Adaptive Training on Basic AR Interactions: Bi-Variate Metrics and Neuroergonomic Evaluation Paradigms

Shantanu Vyas^a, Shivangi Dwivedi^b, Lindsey J. Brenner^b, Isabella Pedron^b, Joseph L. Gabbard^c, Vinayak R. Krishnamurthy^a, Ranjana K. Mehta^d

^aJ. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, Texas; ^bWm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, College Station, Texas; ^c Grado Department of Industrial & Systems Engineering, Virginia Tech, Blacksburg, Virginia; ^d Department of Industrial & Systems Engineering, University of Wisconsin Madison, Madison, Wisconsin

ARTICLE HISTORY

Compiled September 4, 2023

ABSTRACT

Augmented Reality (AR) training is a cost-effective and safe alternative to traditional instructional methods. However, training novices in basic mid-air AR interactions remains challenging. To address this, we aimed to: (a) develop a robust metric to evaluate user performance across different AR interaction techniques and develop adaptation models to predict additional training requirements; (b) evaluate the adaptation models using a neuroergonomics approach. We conduct a two-phase study during which, novice participants perform simple AR interactions: poking and raycasting. In Phase-I, twenty-seven participants' data is used to identify a bi-variate performance metric based on median completion time and consistency. Unsupervised models are trained using this metric to classify participants as low/high performers. In Phase-II, we evaluate the models on twenty-one new participants and analyze the differences in performance, neural activity and heart-rate variability between low/high performers. Our study showcases the effectiveness of our models and further discusses the potential of integrating neuroergonomics for advanced AR-based training applications.

KEYWORDS

Augmented Reality; Adaptive Training; Neurophysiological Analysis

1. Introduction

Augmented reality (AR)-based systems have seen widespread usage over the last few decades due to improvements in hardware and software technologies (Billinghurst, Clark, Lee, et al., 2015). Combining the real and virtual worlds provides unique benefits that have encouraged the integration of AR into various fields, ranging from manufacturing and construction industries (Bottani & Vignali, 2019), to surgery and emergency response (Vávra et al., 2017). A major application of AR in these fields, more specifically, with AR immersive headsets, has been training personnel. Immersing users in complex environments where they can interact with digital objects has led to the development of training applications that can reduce costs, time, and dan-

Corresponding Author: Ranjana K. Mehta. Email: ranjana.mehta@wisc.edu

gers associated with training tasks (Barsom, Graafland, & Schijven, 2016). This is especially true in fields such as emergency response, where training new personnel may be dangerous and require expensive resources (Buckman, 2005; Evarts & Molis, 2018). However, like any emerging technology, AR has its own set of challenges, namely the need for better display technology, richer and more intuitive interactions, improved tracking systems, and technology acceptance (Billinghurst, 2021). In this study, we will focus on the challenge of training users on interactions specific to AR environments that might cause technological barriers to entry.

Gibson's assertion, "perception is for action" (Gibson, 2014), has influenced the design of interactions in computational support systems that align with the confluence of the perception-action-cognition cycle that is inherent in human manipulations. This is particularly relevant in virtual environments where users engage with elements both within and beyond their physical reach (Bowman, 1999). Proximity to the action space is a key consideration in designing interactions that enable effective engagement. However, some AR-based system interactions do not necessarily mimic how humans interact with real-world objects (Billinghurst, 2021) and can feel unnatural to users. While there is continual research on tangible and multi-modal interfaces for AR to make them feel more natural (Mohanty & Krishnamurthy, 2021; Nizam et al., 2018), these interfaces are yet to be fully integrated into fielded applications. As such, users with no prior experience with AR-based systems may need to familiarize themselves with basic interactions in AR like selecting and moving digital objects, before being able to complete more complex tasks like training. In the absence of such "familiarization", adoption of AR applications may be slow, or worse, fail.

1.1. Adaptive Training & Neuroergonomic Evaluation

Training applications in AR are typically domain-specific, such as medical training (Barsom et al., 2016) and assembly (Westerfield, Mitrovic, & Billinghurst, 2015), with a lack of analysis on how novice users' familiarity with basic AR interactions affects the domain-specific tasks. Thus, research is needed that focuses on developing training and evaluation paradigms for accelerating learning on basic AR interactions. This will enable users to effectively interact with domain-specific AR functionalities. A robust training or tutoring system will also have to adapt to a wide range of users with different learning abilities. As a result, having an adaptive training system is preferred over a non-adaptive training system (Peretz et al., 2011). More generally, adaptive training can be defined as a training system where the problem, stimulus, or task can be varied depending on the trainees' performance (Kelley, 1969). Some user-based metrics that help predict the need for adaptation stem from a trainee's raw performance data, their prior expertise, and their working memory capacity (G. Huang et al., 2021). In AR, adaptive training systems have mainly been used for application-specific tasks. For instance, Westerfield et al., (Westerfield et al., 2015) developed an intelligent tutoring system (ITS) for an AR training application on motherboard assembly. The application provided feedback messages to trainees depending on their actions during the assembly task. Similarly, Huang et al. (G. Huang et al., 2021), developed an adaptive tutoring system for machine tasks where the system continually monitored the users' states (e.g., their observations and interactions) and machine component states (e.g., what buttons of the machines were active). The training system then made predictions on both of the future states to decide whether the user required more information to improve their learning outcome. Militello et al. (Militello, Sushereba, Hernandez,

& Patterson, 2019), also provided principles for designing adaptive training in AR that add haptic aspects to visual training, determining the cognitive states of users to adaptively change tasks to reach the desired states, and make simulations appear more realistic. A more personalized approach has also been used where trainees are awarded points based on their performance in order to improve their engagement in the training process (Albayrak, Öner, Atakli, & Ekenel, 2019; Ma et al., 2016).

