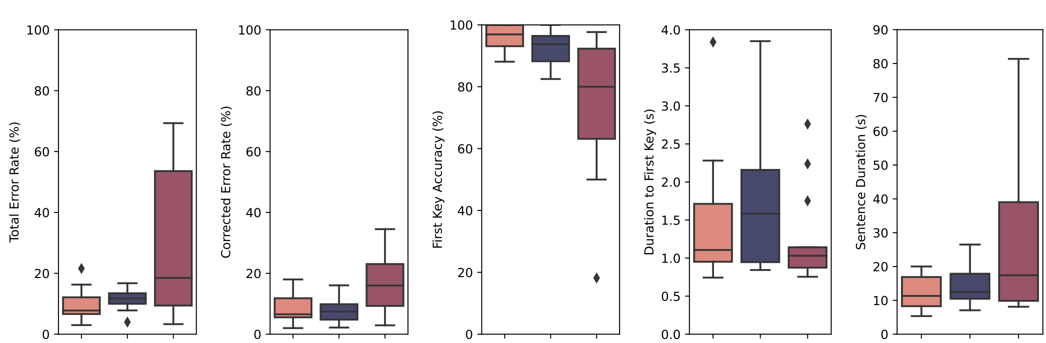
3.5 Participants

A sample of 18 participants voluntarily participated in the study. We excluded four participants due to inadequate EEG data quality, as identified by RANSAC, which revealed that over 50% of electrodes of interest for computing alpha and theta frequencies were classified as "bad," compromising data reliability due to a low SNR as described in Bigdely-Shamlo et al. [17]. Thus, we employ a final sample size of 14 participants (M = 27.9, SD = 4.1; 4 female, 6 male, and 4 diverse). Participants provided written informed consent before their participation. None of the participants reported a history of neurological, psychological, or psychiatric symptoms and all had normal or corrected-to-normal vision. Participants reported spending 113 ± 23 hours per week using computers. The average self-assessed expertise level, on a scale from 1 (*novice*) to 10 (*expert*), was 7.2 ± 1.2, in line with previous work [104]. All participants reported prior experience with AR (M = 4.12, SD = 1.11), AV (M = 1.33, SD = 1.23), and VR (M = 5.4, SD = 2.1), rated on a scale from 1 (not at all familiar) to 7 (extremely familiar) as in previous work [23]. The study met the criteria for fast-track conditions set by the local institutional ethics board.

4 Results

In this section, we first present results on typing metrics and reaction times, heart rate and EEG correlates of attention and task engagement, i.e., Alpha and Theta powers, EEG index of cognitive workload, alpha-to-theta ratio, and, lastly, eye tracking measures of workload and engagement. We employ a Linear Mixed Model (LMM)[5] for all measures on single sentence trials to counteract effects of learning and fatigue [7], as the response in a trial is usually heavily influenced by what happened in the preceding trial. Using mixed-effects models, experimental noise sources are brought under statistical control. We report the effect sizes as for continuous variables (physiological features and typing metrics), we employ the delta total (δ_t) as our effect size measure [61]. We selected the

[5]We used a REML estimation method and Satterthwaite's approximation for degrees of freedom



(a) TER (b) CER (c) First Key Accuracy (d) Duration to First Key (e) Sentence Duration

Fig. 6. Comparison of typing performance measures across conditions including Total Error Rate (TER), Corrected Error Rate (CER), First Key Accuracy, Duration to First Key, and Sentence Duration. Error bars represent Standard Deviation.

Table 1. Results from LMM comparing various metrics across AR, AV, and VR conditions. The table reports the regression coefficients (β), p-values, delta total (δ_t) as a measure of effect size, and the 95% confidence intervals (CI) for the intercept (AR) and comparisons between AR vs. AV and AR vs. VR. The regression coefficient (β) indicates the magnitude and direction of the effect. The delta total (δ_t) quantifies the practical significance of the observed difference, and the 95% confidence interval (CI) provides a range within which the true effect size is expected to lie with 95% confidence. Acronyms used include TER for Total Error Rate, CER for Corrected Error Rate, FKA for First-Key Accuracy, DTFK for Duration To First Key, SDur for Sentence Duration and WPM for Words per Minute.

	Intercept (AR)				**AR vs. AV**				**AR vs. VR**			
	β	p	δ_t	95% CI	β	p	δ_t	95% CI	β	p	δ_t	95% CI
TER	.10	<.001	−.11	[-.26, .05]	.02	.18	.09	[-.04 - .21]	.08	<.001	.45	[.31, .59]
CER	.8	<.001	−.09	[-.32, .13]	-1.7^{-3}	.792	−.09	[-.32, .13]	.05	<.001	.43	[.30, .57]
FKA	.96	<.001	.15	[.03, .26]	−.04	.024	−.15	[-.28, -.02]	−.11	<.001	−.40	[-.54, -.26]
DTFK	1.44	<.001	.08	[-.13, .30]	.16	.213	.08	[-.05, .21]	−.10	<.001	−.05	[-.19, .09]
SDur	12.73	<.001	-1.53^{-3}	[-.22, .21]	1.50	.162	.09	[-.04, .21]	6.08	<.001	.36	[.22, .50]
WPM	34.93	<.001	−.44	[-1.00, .12]	−.43	.74	−.02	[-.14, .10]	−2.43	.085	−.12	[-.25, .02]
α	4.03	<.001	−.13	[-.56, .29]	.08	<.001	.16	[.07, .25]	.07	.003	.15	[.05, .24]
θ	3.66	<.001	−.09	[-.53, .36]	.04,	<.001	.10	[.01, .18]	.03	.203	.06	[-.03, .15]
α/θ	1.08	<.001	−.09	[-.20, .01]	.01	.018	.15	[.03, .27]	.01	.027	.15	[.02, .28]
HR	84.61	<.001	−.06	[-.51, .40]	−1.28	<.001	−.11	[-.15, -.07]	−1.27	<.001	−.11	[-.16, -.06]
IPA	.33	<.001	−.08	[-.18, .02]	.10	.032	.12	[.01, .23]	.01	.799	.02	[-.10, .13]
Sacc. Freq.	4.96	<.001	.26	[-.09, .60]	−.26	<.001	−.17	[-.26, -.08]	−.51	<.001	−.34	[-.43, -.25]

formula: `measure ~ Manifestation + (1 | participant) + (1 | sound)` for the LMM. We summarize our results in Table 1. Finally, for perceived workload evaluated through NASA-TLX, we use one-way repeated measures ANOVA or Friedman's test depending on normality assumption violation as per the Shapiro-Wilk test, see Section 4.5. For effect size, we use partial eta square (η^2) for ANOVA and Cohen's d for post hoc comparisons. We excluded the last sentence typed in each condition as this abruptly ended when the condition duration expired.

4.1 Typing Metrics

4.1.1 Total Error Rate (TER). For the Total Error Rate (TER), see Figure 6a, we chose the AR condition as an intercept for the LMM. The model's total explanatory power is weak (conditional $R^2 = .08$) and the part related to the fixed effects alone (marginal R^2) is .03.Compared to this, the AV condition showed a non-significant increase in TER ($\beta = .02$, $p = .18$), while in contrast, the VR condition significantly increased the TER ($\beta = .08$, $p < .001$). Further analysis revealed that the standardized effect sizes (δ_t) were as follows: for the AV condition, the effect size was minor ($\delta_t = .09$, 95% CI $[-.04, .21]$), indicating minimal practical impact. However, the VR condition exhibited a substantially larger effect size ($\delta_t = .45$, 95% CI $[.31, .59]$), indicating a pronounced increase in error rates when moving from AR to VR. This indicates that participants made more errors in the VR condition compared to the AR and AV conditions. These effect sizes further emphasize the impact of the VR condition on increasing TER.

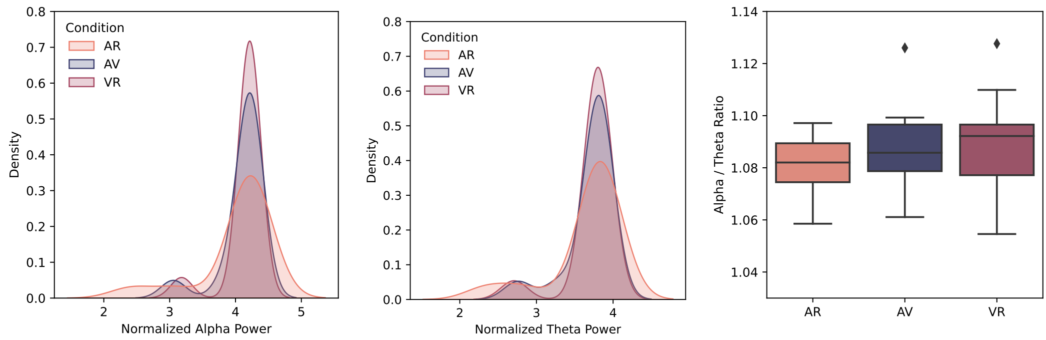
4.1.2 Corrected Error Rate (CER). The base CER, see Figure 6b, for the AR condition was .08 ($p <$ .001). The model's total explanatory power is moderate (conditional $R^2 = .18$) and the part related to the fixed effects alone (marginal $R^2 = .04$). Comparatively, the AV condition did not significantly alter the CER ($beta = -.002$, $p = .792$), indicating no notable difference from the AR condition. On the other hand, the VR condition significantly increased the CER ($beta = .046, p < .001$), highlighting a higher corrected error rate during VR conditions. Analysis of standardized effect sizes showed that the effect size for the AV condition was minor and not statistically significant ($\delta_t = -.02$, 95% CI $[-.14, .10]$), confirming its negligible impact on CER. Conversely, the effect size for the VR condition was substantial ($\delta_t = .43$, 95% CI $[.30, .57]$), showing the VR condition's significant influence on increasing the error rate.