While training applications, both in the traditional and AR settings, focus on performance-based metrics to evaluate individuals, prior works have shown that assessing cognitive load (e.g., intrinsic, germane, and extraneous load) and mental workload can provide a better understanding of an individual's performance (Biddle & Buck, 2019; Tao et al., 2020). It is shown that the cognitive load and mental workload of individuals can affect their ability to learn and recall information (Souchet, Philippe, Lourdeaux, & Leroy, 2022; Steed, Pan, Zisch, & Steptoe, 2016) from different mediums, such as screens and AR headsets. Typically surveys such as the Cognitive Load Theory (Klepsch, Schmitz, & Seufert, 2017) and NASA Task Load Index (TLX) (Hart, 2006) are used to collect these metrics. Additionally, physiological measures such as heart rate and heart rate variability (HRV) have also been used as reliable measures to assess mental effort (Mukherjee, Yadav, Yung, Zajdel, & Oken, 2011) with novel interfaces (Dwivedi et al., 2022) and to differentiate between high and low performers in cognitive tasks such as playing chess, where high performers had significantly higher temporal HRV (Fuentes-García et al., 2019). Studies also show increased cardiovascular load (higher heart rate and lower temporal HRV) among low performers in VR-based training environments (Haves et al., 2022). More recently, neural measures have been used to quantify cognitive load and mental workload, where increased brain activity corresponds to high cognitive load and mental workload (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014; Bunce et al., 2011; Kesedžić, Šarlija, Božek, Popović, & Ćosić, 2021). Neurophysiological measures like brain imaging using functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG) are also used to differentiate between high and low performers during training, where high performers exhibit higher task-related brain activation patterns (Abujelala, Karthikeyan, Tyagi, Du, & Mehta, 2021; Shi, Zhu, Mehta, & Du, 2020; Walia et al., 2022), and where VR-based training under intense stress is associated with reorganization of compensatory brain networks (Tyagi et al., 2021). Therefore, neurophysiological metrics, such as brain activation and connectivity patterns and HRV responses, may be useful in evaluating performance in a variety of tasks, including AR-based training tasks.

One advantage of AR is that it enables training on psychomotor tasks (LaViola et al., 2015). In fact, performance on psychomotor tasks has been found to improve and be more accurate in real-world tasks when users were trained in an AR environment (Henderson & Feiner, 2011). The impact of the cognitive demand of psychomotor tasks on an individual's performance has also been measured by analyzing region-based brain activation and brain functional connectivity (Gevins & Smith, 2003). Changes in activity in different brain regions such as the premotor cortex, supplementary motor area, primary motor, and frontal eye field/cingulate gyrus can indicate different measures of cognitive demand of psychomotor tasks (Micheletti, 1996). Additionally, increased connectivity between functionally different brain regions signals neural efficiency (Tomasi, Wang, & Volkow, 2013). However, whether AR environments alter psychomotor learning and associated neural dynamics remains unexplored. Given that AR environments are associated with distinct changes in cognitive load, capturing brain activity and connectivity can shed light on the neural processes underlying psychomotor learning in AR.

1.2. Personalized Adaptive Training & its Challenges

Recent emphasis on personalization (Mehta et al., 2022) has led to efforts focusing on developing state-based adaptive-training models in AR/VR that incorporate real-time neurophysiological feedback based on the cognitive load experienced by the trainee (Dey, Chatburn, & Billinghurst, 2019; Doswell & Skinner, 2014). The models use heart rate and brain activity to measure trainee cognitive load as they perform the tasks and accordingly decide the tasks' difficulty. However, these studies focus on developing state-based adaptations to facilitate learning of complex tasks that require some prior knowledge on the topic, such as psychomotor surgical training. There has been little to no prior work in the development of performance-based adaptive training techniques that help participants become proficient in fundamental AR tasks, while simultaneously studying the effect of such strategies on their neurophysiological activities. A major challenge here is the lack of guidance on defining proficiency in AR interactions and the lack of transparency in reporting how familiarization with AR interactions impacts subsequent learning in complex domain-specific applications. Examining neurophysiological processes associated with learning may also help ascertain if specific adaptive training on AR interactions is effective for a user's learning process. Given the lack of appropriate and relevant metrics to assess learning of AR interactions, neurophysiological examinations can help assess the utility of such metrics.

1.3. Bi-variate Metric & Two-phase Study Approach

The present study is aimed at providing adaptive training in fundamental AR-based interactions while evaluating its effect on performance and associated neurophysiological responses in learners. We approach this problem by conducting a two-phase study, wherein Phase-I was used to collect performance-based data from novice users performing simple 'selection' tasks using different AR interactions (e.g., poking and raycasting). The aim of Phase-I was to identify performance metrics that could be used to develop and train an adaptation model for identifying low and high-performing individuals. During Phase-II, the adaptation model was tested on novice participants to assess its effectiveness. To this end, we collected neurophysiological data across a training task and an evaluation task. The impact of adaptation on neurophysiological markers of learning were also examined by recording changes in hemodynamic responses from the fronto-motor brain region (i.e., supplementary motor area, premotor cortex, primary motor area, and frontal eye field) using fNIRS, alongside performance data. We hypothesized that neurophysiological changes would occur with task repetition, i.e., strengthening neural connections with performance improvement (Jensen, 2005). We also hypothesized that there would be increased activation and stronger functional connectivity with improved performance (Ghilardi et al., 2000; Micheletti, 1996; Seidler, Noll, & Thiers, 2004). Additionally, we collected heart rate variability as another physiological measure and hypothesized that improvement in performance would result in lower physiological load (Fuentes-García et al., 2019; Hayes et al., 2022; Suriya-Prakash, John-Preetham, & Sharma, 2015). As a result of our two-phase study, we made three key contributions in this paper. First, we develop a multi-dimensional (specifically bi-variate) metric based on time and consistency to evaluate user performance on basic AR interactions. Secondly, we develop an unsupervised adaptation model based on the metric to identify users who require additional training in different interaction techniques. Finally, we show the efficacy of our model through a systematic neurophysiological analysis of the classification capabilities of the adaptation models,

highlighting the differences in functional connectivity, brain activation, heart-rate variability and subjective responses between low and high performing individuals.

[Figure 1 here]

2. General Methods

2.1. Types of AR Interactions

In this work, we focused on the fundamental task of "selection" in AR environments for head-mounted displays (HMD). The "selection" task enables users to interact with digital objects such as buttons and keyboards and is a fundamental building block of many AR applications (Bowman, 1999; Xu, Liang, He, & Wang, 2019). Learning the interaction skills involved in the "selection" task can have a broader impact on a user's overall AR interaction skills, thereby providing a foundation for more complex AR tasks (Looser, Billinghurst, Grasset, & Cockburn, 2007). There are various interaction techniques available to accomplish the "selection" task in AR environments depending on the proximity of digital objects to the user. The 'poking interaction' is used when objects are within arm's reach of the user and involve selecting the virtual object by directly "poking" it. Alternatively, the "raycasting interaction" (or distal pointing) is used when elements are out of arm's reach, and selecting an object involves casting a ray from the middle of the palm onto an object, and then selecting the object by pinching the index finger and thumb. Therefore, familiarizing a user with both of these interaction types is important to learn the 'selection' task in the AR environment.

2.2. Performance Metrics

With the primary goal of adaptively training users on AR interactions (poking and raycasting), it was essential for us to identify users who needed additional training, which has typically been dependent on the context of the task at hand. For instance, Huang et. al. (G. Huang et al., 2021), monitored the state of the machines that learners interacted with to gauge user performance during machine tasks. Similarly, in tasks involving casualty care, the speed with which the learners make clinical decisions can be an indicator for adaptation (Tanaka, Craighead, Taylor, & Sottilare, 2019). As our focus was on the "selection" task in AR, we wanted to identify a metric that could be easily adopted in many different contexts. As such, we defined a bivariate performance metric based on two variables: time and consistency.

The *time variable* corresponded to the median time to complete a unit interaction across all trials. This variable helped us understand how fast the user was able to complete a task using a particular interaction type. Selecting the median time over mean time allowed us to mitigate the effect of outliers that are typically observed in the first few training trials while learning new skills (Anderson, Grossman, Matejka, & Fitzmaurice, 2013). Users who performed the task more quickly were expected to have a lower median completion time compared to slower users. However, only considering the overall median time limited our understanding of the users' learning rates which is better analyzed through the variability across trials (Kast & Leukel, 2016; Wu, Miyamoto, Castro, Ölveczky, & Smith, 2014).

The *consistency variable*, therefore, accounted for the users' change in performance across trials and allowed us to quantifiably assess the change in the users' learning behavior. Prior works have highlighted the importance of analyzing consistency as a way to understand learning behavior in domains such as online courses and virtual learning environments (Młynarska, Greene, & Cunningham, 2016; Sher, Hatala, & Gašević, 2020; Zhou & Bhat, 2021). Taking inspiration from these works, we measured consistency by fitting a quadratic polynomial curve on the users' median times per trial, and then calculating the deviation between the fit curve and the median times per trial using root mean square error (RMSE). A quadratic polynomial curve was chosen over higher degrees to prevent overfitting on users' data. Users who were actively learning across all trials would exhibit more variability and therefore, less consistency compared to users who learnt the skills in the first few trials.

The combination of the time and consistency variables effectively characterized user performance for the given task. A low median completion time but a high RMSE value suggested that the user was still learning and may have completed some trials quickly. Conversely, if their median time was high but their RMSE was low, it indicated that the participant was consistently performing poorly. A high-performing user, however, would exhibit low median time and low RMSE value, indicating that they were able to learn the skills in the given trials.

2.3. Two-Phase Study

The bivariate performance metric helped us evaluate individual users across different interaction types. However, to identify users who needed additional training (low performers) and those who did not (high performers), we needed a baseline to which we could compare each user's metric values. Our goal in this work was to help novices learn and familiarize themselves with the different AR interactions, and not necessarily master them. Comparing them to experts would almost always result in all users requiring additional training. Instead, comparing novice users' performance across themselves can help us identify those who learned the interaction well and those who had not. To achieve this, we conducted two studies, where the Phase-I study involved collecting data on novice users to develop an adaptation model and identify the baseline performance metrics, and the Phase-II study involved testing the adaptation model on new novice users and analyzing the model's effectiveness with respect to the performance metrics, and neural and physiological dynamics of learning.

3. Phase I Study

3.1. Participants

Twenty-seven participants (twelve female; fifteen male) were recruited with a mean age of 30.96 (SD = 15.6) years. Of these, sixteen were from the local student population, while the others were individuals with a background in emergency response. All participants self-reported that they were right-handed and had less than one hour of AR and/or VR experience prior to participating in the study. All procedures were approved by the university's Institutional Review Board and complied with the American Psychological Association Code of Ethics. All participants were equally compensated with \$30 gift cards for their involvement.

3.2. Protocol

Following informed consent, participants completed background questionnaires covering demographic information and the NASA Task Load Index (TLX). Participants then donned an AR-HMD (HoloLens 2, Microsoft Corporation, USA) and the study tasks began. Training: Participants performed two tasks, one corresponding to each of the poking or raycasting interaction. The tasks were ordered in an increasing order of difficulty, where all participants completed the poking interaction task first, followed by the more difficult raycasting interaction task (Argelaguet & Andujar, 2013; Kopper, Bowman, Silva, & McMahan, 2010). For the poking interaction task, participants entered an AR application that displayed ten sequentially numbered buttons in a random order placed within their arm's reach. Participants were instructed to select the buttons in ascending order from 1-10 by poking them with their fingers as quickly and accurately as they could (Figure 1a). Following the poking interaction task, participants entered a similar application for the raycasting interaction task, with the buttons located away from their arm's reach. They used the raycasting approach to sequentially select the buttons in ascending order (Figure 1a). Participants completed eight trials for each of the tasks, where the order of the numbered buttons were randomized after each trial. Following each task, they were asked to complete the NASA TLX (Hart, 2006) and Cognitive Load Theory (CLT) surveys (Klepsch et al., 2017) for the subjective assessment of their workload and cognitive load experienced during tasks. Additionally, participants were queried regarding their subjective experience of fatigue and the underlying factors contributing to it. Participants concluded the study with User Engagement Scale (O'Brien, Cairns, & Hall, 2018) to measure differences in engagement across training and evaluation in an AR environment. The entire protocol took approximately 2 hours to complete.