4.1.3 First-Key Accuracy (FKA). When inspecting the accuracy of typing a first key correctly (see Figure 6c), the linear mixed model showed a significant main effect. The model's total explanatory power is weak (conditional $R^2 = .04$) and the part related to the fixed effects alone (marginal $R^2 = .02$). We observe significantly decreased accuracy in the AV and VR conditions (AV: $beta = -.04$, $p = .024$; VR: $beta = -.11$, $p < .001$, respectively), indicating lower first key correctness in these conditions compared to AR. Effect sizes analysis showed that the effect size for the AV condition was modest but statistically significant ($\delta_t = -.15$, 95% CI $[-.28, -.02]$), suggesting a noticeable decrease in first key accuracy. The effect size for the VR condition was more pronounced ($\delta_t = -.40$, 95% CI $[-.54, -.26]$), significantly emphasizing the substantial impact of the VR condition on reducing first key accuracy.

4.1.4 Duration To First Key (DTFK). For the time needed between the start of the trial and the first keystroke (see Figure 6d), we found no significant differences across conditions. The model's total explanatory power is weak (conditional $R^2 = .12$) and the part related to the fixed effects alone (marginal $R^2 = 2.63 \times 10^{-3}$). The AV and VR conditions demonstrated non-significant effects, with beta values of .16 and -.10, respectively, and correspondingly $p = .213$, $p = .473$. This suggests that the MR manifestations (AV and VR) did not significantly alter the participants' readiness or reaction time in initiating the typing task. Further standardized effect sizes analysis revealed minor effect sizes for both conditions; for AV, the effect size was small and not statistically significant ($\delta_t = .08$, 95% CI $[-.05, .21]$), and for VR, similarly minor and non-significant ($\delta_t = -.05$, 95% CI $[-.19, .09]$).

4.1.5 Sentence Duration. Examining the effects of individual MR conditions, the AV condition showed a non-significant positive effect with a beta value of 1.50 ($p = .162$), while the VR condition demonstrated a significant positive effect, with a beta value of 6.08 ($p < .001$). The model's total explanatory power is moderate (conditional $R^2 = .14$) and the part related to the fixed effects alone (marginal $R^2 = .02$). This reveals that participants took longer to type a complete sentence in VR. Standardized effect sizes analysis showed that the effect size for the AV condition was minor and not statistically significant ($\delta_t = .09$, 95% CI $[-.04, .21]$), suggesting a negligible impact on sentence duration. Conversely, the effect size for the VR condition was substantial ($\delta_t = .36$, 95% CI $[.22, .50]$), significantly indicating that the VR environment greatly increased the time required to complete a sentence, reinforcing the significant increase observed in the statistical model.

4.1.6 Words per Minute (WPM). The model demonstrated substantial explanatory power (conditional $R^2 = .61$) with marginal $R^2 = .0021$. The intercept for AR was estimated at 34.93 (95% CI [23.40, 46.46], $t(645) = 5.95$, $p < .001$), with a standardized effect size (δ_t) of -.44 (95% CI [-1.00, .12]), indicating a moderate negative effect size. The effect of AV was non-significant and negative ($\beta = -.43$, 95% CI [-2.96, 2.11], $t(645) = -.33$, $p = .74$; Std. $\delta_t = -.02$, 95% CI [-.14, .10]), suggesting a moderate effect size. Similarly, the effect for VR was non-significant and negative ($\beta = -2.43$, 95% CI [-5.20, .34], $t(645) = -1.73$, $p = .085$; Std. $\delta_t = -.12$, 95% CI [-.25, .02]), indicating a small effect size. These results confirm that while the model explains a significant amount of variability in *WPM*, the effects of AV and VR were minimal and non-significant.



(a) Alpha (b) Theta (c) Alpha Theta Ratio

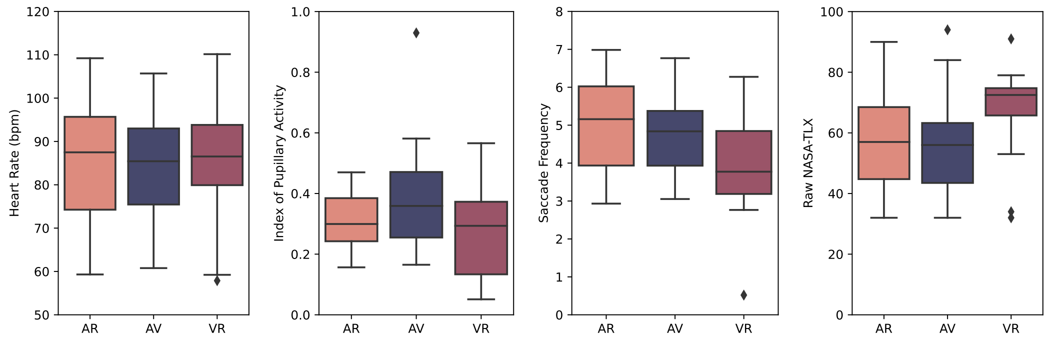
Fig. 7. Analysis of participants' EEG data across MR conditions as an indicator of attention allocation, task engagement, and workload. We present alpha and theta powers in a distribution plot while the alpha-theta ratio is a boxplot. We report increased alpha and Theta in AV, signaling higher internal attention and task-engagement. A lower alpha theta ratio signals increased workload, which is detected in AR condition (c). In (c) error bars represent Standard Deviation.

4.2 Electroencephalogram (EEG)

4.2.1 Alpha power. The model's total explanatory power is substantial (conditional $R^2 = .57$) and the part related to the fixed effects alone (marginal $R^2 = 4.86 \times 10^{-3}$). Within the LMM, the effect of the AV condition was statistically significant and positive, with a beta value of .08 ($95\%\,CI = [.04, .12], p < .001$), see Figure 7a. Additionally, the effect of the VR condition was statistically significant and positive, evidenced by a beta value of .07, $p = .003$. The significant positive effects in alpha power for both AV and VR conditions suggest that participants experienced heightened internal attention and engagement in these MR environments. When inspecting standardized effect sizes, we found that the effect size for the AV condition was substantial ($\delta_t = .16$, 95% CI [.07, .25]), and similarly, the effect size for the VR condition was also considerable ($\delta_t = .15$, 95% CI [.05, .24]). These effect sizes further support the findings that AV and VR conditions significantly enhance participants' internal attention, as reflected in increased alpha power.

4.2.2 Theta power. The model's total explanatory power is substantial (conditional $R^2 = .62$) and the part related to the fixed effects alone (marginal $R^2 = 1.48 \times 10^{-3}$). Regarding fixed effects, the AV condition was found to have a positive and statistically significant influence on theta power, denoted by a beta value of .04, (95% CI [6.05e-03, .08], $p = .022$), see Figure 7b. VR condition showed a positive, albeit not statistically significant, with a beta of .03 (95% CI [-.01, .07], $p = .202$). These results suggest increased engagement in the AV manifestation. Although the VR condition was positive but not significant, it could imply that while the VR environment may influence engagement, it is not as pronounced or consistent as in the AV environment. Further analysis with standardized effect sizes revealed that both AV and VR conditions had considerable effect sizes, with $\delta_t = .15$ for AV ($95\%\,CI = [.03, .27]$) and $\delta_t = .15$ for VR ($95\%\,CI = [.02, .28]$). These effect sizes further underscore the potential influence of both environments on enhancing theta power, which is indicative of heightened engagement, even though the effect in VR was not statistically significant.

4.2.3 Alpha-to-Theta Ratio (Workload Index) . The model's total explanatory power is limited (conditional $R^2 = .02$), and the part related to the fixed effects alone (marginal $R^2 = 5.19 \times 10^{-3}$). When



(a) Heart Rate (b) IPA (c) Saccade Frequency (d) Perceived Workload

Fig. 8. Physiological measures from ECG (Heart Rate) and eye tracking (Index of Pupillary Activity - IPA and Saccade Frequency), and Raw NASA-TLX scores across conditions as indicators for workload. Error bars represent Standard Deviation.

investigating the effects on the EEG workload index, both the AV and VR conditions significantly positively affected the workload index, see Figure 7c. Specifically, the AV condition exhibited a beta value of .01 (95% CI [2.07e-03, .02], $p = .018$). Similarly, the VR condition demonstrated a beta of .01 (95% CI [1.42e-03, .02], $p = .027$). These findings suggest that AV and VR conditions are associated with a decrease in cognitive workload, as indicated by the positive beta values in the EEG workload index, compared to the AR condition. However, further analysis using standardized effect sizes revealed that the effect sizes were negative, albeit not statistically significant for both AV ($\delta_t = -.07$, 95% CI = [-.19, .05]) and VR ($\delta_t = -.007$, $95\% \, CI = [-.14, .13]$).