3.3. Training the Adaptation Models

The temporal performance data collected for each participant served as the training data for developing adaptation models for both of the interaction types. The challenge in developing such models was the lack of prior empirical benchmarks or heuristics that could serve as ground truth references. As such, the models would first need to identify the baseline values for the performance metrics and then classify participants as low and high performers (Figure 1). The problem, therefore, became that of an unsupervised clustering problem resulting in the learning of two clusters, i.e., clusters corresponding to low performers and high performers. We opted to use the k-means algorithm for our adaptation model due to the low complexity of the training data and simple representation of the model output in the form of cluster centroids (Saxena et al., 2017). We first standardized the two features (i.e., time and consistency values) by removing the mean and scaling it to unit variance. Next, we used the scaled data to train the k-means model to compute the two centroids corresponding to the two clusters (which we call the training centroids). The centroid with the lower RMSE value and the median completion time was labeled as the high-performance centroid, whereas the centroid with higher values for the variables was labeled as the low-performance centroid (Figure 2). Participants were classified as low or high performers depending on the label of their closest centroid, measured using the Euclidean distance. Using this method, we trained two different models: one for the poking interaction and another for the ray casting interaction as both interactions are fundamentally disparate, potentially requiring different training strategies.

Here, we note that only the first six trials were used for each task when developing the adaptation models, owing to the results of our fatigue survey, where 33% (9/27) of the participants indicated that the training tasks induced fatigue specifically causing discomfort in their arms and shoulders, highlighting a common drawback of mid-air AR interactions (Brasier, Chapuis, Ferey, Vezien, & Appert, 2020).

In regards to the implementation details, we utilized a laptop equipped with an AMD Ryzen 9 CPU, an NVIDIA 3070 GPU and 16gb RAM to train the adaptation models. The models were trained using the scikit-learn (Pedregosa et al., 2011) and numpy (Harris et al., 2020) libraries available for the Python programming language.

[Figure 2 here]

3.4. Model Training Results

The training of our adaptation model resulted in the clustering of nine and fifteen out of the twenty-seven participants as low performers for the poking and raycasting interactions, respectively (Figure 2). For the poking interaction, median time to press a button (time variable) corresponding to the low and high-performance centroids was found to be 1.36 s and 0.96 s, respectively. Similarly, the RMSE value (consistency) was found to be 0.083 s and 0.197 s for low and high-performance centroids respectively. For raycasting, the values corresponding to the time and consistency variables were found to be 3.59 s and 0.576 s for the low-performance centroid and 1.93 s and 0.223 s for the high-performance centroid respectively. These centroids formed the basis for classifying new participants into low or high-performing clusters.

4. Phase II Study - Methods

[Figure 3 here]

4.1. Participants

For the Phase II study, twenty-one participants (nine females; twelve males) were recruited within the age group of 19 to 37 years (mean = 23.52; SD = 4.31). One participant was excluded due to the incompleteness of the study caused by AR sickness. All participants were from a university community, with educational backgrounds in engineering (n = 16), science (n = 2) and public health & medicine (n = 2). All participants self-reported to be right-handed and have less than one hour of AR and/or VR experience. All procedures were approved by the university's Institutional Review Board and complied with the American Psychological Association Code of Ethics. All participants were equally compensated with \$30 gift cards for their involvement.

[Figure 4 here]

4.2. Protocol

Following informed consent, participants were equipped with relevant bioinstruments and completed background questionnaires covering demographic information and the NASA TLX. Participants were then instructed to sit at rest with their eyes closed for a three-minute baseline collection. After the baseline was completed, participants donned an AR-HMD (HoloLens2, Microsoft Corporation, USA), and the study protocol began. The protocol took approximately 2 hours to complete and is illustrated in detail in Figure 4.

Training Task: Participants completed 6 trials of the poking and raycasting interactions with a two-minute break in between the exercises. The order of the tasks was counterbalanced across the study sample owing to the effect that fatigue might have on the second task. NASA TLX and CLT survey responses were collected after each interaction exercise.

Adaptation Task: Once the participants completed the Training Task, their temporal data was input into the adaptation model that classified them as low or high performers. Participants classified as low performers were given three additional trials of the interaction task, followed by the survey response collection.

Evaluation Task: All participants, irrespective of their performance, completed an evaluation task, which involved three trials of each interaction type. Along with NASA TLX and CLT surveys, participants completed the User Engagement Scale, System Usability Scale, and Device Feedback surveys.

4.3. Bioinstruments

Functional near-infrared spectroscopy (fNIRS) (NIRSport2, NIRx Medical Technologies LLC, USA) was used to monitor cortical hemodynamics at 50 Hz. The fNIRS probe map used an international 10/10 EEG system and captured 20 channels (based on Brodmann areas) using 8 infrared sources and 8 detectors (a total of 16 optodes) operating at two wavelengths ($\lambda = 760$ and 850 nm). This 20-channel layout (Figure 3(b)) mapped hemodynamic activity in 5 regions of interest: frontal eye field/ cingulate cortex (FEF/CG), right/left premotor cortex (R/L PMC), supplementary motor area (SMA), right/left primary motor area (R/L M1). Participants were also equipped with a chest-worn Actiheart device (Actiheart 5, CamNTech, Inc., UK) that recorded an electrocardiogram (ECG) signal at 1024 Hz to quantify physiological load. Overall, participants were equipped with the two bioinstruments described above that provided data for our neuroergonomic evaluation.

4.4. Testing the Adaptation Model

Our adaptation model was tested on the temporal performance data obtained from Phase II participants. Once a participant had completed their training task, their performance metrics were computed. We then added their two-dimensional data point (time and consistency) to the 27 training data points and recomputed the low and high-performance centroids and subsequently found the closest centroid to the new participant's data point. Next, we computed the Euclidean distance between the selected centroid and the training centroids. The new data point was then labeled based on the label of the training centroid corresponding to the shortest distance. While this method is different from classifying the new data point by simply finding its closest training centroid, we were able to better adapt to the effect of potential outliers that commonly affect the k-means clustering algorithm (Saxena et al., 2017). After classifying the new participant as a low or high performer, we provided relevant recommendations for additional training. If the participant was classified as a low performer on a specific interaction (e.g., poking or raycasting, or both), they were asked to complete three additional trials of that interaction before moving on to their evaluation task, whereas, for a high performer no additional trials were required.

4.5. Neuroergonomic Evaluation

The fNIRS signal processing was performed using algorithms from the Homer2 toolbox in MATLAB. Preprocessing included the removal of motion artifacts through peak detection and spline interpolation and the removal of specific bad channels based on recommendations from Homer2 (Dubb & Boas, 2016; Huppert, Diamond, Franceschini, & Boas, 2009). Two participants' data were removed from fNIRS processing due to significant noise. The modified Beer-Lambert Law was applied to determine the change in oxygenated, deoxygenated, and total hemoglobin concentrations. fNIRS signal processing and cleaning was done using an in-built function in MATLAB (Homer2, 2021) (Jahani, Setarehdan, Boas, & Yücel, 2018; Zhu, Rodriguez-Paras, Rhee, & Mehta, 2020). After the application of the processing stream, optical density was converted to concentration to measure oxygenated, deoxygenated, and total hemoglobin. Oxygenated hemoglobin (HbO) was further used to calculate functional connectivity, as it is most responsive to changes in cerebral blood flow due to the motor and working memory tasks (Hoshi, Kobayashi, & Tamura, 2001). A rest period of 3 minutes was provided for the global baseline, which was used to calculate peak activation for each phase (training and evaluation phases) for six different regions of interest (ROIs) for each participant (Zhu et al., 2020). To calculate functional connectivity between different brain regions, Pearson correlation coefficients were calculated by first calculating the global baseline i.e., by averaging HBO values for the last 2 minutes of the 3-minute rest period provided at the beginning of the task, (Tyagi et al., 2021). Followed by the calculation of peak activation for every trial within each phase for each channel (Tyagi et al., 2021). Once peak activation was calculated for each channel, 2s HbO values around the peak were averaged and subtracted from the global baseline, resulting in ΔHBO (Zhu et al., 2020). Followed by averaging each ΔHBO according to each ROI (Tyagi et al., 2021). Thus, resulting in six different ΔHbO values according to each ROI for each participant for each phase. Pearson correlations were transformed into a Fisher z-score to determine the strength of connectivity between different brain regions (Rhee & Mehta, 2018). Followed by replacing any value between -0.4 and +0.4 by 0 to avoid the possibility of false detection of the existence of connectivity between different brain regions (Rhee & Mehta, 2018).

Raw ECG signal was preprocessed to filter out motion artifacts and collect R-peaks via the peak detection algorithm. Time between successive R-R peaks (interbeat interval), was then calculated (Karthikeyan, Smoot, & Mehta, 2021). Low frequency (LF) and high frequency (HF) features were extracted to detect variations in sympathetic and parasympathetic activation (Iizuka, Ohiwa, Atomi, Shimizu, & Atomi, 2020; Melo, Nascimento, & Takase, 2017; Tran, Wijesuriya, Tarvainen, Karjalainen, & Craig, 2009). Root mean square of successive differences (RMSSD) and low-to-high frequency (LF/HF) were extracted to assess cognitive load (Tjolleng et al., 2017).

4.6. Statistical Analyses

For statistical analysis, fNIRS and HRV biosignals were divided into 2 phases (Training and Evaluation). Based on the adaptation model, participants were labeled as low and high performers for each AR interaction. As such, participants were placed into one of two training groups for the poking and raycasting tasks: low performer and high performer. All participants, irrespective of their group assignment also completed evaluation tasks. Three major statistical analyses were performed: (1) To test the hypothesis that the adaptation model successfully identified low from high performers, Mann-Whitney U tests were performed on the performance metric (i.e., median time to complete) during the training tasks between the two groups; (2) To uncover differences in the learning processes between low and high performers, Mann-Whitney U tests were performed on the neural (activation and functional connectivity), physiological (RMSSD, LF/HF ratio, AvgHR, LF, HF), and subjective responses (ratings of workload, cognitive load theory, and user engagement) during the training tasks; (3) To assess whether the adaptation was effective, linear mixed-effect models were employed to test the effects of group (low vs. high performers) and task (training vs. evaluation) on the performance data, with participants serving as random effects to account for individual variability (Bates, Mächler, Bolker, & Walker, 2014); and 4) to assess if the adaptation impacted neurophysiological and perceived load, Mann-Whitney U tests were performed on the neural, physiological, and subjective responses during the evaluation tasks.

5. Phase II Study - Results

5.1. Performance of the Adaptation Model

[Figure 5 here]

Of the twenty new participants who completed the training tasks, thirteen were classified as low performers in at least one of the poking or raycasting interactions. Amongst them, eight were classified as low performers in both interaction types, while three were low performers in poking only, and two were low performers in raycasting only.

[Table 1 here]

Poking Interaction: During the training task, high performers exhibited significantly shorter median time to complete than low performers (p < 0.0001; $\eta^2 = 0.96$). They were 29.76% quicker than the low performers, with an average value of 1.01 s for the median completion time compared to low performers' 1.44 s (Figure 5 (left column)). They were also more consistent with an average RMS value of 0.09 s compared to 0.28 s for low performers.

The results obtained from the linear mixed-effects model showed significant task $(F(1, 18) = 15.895, p < 0.001, \eta^2 = 0.47)$ and group $(F(1, 18) = 27.229, p < 0.001, \eta^2 = 0.60)$ main effects, where longer median completion times were observed during training for high (9.52% longer) and low performers (29.12% longer) when compared to evaluation. The task x group interaction was also found significant $(F(1, 18) = 5.238, p = 0.034, \eta^2 = 0.23)$, where the low performers showed significantly higher median completion times than high performers, however, this was only found in the training task. In the evaluation task, comparable performance was observed between groups (Figure 5 (left column)). Consistency could not be compared in the evaluation stage owing to the reduced number of trials (n = 3).

Raycasting Interaction: High performers were significantly quicker (41.38%; p = 0.001; $\eta^2 = 1.0$) than low performers in the training task with an average median completion time of 2.04 s compared to 3.48 s (Figure 5 (right)). High performers were also more consistent with an average RMSE value of 0.26 s versus 0.85 s for low-performers.