4.3 Heart Rate (HR)

The model's total explanatory power is substantial (conditional $R^2 = .94$) and the part related to the fixed effects alone (marginal $R^2 = 2.04 \times 10^{-3}$). Analyzing the different conditions revealed that AV and VR conditions had statistically significant negative effects on the heart rate (see Figure 8a). Specifically, the AV condition showed a beta value of -1.28 and $p < .001$. Similarly, the VR condition indicated a $\beta = -1.27$ with $p < .001$. This decrease in heart rate could imply that participants were more relaxed or less stressed in AV and VR conditions compared to the AR condition. Further analysis with standardized effect sizes showed that the effect sizes were consistent and significant for both conditions, with $\delta_t = -.11$ for AV ($95\% \, CI = [-.15, -.07]$) and $\delta_t = -.11$ for VR ($95\% \, CI = [-.16, -.06]$). These effect sizes further substantiate the decrease in heart rate, indicating a significant reduction in physiological stress in both AV and VR environments.

4.4 Eye Tracking

4.4.1 Index of Pupillary Activity (IPA). The model's total explanatory power is limited (conditional $R^2 = .02$) and the part related to the fixed effects alone (marginal $R^2 = 3.10 \times 10^{-3}$). Within the generalized linear mixed model, the AV condition significantly positively affected IPA ($\beta = .14$, 95% CI [.05, .23], $p = .002$), as shown in Figure 8b. The VR condition, on the other hand, showed a significantly decreased IPA with such significant influence ($\beta = -.04$, 95% CI [-.07, -.01] $p = .002$). This indicates that the AV environment may have led to higher cognitive engagement and mental effort, as evidenced by the increased IPA. In contrast, the VR environment showed a significantly decreased IPA compared to the AV condition, suggesting that participants experienced reduced cognitive engagement and mental effort. Further analysis using standardized effect sizes revealed that for AV, the effect size was positive and statistically significant (δ_t = .12, $95\%CI[.01, .23]$), aligning with the increase indicated by the beta value. However, the effect size for VR was small and not statistically significant (δ_t = .02, $95\%CI[-.10, .13]$), corresponding with the less pronounced impact on IPA as suggested by its beta value.

4.4.2 Saccade Frequency. The analysis of saccade frequency showed a significant main effect of condition (see Figure 8c). The model's total explanatory power is substantial (conditional $R^2 = .48$) and the part related to the fixed effects alone (marginal $R^2 = .02$). Specifically, the AV condition was associated with a significant decrease in saccade frequency, with a coefficient of $\beta = -.26$, (95% CI [-0.38, -.13], $p < .001$). Conversely, the VR condition demonstrated a pronounced negative effect on saccade frequency, evidenced by $\beta = -.51$ (95% CI [-.65, -.37], $p < .001$). This indicates that both AV and VR conditions led to reduced saccade frequency, with a more pronounced effect in VR, suggesting higher visual focus in these environments compared to the baseline AR condition. Inspecting standardized effect sizes revealed significant and negative effects for both conditions: for AV, the effect size was considerable ($\delta_t = -.17$, $95\%CI[-.26, -.08]$), and for VR, even more substantial ($\delta_t = -.34$, $95\%CI[-.43, -.25]$). These effect sizes further substantiate the findings that both AV and VR environments significantly reduce saccade frequency, enhancing visual focus and potentially reducing cognitive load.

4.5 Perceived Workload

Shapiro-Wilk testing showed a normal distribution of Raw TLX scores ($W = .962, p = .182$). Thus, we conducted a repeated-measures ANOVA to investigate the influence of MR manifestations on perceived workload. The analysis revealed a significant effect of the MR condition ($F(2, 26) = 4.03, p = .029, \eta^2 = .065$). The partial eta squared value of .065 indicates a moderate effect size, suggesting that the MR environments have a noticeable impact on perceived workload overall. This effect size reflects the variance in perceived workload attributed to the differences between the MR conditions. We applied post-hoc pairwise comparisons utilizing Bonferroni-adjusted paired t-tests to discern the differences between conditions. The outcomes indicate no significant distinction between the AR condition and both others (AR AV: $p = 1.0$; AR VR: $p = .220$). However, a significant difference was observed between the AV and VR conditions ($p = .032$), as illustrated in Figure 8d. Those results suggest that while there was no significant difference in perceived workload between the AR condition and either the AV or VR conditions, participants perceived a significantly higher workload in the VR condition compared to the AV condition. In our analysis, we computed Cohen's d to measure the effect size for the difference between the AV and VR conditions on perceived workload. Cohen's d was $-.80$ ($95\%CI[-2.04, .45]$), indicating a large and negative effect, with the VR condition associated with a lower perceived workload than the AV condition. The 95% confidence interval suggests variability in the effect size estimate but maintains a tendency toward a negative effect.

5 Discussion

We conducted a study to investigate the influence of MR environments on typing performance, subjective workload, and physiological correlates of workload and task engagement. We aimed to understand better how engagement and workload vary across the continuum using physiological features not previously investigated in a typing task in MR to inform the design of current and future MR systems. We did this to first replicate the results of the seminal work by McGill et al. [102] and extend it by integrating physiological correlates of workload and task engagement varying across MR manifestations. We first summarize our results, then relate our findings and replication of previous work and its relation with physiological workload and task engagement. We then conclude with insights from interruption science to adaptive transitional interfaces and highlight future fieldwork.

5.1 Summary of Results

5.1.1 Typing Metrics. The evaluation of typing metrics across AR, AV, and VR conditions revealed significant differences, particularly between AR and VR. The VR condition consistently hindered performance, leading to a significant increase in both TER and CER, as well as a marked decline in FKA. While the AV condition showed some negative effects, these were less pronounced than VR. In terms of timing, VR significantly increased the time taken to complete a sentence, but had no significant effect on the duration to the DTFK. Neither AV nor VR significantly impacted WPM, suggesting minimal effects on typing speed.

5.1.2 Physiological Features.

EEG. The EEG metrics across AR, AV, and VR conditions showed significant differences, particularly in AV and VR environments. Both AV and VR significantly increased alpha power, indicating heightened attention and engagement. AV also significantly increased theta power, while VR's increase was not statistically significant, though both had considerable effect sizes. The alpha-to-theta ratio suggested reduced cognitive workload in AV and VR compared to AR.

Heart Rate and Eye Tracking. AV and VR conditions significantly lowered heart rate compared to AR, suggesting reduced stress and greater relaxation in these environments. For eye-tracking metrics, AV increased the IPA, indicating higher cognitive engagement, while VR decreased IPA, implying reduced engagement. Additionally, both AV and VR significantly reduced saccade frequency, with VR having a stronger effect, indicating enhanced visual focus and potentially lower cognitive load in these manifestations.

5.1.3 Perceived Workload. The results showed that MR conditions significantly affected perceived workload, with participants reporting a higher workload in VR compared to AV. No significant difference was found between AR and the other conditions.

5.2 Replication of Typing Performance Across MR Manifestations

McGill et al. [102] reported how augmenting a virtual environment with a real keyboard reduced the error rate and allowed users to better identify overall keyboard position and single keystrokes compared to VR. However, low stereoscopic resolution and latency in VR led to negligible gains in AV. We come to a similar conclusion from the typing performance, as AV showed a decreased error rate while still having lower first-key accuracy.

Adding to this, Pham and Stuerzlinger [115] demonstrated that their HawKEY system, which utilizes a portable keyboard setup with various visualization methods, significantly improves text entry rates and reduces error rates in VR. Their study found that the Video condition, which displays the keyboard only when the user looks down, and the Point Cloud condition, similar to our video-cutout, were particularly effective. These methods yielded text entry speeds close to those in real-world settings and minimized errors, highlighting the importance of effective keyboard visualization in AV.

Our results further contribute to the comparison with the AR reality condition. In our setup, the AR condition mimicked real-world MR productivity settings, with augmented display and environmental noise, while still being kept under statistical control. Here, we did not find any significant difference between the AV and AR conditions regarding performance accuracy measures (TER and CER) and the time needed to type a single sentence. This result is confirmed by the reports on perceived workload, where no significant differences were found in the AR-AV contrast but more interestingly also in the AR-VR contrast. We argue that here, subjective measures point towards an increased workload for VR and no difference when comparing a condition where the real environment is displayed (AR) and a scenario where only task-relevant objects, i.e., the keyboard and the hands, are displayed (AV).

Our research builds upon the findings of McGill et al. [102] and Pham and Stuerzlinger [115] by extending the evaluation of text entry methods to more realistic MR productivity settings and incorporating augmented environmental noise. While McGill et al. [102] demonstrated the benefits of real keyboard augmentation in VR, our study explored typing behavior across the full virtuality–reality continuum, including AV and AR. Furthermore, the work by Pham and Stuerzlinger [115] on the HawKEY system highlighted effective visualization techniques for VR text entry, which we confirm and extend by demonstrating their applicability and effectiveness in AR settings as well. Our findings suggest that the advantages of advanced keyboard visualization techniques, such as those used in HawKEY, are robust across different MR environments, thereby enhancing user productivity and comfort in VR and AR scenarios.

5.3 Physiological Features Highlight Improved Engagement Without Task Overload in AV

Next to replicating previous work, our contribution complements and integrates behavioral results with implicit measures of task engagement, attention allocation, and workload. Physiological correlates of engagement and workload discriminated across conditions, pointing towards an improved task engagement in AV while showing increased workload and external attention allocation in AR. Alpha power increased in the AV condition, indexing either a decreased workload [123] and an internal attention state [13]. Here, first, the AV condition shielded users from environmental distractors while engaging them in a virtual environment that allowed them to focus on the task, i.e., an internal attention state. Internal attention proves beneficial in typing tasks as it facilitates a focused and uninterrupted workflow. This interpretation is supported by the not significant measures in typing measures, where AV promoted comparable accuracy and speed to AR, minimizing the distractions and cognitive load associated with processing external stimuli. Similarly, EEG correlates of engagement, i.e., increased theta power, supporting the increased internal state in the AV condition [35]. On the other side, an opposite direction between alpha and theta is used to compute the Workload index [80, 128]. Here, the AR condition showed increased physiological workload as compared to AV.

The eye tracking features complement results on physiological engagement and workload derived from EEG. Here, IPA showed higher AV values than AR, while VR did not show significant variations from AR. IPA can be interpreted as a measure of mental effort, and sensitive to mental fatigue; the more fatigue is experienced, the smaller is the IPA [1]. Moreover, recent work proved how constant or increasing IPA changes if participants were positively engaged in solving a task [113] and correlating with EEG-related measures [39]. Moreover, we found reduced saccade frequency in AR compared to AV and VR. Here, reduced saccade frequency in AV and VR can reflect decreased visual exploration, implying that users found it easier to focus on relevant information, thus possibly reducing cognitive load and enhancing task engagement [41, 82]. Conversely, the higher saccade frequency observed in the AR condition might indicate a heightened level of visual exploration and cognitive workload, as users might be trying to integrate and process a higher volume of both virtual and real-world stimuli, which can potentially increase the mental effort and fatigue experienced during the task [138].

Finally, inspecting peripheral physiological stress and arousal measures, heart rate increased in the AR condition. AR integrated a more complex blend of real and virtual elements, increasing mental effort and ultimately increasing physiological stress. At the same time, AV allowed for optimal information presentation, and decreasing the visual load did not majorly impact the participants' physiological arousal.

Taken together, physiological data show converging trends across the MR conditions. In AV and VR, increased alpha and theta power suggest enhanced internal attention and engagement, supported by lower saccade frequencies indicating focused cognitive states. This consistency in alpha and theta power enhancements across AV and VR highlights that immersive environments, which reduce real-world distractions, promote higher engagement and internal focus. Conversely, AR showed higher heart rates and saccade frequencies, indicating elevated cognitive load and stress. The need to manage both virtual and real elements in AR contributed to this divergence. While IPA was higher in AV, indicating sustained mental effort, VR did not significantly differ from AR, suggesting that VR may not consistently sustain the same cognitive effort as AV. The primary reason for these differences is how each environment integrates real and virtual elements. AR demands greater cognitive resources to process mixed stimuli, reflected in increased stress and external attention. In contrast, AV and VR provide controlled environments that reduce distractions, allowing users to maintain focus more effectively. By understanding these trends, we can better design MR environments to balance engagement and cognitive load, optimizing them for productivity tasks.

5.4 From Interruption Science to Physiologically-Adaptive Mixed Reality

Our study's findings in the AR and AV conditions can be contextualized within the broader field of interruption science, which examines how external stimuli and interruptions affect task performance and cognitive processes [8]. The presence of environmental distractors in the AR condition mimicked real-world productivity settings and allowed us to explore how such stimuli impact text entry performance. Interestingly, we found no significant differences in performance accuracy measures (TER and CER) or the time needed to type a single sentence between the AR and AV conditions. This suggests that the AV condition effectively mitigates the negative effects of environmental distractors by shielding users from visual distractions and displaying only task-relevant objects (i.e., the keyboard and hands).

Interruption science indicates that distractions and interruptions significantly impair cognitive performance by breaking the user's focus and requiring additional cognitive resources to resume the task at hand [11]. Our findings align with this research, demonstrating that reducing visual distractions from the real world in AR, in the AV condition helps maintain user attention and enhances task engagement. The study by Bailey and Konstan [11] further supports this by showing that deferring interruptions until natural task boundaries lead to higher task performance and reduced annoyance and stress. These insights suggest that designing MR systems that adapt to users' cognitive load and engagement levels by deferring interruptions and minimizing unnecessary stimuli can significantly improve user productivity and comfort.

Here, we position our work within the control theoretic perspective envisioned by McGill et al. [102], where the MR control loop encompasses inputs from reality and feedback from the virtual environment. This high-level system is analogous to on-screen feedback observed during typing tasks. As users engage more intensively with real interactive elements like keyboards or individuals, these elements can be progressively integrated into the VR view. Our approach integrates the physiological computing perspective, which involves processing physiological features, translating them into system responses, and shaping future psychophysiological reactions from the user.

We found that when typing on a physical, real-world keyboard in AV, users have comparable performance but improved engagement compared to AR. These results can be employed to promote transitions within the continuum and allow for the gradual fading of reality [24]. Such transitions need to be context-aware and consider multivariate patterns in the user. Consequently, we adopt a sensor fusion strategy to gauge cognitive load and engagement to offer reliable real-time results. The features we computed have already been employed in previous adaptive systems in AR for IPA [88], ECG in VR [74], and EEG [28] using real-time estimation of different states. We propose to investigate this in the future and expand our framework with sensor fusion approaches and multimodal user state analysis.

Our results have implications for designing an attention-aware system, which seeks to balance a user's need for minimal disruption with an application's need to deliver information effectively. An attention-aware system based on physiological features such as EEG alpha power or IPA could employ a temporal strategy, manipulating when to hide information or render attentional cues in peripheral displays [92, 93, 144]. By deferring presentation until appropriate points during task execution, such a system could mitigate the negative effects of interruptions.

Developing an attention-aware system in the context of MR would require mechanisms to specify user tasks, observe their performance, and learn a task execution model to forecast appropriate interruption points [10]. In MR environments, this means continuously monitoring physiological signals to adaptively manage visual and auditory stimuli based on the user's cognitive state. While building such a system requires significant research, our findings show that this effort is warranted. An effective system could enable users to perform tasks faster, commit fewer errors, and experience less annoyance and physiological stress.

5.5 Limitation and Future Work

Our study, though examining the interactions between cognitive load, and engagement, across behavioral subjective, and physiological variables in MR environments, recognizes certain limitations and foresees promising future work in the field of MR for productivity.

5.5.1 Task Complexity, Ecological Productivity Settings, and Text-Input Techniques. While ecologically valid, the typing task utilized in this study might not fully include the many variables of primary office tasks. Future investigations could thus encompass more complex tasks, such as creating slide presentations or coding, to simulate real-world office environments better. This also applies to interaction with other agents, i.e., human in MR [102] or virtual avatars in VR and AR [31, 67] and the effect varying levels of visual detail and lighting could affect readability and the ease with which users can locate and interact with virtual elements, potentially impacting typing speed and accuracy [55, 126]. Further, we evaluate merely one of several possible text-input techniques, namely typing, as this remains the predominant modality in day-to-day office work. However, MR technology may well bring about a paradigm change for text-input, fueled by further research on alternative modalities (e.g., eye tracking or voice-based).

5.5.2 Environmental Control and Dataset Size. We controlled the visual load and embedded the task at a real office level to maintain ecological validity, VR and AV environments might allow larger opportunities to isolate participants better from potential distractions. This invites further research to evaluate the effects of manipulating sound environments and their congruence with the visual presentation of AV and VR environments. Additionally, while most of our reported effect sizes are small to medium, indicating low power to detect meaningful effects, some smaller effect sizes suggest that larger sample sizes could provide more robust findings. To address this, future research should consider increasing the dataset size to ensure that subtle but significant effects are detected reliably.

5.5.3 Blending of Virtuality and Reality. Our approach focused on three primary conditions: AR, AV, and VR, to explore the continuum of virtuality and reality. Specifically, we intended the VR condition as a purely virtual environment, including the input device, while we designed the

AV condition based on previous work [102, 115], implementing physical information only when task-relevant. This decision was based on balancing ecological validity and the practical constraints of the study. However, we acknowledge that there are many ways to balance virtuality and reality when typing, and our chosen conditions represent just a few points along this continuum. The VR condition tested in this study is further from AV compared to minimal blending, which might limit the resolution of our understanding of the impacts compared to prior work [24, 62, 105]. Alternative approaches to designing AV environments include the work by Cheng et al. [24], which investigates the concept of Diminished Reality (DR). DR involves removing content from a user's visual environment to reduce information overload and improve task performance by diminishing non-relevant objects. By removing virtual objects rather than adding relevant physical elements, DR provides an alternative strategy for balancing virtual and physical content, allowing for improved awareness when interacting in AV. Future research should explore how degrees of virtuality, from minimal blending and other intermediate levels of virtuality, impact attention allocation, task performance and engagement. Alternatively, Hettiarachchi and Wigdor [62] suggests continuously scanning the user's surroundings, selecting physical objects similar to given virtual objects, and overlaying the virtual models onto the selected physical ones. This approach emphasizes the integration of haptic feedback by matching virtual objects to real-world counterparts, enhancing the engagement.