The linear mixed-effects model results showed significant task $(F(1, 18) = 47.107, p < 0.001, \eta^2 = 0.72)$ and group $(F(1, 18) = 33.180, p < 0.001, \eta^2 = 0.65)$ main effects, where similar to the poking interaction, longer median completion times were

observed during training for high (18.01% longer) and low performers (56.22% longer) when compared to evaluation. The task x group interaction was also found significant $(F(1,18) = 17.143, p < 0.001, \eta^2 = 0.49)$, where the low performers showed significantly higher median completion times than high performers in the training task, however, comparable performance was observed between groups during the evaluation task (Figure 5 (right column)).

5.2. Functional Connectivity

[Figure 6 here]

For poking, compared to low performers, high performers exhibited greater connectivity strengths across FEF/CG-RPMC (p = 0.01), LPMC-RPMC (p = 0.015), and LPMC-RM1 (p = 0.014) during training. During evaluation, functional connectivity between different brain regions was found to be comparable across both groups (all p > 0.154), except for FEF/CG-SMA (p = 0.028), where high performers exhibited greater connectivity than low performers (Figure 6 (g)).

For raycasting, high performers exhibited stronger functional connectivity across FEF/CG-SMA (p = 0.03), FEF/CG-RPMC (p = 0.049), and FEF/CG-LM1 (p = 0.024) during training. During evaluation, functional connectivity between FEF/CG-LPMC (p = 0.013), FEF/CG-LM1 (p = 0.011), and LPMC-LM1 (p = 0.023) was observed to be stronger among high performers compared to low performers (Figure 7g).

5.3. Neural Activation

[Figure 7 here]

For poking (Figure 6 a-f), both low and high performers exhibited comparable brain activation levels across the six ROIs (all p > 0.277) during the training phase. In the evaluation phase, low performers exhibited greater activation in the FEF/CG region than the high performers, though this was marginal (p = 0.07), while no other differences were observed in other ROIs (all p > 0.178).

For raycasting (Figure 7 a-f), high performers exhibited greater activation in the RM1 region than the low performers (p = 0.011) during the training phase, while no other differences were observed in other ROIs (all p > 0.26). In the evaluation phase, low performers exhibited greater activation in the FEF/CG region than the high performers (p = 0.021) and a marginal increase in the SMA (p = 0.075), while no other differences were observed in other ROIs (all p > 0.25).

5.4. Heart Rate Variability

[Table 2 here]

For poking, RMSSD was significantly lower in the high performers than the low performers (p = 0.043) during training, while no other HR/V features showed any group differences (all p > 0.122). HR/V features were found to be comparable across groups during the evaluation phase (all p > 0.083)(Figure 2 HR/V row).

For raycasting, no group differences were seen across any HR/V features during training or evaluation (all p > 0.243).

5.5. Subjective Responses

[Figure 8 here]

For poking interaction, no group differences were seen across any NASA TLX subscales during training or evaluation (all p > 0.283). Both groups reported comparable scores across all subscales of the cognitive load theory and user engagement scale during training and evaluation (all p > 0.091) (Figure 8 top row).

For raycasting, low performers reported higher demands across mental (p = 0.058), performance (p = 0.058), and frustration (p = 0.049) sub-scales than low performers during training (Figure 8 denoted with *). However, during evaluation, group differences were not found significant on any NASA TLX sub-scale (all p > 0.1). Both groups reported comparable scores across all subscales of the cognitive load theory and user engagement scale during training and evaluation (all p > 0.161) (Figure 2 UES and CLT rows).

6. Limitations

Our adaptation models were trained using the temporal performance data from twentyseven novice participants. While the sample size is relatively small compared to the data used to train large scale machine learning models, our choice of the unsupervised k-means clustering technique (Saxena et al., 2017), with a small and known number of clusters (k = 2 for low and high performers) performed well on new participants. We showed the effectiveness of the models in our Phase II study, where participants classified by the model as high and low performers showed significant difference in their performance metric as well as differences in their neuroergonomic patterns (functional connectivity and neural activation). Further studies, with a larger number of participants, can be conducted using the insights from our current work, to build more robust adaptation models. Additionally, while we did not consider any individual or group differences (such as gender, background, etc.) between participants in training our models, future studies could incorporate these differences into the development of more personalized adaptation models. Recognizing that individual differences (owing to user-specific learning processes, their motoric interaction quality, etc.) may impact model evaluation, we incorporated participants as a random effect in the linear mixed model analyses performed for the Phase II study.

Another limitation in our study was our inability to compute consistency of users in the Phase II evaluation tasks owing to the reduced number of trials (n = 3). Our choice for the reduced evaluation trials was to prevent fatigue in users, that was reported by some participants during their last couple of training trials in the Phase I study. However, based on our Phase II subjective analysis that showed no significant changes in cognitive loads and user engagement after additional training, further studies can be conducted with more number of evaluation trials to give us a better understanding of the change in consistency between the training and evaluation tasks across low and high-performing users. We would also like to note here that in our Phase-I study, we had a fixed order for the training tasks (i.e., raycasting after poking) which might have contributed to the fatigue issues in some users. We rectified this in our Phase-II study where the order of tasks was counterbalanced across the study sample. Further studies can use the counter-balanced approach during both the training and testing phases.

We also observed limitations in our HRV analysis, where, due to the short period of the unit interactions in our selection task (i.e., pressing a button), it was difficult to obtain stable heart rate and perform ultra short analysis for the short duration. This prevented us from observing many meaningful differences in heart rate variability between low and high performers. Future studies might have to utilize HRV analysis primarily for longer interaction techniques (e.g., scrolling or moving).

7. Discussion

Our primary focus in this work lies in the use of AR/VR systems as a training tool. In this specific context, our studies reveal a fundamental conceptual gap across how users perceive and learn unit (basic) interactions in AR/VR, how those interactions are designed, and how visuo-motor behaviour is mapped to user efficacy. For instance, it was already known that raycasting is complex from a visuo-motor perspective (Argelaguet & Andujar, 2013). Work by Kopper et al. (Kopper et al., 2010) showed that the motor behavior model for raycasting (or *distal pointing*), unlike its 2D counterpart, follows a non-linear (quadratic) relationship between task difficulty and angular target size in contrast to Fitt's law (linear growth in difficulty with *distance*).

From the neuroergonomics perspective, increased connectivity between brain regions, i.e., functional integration, has shown to be associated with efficient neural strategies typically adopted with expertise development, while increased activation of select brain regions, i.e., functional specialization, implicates increased requirements of cognitive resources to maintain performance (Tyagi et al., 2021). In this respect, our work offers a multi-dimensional metric of performance (what happened at the *hand*) that is supported by neuroergonomic data (what happened in the *brain*). Specifically, we show that the difference in motor strategies is also accompanied by quantifiable differences in neural strategies across the two interactions. Below we discuss the implications of our study within this broader context.

7.1. Training with AR — What is the user learning?

"Context is everything" is often used as an underlying adage for evaluating user interactions in AR/VR/MR interfaces focused on training and adaptation. As a result, the design of the training tasks and the metrics to evaluate user performance in those tasks are naturally considered as the key factors that influence the development of adaptive training (Kelley, 1969).

Recent studies that have employed adaptive AR-based training systems (G. Huang et al., 2021; Westerfield et al., 2015) mainly focus on training outcomes that are governed by the training context, e.g., medical tasks or machine tasks, and thereby focus on metrics such as task completion speed or errors. This can confound key usability metrics for fielded AR applications (Dwivedi et al., 2022). For instance, notion of completion time is dependent on how the task is defined in the context of the application rather than the unit action taken by the users to reach and interact with (touch, click, or point, etc.) with the virtual object. Similarly, the notion of accuracy and the measurement of "success" is also completely contextual (e.g. did the welding happen properly? vs. did the user reach the target spot within a certain positional and orientational threshold?). As a result, the user's prior psychomotor capabilities to effectively perform the AR-based interactions were either not taken into account or controlled for by offering structured training.

In principle, our work does not lead to a disagreement with contextual investigation of AR interactions. However, our studies call for taking a step back to re-consider what it means for a user to be a learner or an expert in a real-world task (say triage in emergency response) versus a learner or an expert in using AR as a tool for training. Specifically, our psychomotor and neuroergonomic analyses, when taken in conjunction, offer key insights that could be instrumental in developing personalized feedback and adaptation strategies for AR-based training. For instance, while the low performers improved in their performance after additional training, they still took longer to complete than the high performers. For poking, which is the simpler of the two interactions we considered, the neural activation and connectivity patterns (i.e., higher frontal eye field activation and its lack of connectivity with the supplementary motor area) in low performers indicate that their visual attention and visuomotor search efficiencies were still subpar to the high performers (Gitelman, Parrish, Friston, & Mesulam, 2002). This was more evident with raycasting or distal pointing, wherein high performers maintained strong connectivity between visual attention and motor function regions that low performers were unable to match even after additional trials.

While our bi-variate performance metric was useful in training adaptation models for both poking and raycasting interactions, we observed some critical differences between the two interactions. The raycasting adaptation model labeled over half (55%) of Phase I participants as low performers, while the poking adaptation model labeled only one-third as low performers. This suggested that more users experienced the raycasting interaction as comparatively harder to learn. The observation was subsequently supported through user perceptions of workload during Phase II, where low-performing participants reported significantly higher mental demand and frustration than highperforming users for the raycasting interactions, whereas no significant difference was observed for poking Figure 8.

7.2. Multi-dimensional Metrics in AR — How to assess learning?

The characterization of learning and its relation with AR-based interaction evaluation and design has largely relied on a *bag-of-metrics* approach (i.e. multiple measures such as completion time, accuracy, etc. taken individually). In contrast, our choice to model user performance (phase I) as a joint (bi-variate) measure and especially the inclusion of consistency reflects our focus on characterizing users' expertise in and learning of the AR itself. Even though this seems to be a reasonable (if not an obvious) choice, our studies strongly point to a need for re-visiting the metrics for assessing user efficacy in AR-based interactions in the form of multi-dimensional metrics.

This need for multi-dimensional metrics is also informed through literature on active learning. As an example, the time variable provided insights into the users' instantaneous performance, while the consistency variable provided insights into the learning behavior, where performance variability indicated active learning (Sher et al., 2020; Wu et al., 2014).

The key evidence that multi-dimensionality could be an important step for future AR-systems comes from the neuroergonimcs perspective. Specifically, when we compared high and low performers, the low performers exhibited long time to complete during the training phase and this was accompanied by weaker functional brain connectivity between the frontal eye field region and left/right premotor cortex, as well as between left and right premotor cortex (Figure 6(g), Figure 7(g)). Interestingly, despite differences in training performances, neural activation patterns of each of the regions of interest in the frontal and motor areas were largely comparable between the two groups. This implies distinct neural strategies adopted by high performers who

exhibited quicker completion times during the training phase, which signaled efficient brain network integration, rather than functional specialization (Tyagi et al., 2021), between regions that regulate visuospatial attention and motor preparation (Simon et al., 2002). That the bi-variate performance metric was able to differentiate between users who adopted distinctly different neural strategies further strengthens the utility of the metric.

While there is ongoing research focused on improving mid-air interactions through the use of tangible interfaces (H.-M. Huang, Huang, & Cheng, 2021; Nizam et al., 2018), their widespread adoption in fielded AR applications is still lacking (Masood & Egger, 2019). This raises an unanswered question that likely impacts evaluations of fielded AR applications — should performance metrics during task performance and/or training on AR skills be interaction-specific? For example, object selection using raycasting interaction requires significantly greater hand-eye coordination than poking, as the objects are away from the user and can occlude one another (Herndon, Van Dam, & Gleicher, 1994). We see this as an immense opportunity for establishing a richer metric space that includes other modalities such as gaze-related metrics to capture to understand how different users learn complex AR interactions (Argelaguet & Andujar, 2009; Wolf, Lohmeyer, Holz, & Meboldt, 2021). This may particularly benefit the development of new interaction-specific learning assessment and adaptation models for AR interactions that take place outside the user's reach (Ye et al., 2022).