5.5.4 Adaptive MR Systems and Transitions. To develop and evaluate a working MR adaptive system, we need to consider how transitions are accepted by users and their effect on attention, engagement and workload. In previous work in VR, users seemed to prefer simple and short transitions [46]. However, it is not clear if the same either applies across MR manifestations and if a sudden change in sensory information, either by enriching or diminishing the environment, might impact physiological arousal and engagement. Thus, we propose to evaluate different design approaches for transitions in MR and which "direction" across the continuum might better support users' goals in the individual task.

Future Directions in Non-Invasive Monitoring of Engagement. Our study utilized a multimodal approach to collect and validate multiple physiological indicators of engagement. However, the invasive nature of methods like EEG, along with their lengthy setup times, prompts the need for non-invasive alternatives suitable for real-time applications in engagement-aware systems.

Future research should focus on less intrusive technologies such as functional near-infrared spectroscopy [37, 75] and other non-invasive physiological monitoring techniques, such as dry EEG setups [15] or sensing systems directly embedded in a MR system [27]. Additionally, we recognize the need to optimize channel selection from the total EEG channels used in our study. While our current focus was on electrodes in the frontal, parietal, and occipital regions, future work will investigate how each electrode and combination of electrodes can best contribute to classification accuracy. Methodologies like Recursive Feature Elimination and Genetic Algorithms, along with advanced feature importance techniques such as mutual information, permutation importance, shapley additive explanations, or ablation studies, will be employed to systematically evaluate and identify the most relevant EEG channels [89, 124].

6 Conclusion

In our work, we explored and expanded on how varying MR manifestations might influence productivity and engagement at behavioral and physiological levels. We engaged participants in a typing task in three different mixed reality environments: AR, AV, and VR, while monitoring EEG, ECG, and eye tracking correlates of engagement, workload, and attention. Our results replicate findings of prior work in MR, showing that typing performance was optimal when users' could rely on both physical and visual information, i.e., in AR and AV. However, the best typing performance in AR corresponded with increased physiological workload and decreased engagement. Interestingly, AV also showed opposite, positive results in terms of better comfort and focus on the task. Identifying distinct physiological patterns across the reality-–virtuality continuum opens the field to adaptive MR systems grounded in physiological computing and control theory. Our results can thereby inform the design of future dynamic interfaces that adapt to the users' physiological signals, to ensure effective interaction with MR systems.

7 Open Science

We encourage readers to review, reproduce, and extend our results and analysis methods. To achieve this goal, we make available our collected datasets, sentence stimuli, MR environments, and experiment setup and analysis scripts at this link https://osf.io/juk9x/.

Acknowledgments

Francesco Chiossi, Tiare Feuchtner, Abdelrahman Zaky, and Sven Mayer were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with Project ID 251654672 TRR 161. This work has been funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 952026 (https://www.humane-ai.eu/).

References

[1] John L Andreassi. 2013. *Psychophysiology: Human behavior & physiological response.* Psychology Press.

[2] Stephanie Arevalo Arboleda, Franziska Rücker, Tim Dierks, and Jens Gerken. 2021. Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 728, 14 pages. https://doi.org/10.1145/3411764.3445398

[3] Stefan Arnau, Tina Brümmer, Nathalie Liegel, and Edmund Wascher. 2021. Inverse effects of time-on-task in task-related and task-unrelated theta activity. *Psychophysiology* 58, 6 (2021).

[4] Rawin Assabumrungrat, Soravitt Sangnark, Thananya Charoenpattarawut, Wipamas Polpakdee, Thapanun Sudhawiyangkul, Ekkarat Boonchieng, and Theerawit Wilaiprasitporn. 2021. Ubiquitous affective computing: A review. *IEEE Sensors Journal* 22, 3 (2021), 1867–1881. https://doi.org/10.1109/JSEN.2021.3138269

[5] Gary Aston-Jones and Jonathan D Cohen. 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28 (2005), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

[6] Jonas Auda, Uwe Gruenefeld, Sarah Faltaous, Sven Mayer, and Stefan Schneegass. 2023. A Scoping Survey on Cross-reality Systems. *ACM Comput. Surv.* 56, 4, Article 83 (Oct. 2023), 38 pages. https://doi.org/10.1145/3616536

[7] R Harald Baayen, Douglas J Davidson, and Douglas M Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* 59, 4 (2008), 390–412.

[8] Anja Baethge, Thomas Rigotti, and Robert A Roe. 2015. Just more of the same, or different? An integrative theoretical framework for the study of cumulative interruptions at work. *European Journal of Work and Organizational Psychology* 24, 2 (2015), 308–323. https://doi.org/10.1080/1359432X.2014.897943

[9] Tanya Bafna, John Paulin Paulin Hansen, and Per Baekgaard. 2020. Cognitive Load during Eye-Typing. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 23, 8 pages. https://doi.org/10.1145/3379155.3391333

[10] Brian P Bailey, Piotr D Adamczyk, Tony Y Chang, and Neil A Chilson. 2006. A framework for specifying and monitoring user tasks. *Computers in Human Behavior* 22, 4 (2006), 709–732.

[11] Brian P Bailey and Joseph A Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* 22, 4 (2006), 685–708.

[12] Diana Batista, Hugo Plácido da Silva, Ana Fred, Carlos Moreira, Margarida Reis, and Hugo Alexandre Ferreira. 2019. Benchmarking of the BITalino biomedical toolkit against an established gold standard. *Healthcare technology letters* 6, 2 (2019), 32–36. https://doi.org/10.1049/htl.2018.5037

[13] Mathias Benedek, Rainer J Schickel, Emanuel Jauk, Andreas Fink, and Aljoscha C Neubauer. 2014. Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia* 56 (2014), 393–400. https://doi.org/10.1016/j.neuropsychologia.2014.02.010

[14] Lisanne Bergefurt, Rianne Appel-Meulenbroek, and Theo Arentze. 2023. How salutogenic workplace characteristics influence psychological and cognitive responses in a virtual environment. *Ergonomics* just-accepted (2023), 1–22.

[15] Guillermo Bernal, Nelson Hidalgo, Conor Russomanno, and Pattie Maes. 2022. Galea: A physiological sensing system for behavioral research in Virtual Environments. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 66–76. https://doi.org/10.1109/VR51125.2022.00024

[16] Verena Biener, Snehanjali Kalamkar, Negar Nouri, Eyal Ofek, Michel Pahud, John J Dudley, Jinghui Hu, Per Ola Kristensson, Maheshya Weerasinghe, Klen Čopič Pucihar, et al. 2022. Quantifying the effects of working in VR for one week. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3810–3820.

[17] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A Robbins. 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in neuroinformatics* 9 (2015), 16. https://doi.org/10.3389/fninf.2015.00016

[18] Benjamin Blankertz, Michael Tangermann, Carmen Vidaurre, Siamac Fazli, Claudia Sannelli, Stefan Haufe, Cecilia Maeder, Lenny E Ramsey, Irene Sturm, Gabriel Curio, et al. 2010. The Berlin brain–computer interface: non-medical uses of BCI technology. *Frontiers in neuroscience* (2010), 198. https://doi.org/10.3389/fnins.2010.00198

[19] Andrew C Boud, David J Haniff, Chris Baber, and SJ Steiner. 1999. Virtual reality and augmented reality as a training tool for assembly tasks. In *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*. IEEE, New York, NY, USA, 32–36.

[20] Jhon Alexander Bueno-Vesga, Xinhao Xu, and Hao He. 2021. The effects of cognitive load on engagement in a virtual reality learning environment. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 645–652.

[21] Evan A Byrne and Raja Parasuraman. 1996. Psychophysiology and adaptive automation. *Biological psychology* 42, 3 (1996), 249–268.

[22] Benjamin T Carter and Steven G Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (2020), 49–62.

[23] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 282–297. https://doi.org/10.1145/3472749.3474750

[24] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards Understanding Diminished Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 549, 16 pages. https://doi.org/10.1145/3491102.3517452

[25] Aline Chevalier and Maud Kicka. 2006. Web designers and web users: Influence of the ergonomic quality of the web site on the information search. *International Journal of Human-Computer Studies* 64, 10 (2006), 1031–1048.

[26] Francesco Chiossi, Uwe Gruenefeld, Baosheng James Hou, Joshua Newn, Changkun Ou, Rulu Liao, Robin Welsch, and Sven Mayer. 2024. Understanding the Impact of the Reality-Virtuality Continuum on Visual Search Using Fixation-Related Potentials and Eye Tracking Features. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 281 (Sept. 2024), 33 pages. https://doi.org/10.1145/3676528

[27] Francesco Chiossi, Thomas Kosch, Luca Menghini, Steeven Villa, and Sven Mayer. 2023. SensCon: Embedding Physiological Sensing into Virtual Reality Controllers. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 223 (sep 2023), 32 pages. https://doi.org/10.1145/3604270

[28] Francesco Chiossi, Changkun Ou, Carolina Gerhardt, Felix Putze, and Sven Mayer. 2023. Designing and Evaluating an Adaptive Virtual Reality System using EEG Frequencies to Balance Internal and External Attention States. *arXiv preprint arXiv:2311.10447* (2023). https://doi.org/10.48550/arXiv.2311.10447

[29] Francesco Chiossi, Changkun Ou, and Sven Mayer. 2023. Exploring Physiological Correlates of Visual Complexity Adaptation: Insights from EDA, ECG, and EEG Data for Adaptation Evaluation in VR Adaptive Systems. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 118, 7 pages. https://doi.org/10.1145/3544549.3585624

[30] Francesco Chiossi, Ines Trautmannsheimer, Changkun Ou, Uwe Gruenefeld, and Sven Mayer. 2024. Searching Across Realities: Investigating ERPs and Eye-Tracking Correlates of Visual Search in Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–11. https://doi.org/10.1109/TVCG.2024.3456172

[31] Francesco Chiossi, Yagiz Turgut, Robin Welsch, and Sven Mayer. 2023. Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 196 (sep 2023), 26 pages. https://doi.org/10.1145/3604243

[32] Francesco Chiossi, Robin Welsch, Steeven Villa, Lewis Chuang, and Sven Mayer. 2022. Virtual Reality Adaptation Using Electrodermal Activity to Support the User Experience. *Big Data and Cognitive Computing* 6, 2 (2022), 19.

https://doi.org/10.3390/bdcc6020055

[33] Witold X Chmielewski, Moritz Mückschel, Ann-Kathrin Stock, and Christian Beste. 2015. The impact of mental workload on inhibitory control subprocesses. *NeuroImage* 112 (2015), 96–104.

[34] Witold X Chmielewski, Moritz Mückschel, Tjalf Ziemssen, and Christian Beste. 2017. The norepinephrine system affects specific neurophysiological subprocesses in the modulation of inhibitory control by working memory demands. *Human Brain Mapping* 38, 1 (2017), 68–81. https://doi.org/10.1002/hbm.23344

[35] Giorgia Cona, Francesco Chiossi, Silvia Di Tomasso, Giovanni Pellegrino, Francesco Piccione, Patrizia Bisiacchi, and Giorgio Arcara. 2020. Theta and alpha oscillations as signatures of internal and external attention to delayed intentions: A magnetoencephalography (MEG) study. *NeuroImage* 205 (2020), 116295. https://doi.org/10.1016/j.neuroimage.2019.116295

[36] Daniele De Massari, Daniel Pacheco, Rahim Malekshahi, Alberto Betella, Paul FMJ Verschure, Niels Birbaumer, and Andrea Caria. 2014. Fast mental states decoding in mixed reality. *Frontiers in behavioral neuroscience* 8 (2014), 415. https://doi.org/10.3389/fnbeh.2014.00415

[37] Frederic Dehais, Alban Dupres, Gianluca Di Flumeri, Kevin Verdiere, Gianluca Borghini, Fabio Babiloni, and Raphalle Roy. 2018. Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, New York, NY, USA, 544–549. https://doi.org/10.1109/SMC.2018.00102

[38] Vivek Dhakal, Anna Maria Feit, Per Ola Kristensson, and Antti Oulasvirta. 2018. Observations on Typing from 136 Million Keystrokes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174220

[39] Gabriel Dippel, Moritz Mückschel, Tjalf Ziemssen, and Christian Beste. 2017. Demands on response inhibition processes determine modulations of theta band activity in superior frontal areas and correlations with pupillometry–Implications for the norepinephrine system during inhibitory control. *NeuroImage* 157 (2017), 575–585. https://doi.org/10.1016/j.neuroimage.2017.06.037

[40] Kevin Doherty and Gavin Doherty. 2018. Engagement in HCI: conception, theory and measurement. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–39. https://doi.org/10.1145/3234149

[41] Jan Drewes, Sascha Feder, and Wolfgang Einhäuser. 2021. Gaze during locomotion in virtual reality and the real world. *Frontiers in Neuroscience* 15 (2021), 656913. https://doi.org/10.3389/fnins.2021.656913

[42] John J. Dudley, Keith Vertanen, and Per Ola Kristensson. 2018. Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion. *ACM Trans. Comput.-Hum. Interact.* 25, 6, Article 30 (dec 2018), 40 pages. https://doi.org/10.1145/3232163

[43] Maria K Eckstein, Belén Guerra-Carrillo, Alison T Miller Singley, and Silvia A Bunge. 2017. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental cognitive neuroscience* 25 (2017), 69–91.

[44] Stephen H Fairclough, Louise Venables, and Andrew Tattersall. 2005. The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology* 56, 2 (2005), 171–184.

[45] Johanna R Falk, Peter M Gollwitzer, Gabriele Oettingen, and Guido HE Gendolla. 2022. Task choice shields against incidental affective influences on effort-related cardiovascular response. *Psychophysiology* 59, 7 (2022), e14022.

[46] Nico Feld, Pauline Bimberg, Benjamin Weyers, and Daniel Zielasko. 2024. Simple and Efficient? Evaluation of Transitions for Task-Driven Cross-Reality Experiences. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[47] Daniel Immanuel Fink, Johannes Zagermann, Harald Reiterer, and Hans-Christian Jetter. 2022. Re-locations: Augmenting Personal and Shared Workspaces to Support Remote Collaboration in Incongruent Spaces. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (2022), 1–30.

[48] John J Foxe and Adam C Snyder. 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in psychology* 2 (2011), 154. https://doi.org/10.3389/fpsyg.2011.00154

[49] Jared A. Frank, Matthew Moorhead, and Vikram Kapila. 2016. Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New York, NY, USA, 302–307. https://doi.org/10.1109/ROMAN.2016.7745146

[50] Anthony G Gallagher, Neal E Seymour, Julie-Anne Jordan-Black, Brendan P Bunting, Kieran McGlade, and Richard Martin Satava. 2013. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition.

[51] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *Proceedings of the 2021 CHI Conference on*

Human Factors in Computing Systems (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 662, 14 pages. https://doi.org/10.1145/3411764.3445576

[52] Alan Gevins and Michael E Smith. 2003. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical issues in ergonomics science* 4, 1-2 (2003), 113–131. https://doi.org/10.1080/14639220210159717

[53] Mark S Gilzenrat, Sander Nieuwenhuis, Marieke Jepma, and Jonathan D Cohen. 2010. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience* 10, 2 (2010), 252–269.

[54] Ary Louis Goldberger and Emanuel Gold-berger. 1981. Clinical electrocardiography, a simplified approach. *Critical Care Medicine* 9, 12 (1981), 891–892.

[55] Guilherme Goncalves, Pedro Monteiro, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Systematic review on realism research methodologies on immersive virtual, augmented and mixed realities. *IEEE Access* 9 (2021), 89150–89161. https://doi.org/10.1109/ACCESS.2021.3089946

[56] Mar Gonzalez-Franco, Diar Abdlkarim, Arpit Bhatia, Stuart Macgregor, Jason Alexander Fotso-Puepi, Eric J Gonzalez, Hasti Seifi, Massimiliano Di Luca, and Karan Ahuja. 2024. Hovering Over the Key to Text Input in XR. *arXiv preprint arXiv:2406.09579* (2024). https://doi.org/10.48550/arXiv.2406.09579

[57] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, and Lauri Parkkonen. 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience* 7 (2013), 267. https://doi.org/10.3389/fnins.2013.00267

[58] Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Effects of hand representations for typing in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 151–158.

[59] Patrick Hamilton. 2002. Open source ECG analysis. In *Computers in Cardiology*. IEEE, Memphis, TN, USA, 101–104. https://doi.org/10.1109/CIC.2002.1166717

[60] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[61] Larry V Hedges. 2007. Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics* 32, 4 (2007), 341–370. https://doi.org/10.3102/1076998606298

[62] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1957–1967. https://doi.org/10.1145/2858036.2858134

[63] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine* 29, 6 (2012), 82–97.

[64] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, and Joost van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford University Press, Oxford, United Kingdom. 560 pages.

[65] Jesper F Hopstaken, Dimitri Van Der Linden, Arnold B Bakker, and Michiel AJ Kompier. 2015. The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological psychology* 110 (2015), 100–106.

[66] Lode Hoste and Beat Signer. 2013. SpeeG2: a speech- and gesture-based interface for efficient controller-free text input. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, Australia) *(ICMI '13)*. Association for Computing Machinery, New York, NY, USA, 213–220. https://doi.org/10.1145/2522848.2522861

[67] Ann Huang, Pascal Knierim, Francesco Chiossi, Lewis L Chuang, and Robin Welsch. 2022. Proxemics for Human-Agent Interaction in Augmented Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 421, 13 pages. https://doi.org/10.1145/3491102.3517593

[68] Sebastian Hubenschmid, Johannes Zagermann, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. 2023. ARound the Smartphone: Investigating the Effects of Virtually-Extended Display Size on Spatial Memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 527, 15 pages. https://doi.org/10.1145/3544548.3581438

[69] Darryl G Humphrey and Arthur F Kramer. 1994. Toward a psychophysiological assessment of dynamic changes in mental workload. *Human Factors* 36, 1 (1994), 3–26. https://doi.org/10.1177/001872089403600101

[70] Jan Jarvis, Felix Putze, Dominic Heger, and Tanja Schultz. 2011. Multimodal person independent recognition of workload related biosignal patterns. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. Association for Computing Machinery, New York, NY, USA, 205–208. https://doi.org/10.1145/2070481.2070516

[71] Ole Jensen and Ali Mazaheri. 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *In Human Neuroscience* 4 (2010), 186. https://doi.org/10.3389/fnhum.2010.00186

[72] Marieke Jepma and Sander Nieuwenhuis. 2011. Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of cognitive neuroscience* 23, 7 (2011), 1587–1596.

[73] Christoph Albert Johns, João Marcelo Evangelista Belo, Clemens Nylandsted Klokmose, and Ken Pfeuffer. 2023. Pareto Optimal Layouts for Adaptive Mixed Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 223, 7 pages. https://doi.org/10.1145/3544549.3585732

[74] Sorelle Audrey Kamkuimo K, Benoît Girard, and Bob-Antoine J Menelas. 2020. Dynamic Difficulty Adjustment Through real-time physiological feedback for a more adapted Virtual Reality Exposure Therapy. In *International Conference on Games and Learning Alliance*. Springer, Springer, Cham, 102–111.

[75] Alexander J Karran, Théophile Demazure, Pierre-Majorique Leger, Elise Labonte-LeMoyne, Sylvain Senecal, Marc Fredette, and Gilbert Babin. 2019. Toward a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS. *Frontiers in Human Neuroscience* 13 (2019), 393. https://doi.org/10.3389/fnhum.2019.00393

[76] Hannes Kaufmann, Dieter Schmalstieg, and Michael Wagner. 2000. Construct3D: a virtual reality application for mathematics and geometry education. *Education and information technologies* 5 (2000), 263–276.

[77] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235.

[78] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173919

[79] Jens Kohlmorgen, Guido Dornhege, Mikio Braun, Benjamin Blankertz, Klaus-Robert Müller, Gabriel Curio, Konrad Hagemann, Andreas Bruns, Michael Schrauf, Wilhelm Kincses, et al. 2007. Improving human performance in a real operating environment through real-time mental workload detection. *Toward brain-computer interfacing* 409422 (2007), 409–422.

[80] Thomas Kosch, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. *ACM Comput. Surv.* 55, 13s, Article 283 (July 2023), 39 pages. https://doi.org/10.1145/3582272

[81] Thomas Kosch, Albrecht Schmidt, Simon Thanheiser, and Lewis L. Chuang. 2020. One Does Not Simply RSVP: Mental Workload to Select Speed Reading Parameters Using Electroencephalography. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376766

[82] George Alex Koulieris, Kaan Akşit, Michael Stengel, Rafał K Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-eye display and tracking technologies for virtual and augmented reality. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, Hoboken, NJ, USA, 493–519.

[83] Matthias Kraus, Johannes Fuchs, Björn Sommer, Karsten Klein, Ulrich Engelke, Daniel Keim, and Falk Schreiber. 2022. Immersive analytics with abstract 3D visualizations: A survey. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, Hoboken, NJ, USA, 201–229.

[84] Per Ola Kristensson and Keith Vertanen. 2012. Performance comparisons of phrase sets and presentation styles for text entry evaluations. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (Lisbon, Portugal) *(IUI '12)*. Association for Computing Machinery, New York, NY, USA, 29–32. https://doi.org/10.1145/2166966.2166972

[85] Robert Kurzban, Angela Duckworth, Joseph W Kable, and Justus Myers. 2013. An opportunity cost model of subjective effort and task performance. *Behavioral and brain sciences* 36, 6 (2013), 661–679.

[86] Te-Won Lee, Mark Girolami, and Terrence J Sejnowski. 1999. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation* 11, 2 (1999), 417–441.

[87] David Lindlbauer. 2022. The future of mixed reality is adaptive. *XRDS: Crossroads, The ACM Magazine for Students* 29, 1 (2022), 26–31. https://doi.org/10.1145/3558191

[88] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 147–160. https://doi.org/10.1145/3332165.3347945

[89] Xingyu Long, Sven Mayer, and Francesco Chiossi. 2024. Multimodal Detection of External and Internal Attention in Virtual Reality using EEG and Eye Tracking Features. In *Proceedings of Mensch Und Computer 2024* (Karlsruhe, Germany) *(MuC '24)*. Association for Computing Machinery, New York, NY, USA, 29–43. https://doi.org/10.1145/3670653.3670657

[90] 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* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 815–825. https://doi.org/10.1145/3472749.3474788

[91] Rasmus S Lunding, Mathias N Lystbæk, Tiare Feuchtner, and Kaj Grønbæk. 2023. AR-supported Human-Robot Collaboration: Facilitating Workspace Awareness and Parallelized Assembly Tasks. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, New York, NY, USA, 1064–1073.

[92] Blair MacIntyre, Elizabeth D. Mynatt, Stephen Voida, Klaus M. Hansen, Joe Tullio, and Gregory M. Corso. 2001. Support for multitasking and background awareness using interactive peripheral displays. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (Orlando, Florida) *(UIST '01)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/502348.502355

[93] Paul P. Maglio and Christopher S. Campbell. 2000. Tradeoffs in displaying peripheral information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 241–248. https://doi.org/10.1145/332040.332438

[94] Elisa Magosso, Francesca De Crescenzio, Giulia Ricci, Sergio Piastra, and Mauro Ursino. 2019. EEG alpha power is modulated by attentional changes during cognitive tasks and virtual reality immersion. *Comp. Intelligence and Neuroscience* (2019). https://doi.org/10.1155/2019/7051079

[95] Elisa Magosso, Giulia Ricci, and Mauro Ursino. 2021. Alpha and theta mechanisms operating in internal-external attention competition. *Journal of Integrative Neuroscience* 20, 1 (2021), 1–19. https://doi.org/10.31083/j.jin.2021.01.422

[96] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods* 53, 4 (2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[97] Sotiris Makris, Panagiotis Karagiannis, Spyridon Koukas, and Aleksandros-Stereos Matthaiakis. 2016. Augmented reality system for operator support in human–robot collaborative assembly. *CIRP Annals* 65, 1 (2016), 61–64.

[98] Sebastian Marwecki, Andrew D. Wilson, Eyal Ofek, Mar Gonzalez Franco, and Christian Holz. 2019. Mise-Unseen: Using Eye Tracking to Hide Virtual Reality Scene Changes in Plain Sight. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 777–789. https://doi.org/10.1145/3332165.3347919

[99] Nathan Matsuda, Alex Chapiro, Yang Zhao, Clinton Smith, Romain Bachy, and Douglas Lanman. 2022. Realistic Luminance in VR. In *SIGGRAPH Asia 2022 Conference Papers* (Daegu, Republic of Korea) *(SA '22)*. Association for Computing Machinery, New York, NY, USA, Article 21, 8 pages. https://doi.org/10.1145/3550469.3555427

[100] Gerald Matthews, Joel S Warm, Lauren E Reinerman, Lisa K Langheim, and Dyani J Saxby. 2010. Task engagement, attention, and executive control. *Handbook of individual differences in cognition: Attention, memory, and executive control* (2010), 205–230. https://doi.org/10.1007/978-1-4419-1210-7_13

[101] Elisabeth Mayer, Francesco Chiossi, and Sven Mayer. 2024. Crossing Mixed Realities: A Review for Transitional Interfaces Design. In *Proceedings of Mensch Und Computer 2024* (Karlsruhe, Germany) *(MuC '24)*. Association for Computing Machinery, New York, NY, USA, 629–634. https://doi.org/10.1145/3670653.3677481

[102] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2143–2152. https://doi.org/10.1145/2702123.2702382

[103] Mark McGill, Stephen Brewster, Daniel Pires De Sa Medeiros, Sidney Bovet, Mario Gutierrez, and Aidan Kehoe. 2022. Creating and Augmenting Keyboards for Extended Reality with the Keyboard Augmentation Toolkit. *ACM Trans. Comput.-Hum. Interact.* 29, 2, Article 15 (jan 2022), 39 pages. https://doi.org/10.1145/3490495

[104] Mark McGill, Stephen Brewster, Daniel Pires De Sa Medeiros, Sidney Bovet, Mario Gutierrez, and Aidan Kehoe. 2022. Creating and Augmenting Keyboards for Extended Reality with the Keyboard Augmentation Toolkit. *ACM Trans. Comput.-Hum. Interact.* 29, 2, Article 15 (jan 2022), 39 pages. https://doi.org/10.1145/3490495

[105] Mark Mcgill, Aidan Kehoe, Euan Freeman, and Stephen Brewster. 2020. Expanding the Bounds of Seated Virtual Workspaces. *ACM Trans. Comput.-Hum. Interact.* 27, 3, Article 13 (May 2020), 40 pages. https://doi.org/10.1145/3380959

[106] Leonel Merino, Magdalena Schwarzl, Matthias Kraus, Michael Sedlmair, Dieter Schmalstieg, and Daniel Weiskopf. 2020. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, New York, NY, USA, 438–451. https://doi.org/10.1109/ISMAR50242.2020.00069

[107] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[108] Peter R Murphy, Ian H Robertson, Joshua H Balsters, and Redmond G O'connell. 2011. Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology* 48, 11 (2011), 1532–1543.

[109] Shivsevak Negi and Ritayan Mitra. 2018. EEG Metrics to Determine Cognitive Load and Affective States: A Pilot Study. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 182–185. https://doi.org/10.1145/3267305.3267618

[110] Peter Nickel and Friedhelm Nachreiner. 2003. Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human factors* 45, 4 (2003), 575–590. https://doi.org/10.1518/hfes.45.4.575.27094

[111] Heather L O'Brien and Elaine G Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology* 59, 6 (2008), 938–955. https://doi.org/10.1002/asi.20801

[112] Fred Paas, Alexander Renkl, and John Sweller. 2003. Cognitive load theory and instructional design: Recent developments. *Educational psychologist* 38, 1 (2003), 1–4. https://doi.org/10.1207/S15326985EP3801_1

[113] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 141–144. https://doi.org/10.1145/1743666.1743701

[114] Max Pascher, Uwe Gruenefeld, Stefan Schneegass, and Jens Gerken. 2023. How to Communicate Robot Motion Intent: A Scoping Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 409, 17 pages. https://doi.org/10.1145/3544548.3580857

[115] 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* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 21, 11 pages. https://doi.org/10.1145/3359996.3364265

[116] Sebastian Pick, Andrew S Puika, and Torsten W Kuhlen. 2016. SWIFTER: Design and evaluation of a speech-based text input metaphor for immersive virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, New York, NY, USA, 109–112. https://doi.org/10.1109/3DUI.2016.7460039

[117] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. 2019. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* 198 (2019), 181–197. https://doi.org/10.21105/joss.04484

[118] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3132787.3139200

[119] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI* 6 (2019), 5. https://doi.org/10.3389/frobt.2019.00005

[120] Henning Pohl and Roderick Murray-Smith. 2013. Focused and casual interactions: allowing users to vary their level of engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2223–2232. https://doi.org/10.1145/2470654.2481307

[121] Alan T Pope, Edward H Bogart, and Debbie S Bartolome. 1995. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology* 40, 1-2 (1995), 187–195. https://doi.org/10.1016/0301-0511(95)05116-3

[122] Lawrence J Prinzel III, Raja Parasuraman, Frederick G Freeman, Mark W Scerbo, Peter J Mikulka, and Alan T Pope. 2003. *Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centered adaptive automation design.* Technical Report. NASA.

[123] Sébastien Puma, Nadine Matton, Pierre-V. Paubel, Éric Raufaste, and Radouane El-Yagoubi. 2018. Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology* 123 (2018), 111–120. https://doi.org/10.1016/j.ijpsycho.2017.10.004

[124] Felix Putze and Hendrik Eilts. 2023. Analyzing the Importance of EEG Channels for Internal and External Attention Detection. In *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, New York, NY, USA, 4752–4757. https://doi.org/10.1109/SMC53992.2023.10394322

[125] Felix Putze, Dennis Weiß, Lisa-Marie Vortmann, and Tanja Schultz. 2019. Augmented reality interface for smart home control using SSVEP-BCI and eye gaze. In *2019 IEEE international conference on systems, man and cybernetics (smc)*. IEEE, New York, NY, USA, 2812–2817.

[126] Eric D Ragan, Doug A Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P McMahan. 2015. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics* 21, 7 (2015), 794–807.

[127] Troels Rasmussen, Tiare Feuchtner, Weidong Huang, and Kaj Grønbæk. 2022. Supporting workspace awareness in remote assistance through a flexible multi-camera system and Augmented Reality awareness cues. *Journal of Visual Communication and Image Representation* 89 (2022), 103655.

[128] Bujar Raufi and Luca Longo. 2022. An Evaluation of the EEG alpha-to-theta and theta-to-alpha band Ratios as Indexes of Mental Workload. *Frontiers in Neuroinformatics* 16 (2022), 44. https://doi.org/10.3389/fninf.2022.861967

[129] Jacob Reimer, Matthew J McGinley, Yang Liu, Charles Rodenkirch, Qi Wang, David A McCormick, and Andreas S Tolias. 2016. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature communications* 7, 1 (2016), 13289.

[130] Anke Verena Reinschluessel, Thomas Muender, Roland Fischer, Valentin Kraft, Verena Nicole Uslar, Dirk Weyhe, Andrea Schenk, Gabriel Zachmann, Tanja Döring, and Rainer Malaka. 2023. Versatile Immersive Virtual and Augmented Tangible OR – Using VR, AR and Tangibles to Support Surgical Practice. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 477, 5 pages. https://doi.org/10.1145/3544549.3583895

[131] Giulia Ricci, Francesca De Crescenzio, Sandhya Santhosh, Elisa Magosso, and Mauro Ursino. 2022. Relationship between electroencephalographic data and comfort perception captured in a Virtual Reality design environment of an aircraft cabin. *Scientific Reports* 12, 1 (2022), 10938.

[132] Radiah Rivu, Ville Mäkelä, Mariam Hassib, Yomna Abdelrahman, and Florian Alt. 2021. Exploring How Saliency Affects Attention in Virtual Reality. In *IFIP Conference on Human-Computer Interaction*. Springer, Cham, Switzerland, 147–155. https://doi.org/10.1007/978-3-030-85607-6_10

[133] Paul Sauseng, Wolfgang Klimesch, Manuel Schabus, and Michael Doppelmayr. 2005. Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International journal of Psychophysiology* 57, 2 (2005), 97–103. https://doi.org/10.1016/j.ijpsycho.2005.03.018

[134] Dieter Schmalstieg, Anton Fuhrmann, Gerd Hesina, Zsolt Szalavári, L Miguel Encarnaçao, Michael Gervautz, and Werner Purgathofer. 2002. The studierstube augmented reality project. *Presence: Teleoperators & Virtual Environments* 11, 1 (2002), 33–54.

[135] Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal technologies* 4 (2000), 191–199.

[136] Sofia Seinfeld, Tiare Feuchtner, Antonella Maselli, and Jörg Müller. 2021. User representations in human-computer interaction. *Human–Computer Interaction* 36, 5-6 (2021), 400–438.

[137] Richard Skarbez, Missie Smith, and Mary C Whitton. 2021. Revisiting Milgram and Kishino's reality-virtuality continuum. *Frontiers in Virtual Reality* 2 (2021), 647997.

[138] Alexis D Souchet, Mamadou Lamarana Diallo, and Domitile Lourdeaux. 2022. Cognitive load Classification with a Stroop task in Virtual Reality based on Physiological data. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, New York, NY, USA, 656–666.

[139] R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 113–120. https://doi.org/10.1145/642611.642632

[140] 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* (, Montreal QC, Canada,) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174221

[141] Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3643–3652. https://doi.org/10.1145/2702123.2702136

[142] Paul Streli, Jiaxi Jiang, Andreas Rene Fender, Manuel Meier, Hugo Romat, and Christian Holz. 2022. TapType: Ten-Finger Text Entry on Everyday Surfaces via Bayesian Inference. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 497, 16 pages. https://doi.org/10.1145/3491102.3501878

[143] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/642611.642626

[144] Maarten van Dantzich, Daniel Robbins, Eric Horvitz, and Mary Czerwinski. 2002. Scope: providing awareness of multiple notifications at a glance. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Trento, Italy) *(AVI '02)*. Association for Computing Machinery, New York, NY, USA, 267–281. https://doi.org/10.1145/1556262.1556306

[145] Pauline van der Wel and Henk Van Steenbergen. 2018. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic bulletin & review* 25 (2018), 2005–2015.

[146] Akos Vetek and Saija Lemmelä. 2011. Could a dialog save your life? analyzing the effects of speech interaction strategies while driving. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. Association for Computing Machinery, New York, NY, USA, 145–152. https://doi.org/10.1145/2070481.2070506

[147] Lisa-Marie Vortmann and Felix Putze. 2021. Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 80 (June 2021), 27 pages. https://doi.org/10.1145/3463507

[148] Rui I. Wang, Brandon Pelfrey, Andrew T. Duchowski, and Donald H. House. 2012. Online Gaze Disparity via Bioncular Eye Tracking on Stereoscopic Displays. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*. IEEE, New York, NY, USA, 184–191. https://doi.org/10.1109/3DIMPVT.2012.37

[149] Glenn F Wilson and F Thomas Eggemeier. 1991. Psychophysiological assessment of workload in multi-task environments. *Multiple-task performance* 329360 (1991), 34.

[150] Frederik Winther, Linoj Ravindran, Kasper Paabøl Svendsen, and Tiare Feuchtner. 2020. Design and Evaluation of a VR Training Simulation for Pump Maintenance Based on a Use Case at Grundfos. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, New York, NY, USA, 738–746. https://doi.org/10.1109/VR46266.2020.00097

[151] Jie Yin, Nastaran Arfaei, Piers MacNaughton, Paul J Catalano, Joseph G Allen, and John D Spengler. 2019. Effects of biophilic interventions in office on stress reaction and cognitive function: A randomized crossover study in virtual reality. *Indoor air* 29, 6 (2019), 1028–1039. https://doi.org/10.1111/ina.12593

[152] Sijia Zhao, Maria Chait, Fred Dick, Peter Dayan, Shigeto Furukawa, and Hsin-I Liao. 2019. Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nature Communications* 10, 1 (2019), 4030. https://doi.org/10.1038/s41467-019-12048-1

[153] Andrey D Zhdanov, Dmitry D Zhdanov, Nikolai Nikolaevich Bogdanov, Igor Stanislavovich Potemin, Vladimir A Galaktionov, and Maxim I Sorokin. 2019. Discomfort of visual perception in virtual and mixed reality systems. *Programming and Computer Software* 45 (2019), 147–155. https://doi.org/10.1134/S036176881904011X