7.3. Adaptation Strategies in AR — Who, how, and how much?

The choice of a bi-variate (time+consistency) performance space in combination with cluster analysis worked successfully toward identifying users in need for adaptation. This is also supported by the neuroergonomics data. Specifically, the high performers were significantly quicker and more consistent in their training tasks compared to low performers, signifying that low performers were still actively learning at the end of their training task. As a result, the additional training provided to the low performers ers contributed to significant time-based improvements during their evaluation tasks. Having said this, we were concerned that enforcing additional repetitions as a means for adaptation might add to their physical and cognitive loads. To our surprise, this was not the case. Interestingly, perceived workload, cognitive load elements (i.e., intrinsic, germane, and extraneous load), and user engagement ratings were comparable between high and low performers during the evaluation task, indicating that the *additional training did not additionally burden low performers*. This observation further supported our bi-variate performance metric that identified actively learning users who were more accepting of additional trials.

Note that while the additional trials helped low performers improve their median completion times significantly during the evaluation tasks, they were still slower compared to the high performers (Figure 5). A possible reason for this difference could be our adaptation strategy of providing a fixed number of similar trials to all low performers. Prior research on motor learning has demonstrated that variable repetition on consecutive training trials enhances the process of skill learning when compared to practice on similar trials (Vleugels, Swinnen, & Hardwick, 2020), such as that experienced by users during adaptation in the present study. However, the current results indicate that participants' performance may plateau after a number of repetitions, after which additional repetitions may not help. For instance, high performers did not see the same percentage of improvement in their evaluation tasks compared to the low performers (Figure 5). An alternative approach for adaptive training, could potentially be to embrace variable training strategies (Bonney, Jelsma, Ferguson, & Smits-Engelsman, 2017), which focus on introducing *contextual interference* by modifying stimuli information while keeping the repetitions consistent. However, the impacts of variable training paradigms in AR require additional investigations as they have been shown to increase cognitive load (Guadagnoli & Lee, 2004). With more complex AR interactions, such as raycasting, trade-offs between adaptation strategies (e.g., variable training) and user cognitive load need to be fully contemplated, as these can impact perceived usefulness, usability, and subsequent acceptance of adaptive training systems (Dwivedi et al., 2022).

Another perspective towards adaptive training strategies would be to take a flexible approach towards clustering and classifying users. In our work, we took a binary approach of classifying users as low and high performers to maintain uniformity across poking and raycasting interactions. However, for a more interaction-specific classification, the number of clusters can be dynamically chosen (Shafeeq & Hareesha, 2012), that might generate clusters with specific characteristics (e.g., a cluster categorizing dominantly inconsistent users) which can then be used to provide more personalized training (e.g., more trials to improve consistency). For applications that involve complex interactions requiring higher cognitive and physical demands, such as ER and surgical training (Nunes, Lucas, Simões-Marques, & Correia, 2018; Vávra et al., 2017), it may be necessary to consider multi-dimensional metrics that integrate performance and neurophysiological measures. In such applications, methods like hierarchical and grid-based clustering algorithms (Saxena et al., 2017) can better characterize the highdimensional modalities, thereby, providing opportunities for more dynamic adaptation strategies.

8. Conclusion

In this work, we took a two-phase study approach toward training users in the basic AR-interactions. First, we identified a bi-variate performance metric and developed adaptation models to identify users who would need additional training. Next, we tested the performance of these models and carried out neuroergonomic evaluations to understand their effects on the users' neurophysiological processes. Through our Phase-I study, we showed the successful use of our bi-variate performance metric, derived from time and consistency, in developing adaptation models that clustered users into low and high-performing groups for each of the interaction types. In Phase-II, we tested these models on new users to identify the low performers. Given that the focus of this research was primarily to identify users who needed adaptation, the study adopted a basic adaptation strategy of providing additional repetitions, shown to be a hallmark characteristic of expertise development (Gupta & Cohen, 2002). However, by adopting a neuroergonomics perspective, i.e., examining what happened in the *brain*, this study was able to capture the different neural strategies exhibited by low and high-performing users, which revealed (for the first time) key information and spatial processing characteristics of AR interactions that can guide the development of novel paradigms for personalized training in AR. We believe integrating psychomotor and neuroergonomic approaches offers a rich research platform for AR/VR interfaces focused on providing visual cues or guides to support visual attention and facilitate motor preparation.

Acknowledgement(s)

This project was funded by the National Science Foundation awards #2033592 and #2013122.

Disclosure statement

The authors report there are no competing interests to declare.

References

- Abujelala, M., Karthikeyan, R., Tyagi, O., Du, J., & Mehta, R. K. (2021). Brain activity-based metrics for assessing learning states in vr under stress among firefighters: An explorative machine learning approach in neuroergonomics. *Brain Sciences*, 11(7), 885.
- Albayrak, M. S., Öner, A., Atakli, I. M., & Ekenel, H. K. (2019). Personalized training in fast-food restaurants using augmented reality glasses. In 2019 international symposium on educational technology (iset) (pp. 129–133).
- Anderson, F., Grossman, T., Matejka, J., & Fitzmaurice, G. (2013). Youmove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual* acm symposium on user interface software and technology (pp. 311–320).
- Argelaguet, F., & Andujar, C. (2009). Efficient 3d pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, 29(6), 34–43.
- Argelaguet, F., & Andujar, C. (2013). A survey of 3d object selection techniques for virtual environments. Computers & Graphics, 37(3), 121–136.
- Barsom, E. Z., Graafland, M., & Schijven, M. P. (2016). Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical endoscopy*, 30, 4174–4183.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
- Biddle, E., & Buck, B. (2019). Adaptive team training for one. In Adaptive instructional systems: First international conference, ais 2019, held as part of the 21st hci international conference, hcii 2019, orlando, fl, usa, july 26–31, 2019, proceedings 21 (pp. 15–27).
- Billinghurst, M. (2021). Grand challenges for augmented reality. *Frontiers in Virtual Reality*, 12.
- Billinghurst, M., Clark, A., Lee, G., et al. (2015). A survey of augmented reality. Foundations and Trends (R) in Human-Computer Interaction, 8(2-3), 73–272.
- Bonney, E., Jelsma, L. D., Ferguson, G. D., & Smits-Engelsman, B. C. (2017). Learning better by repetition or variation? is transfer at odds with task specific training? *PLoS One*, 12(3), e0174214.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, 58-75. Retrieved from https://www.sciencedirect.com/science/article/pii/S0149763412001704 (Applied Neuroscience: Models, methods, theories, reviews. A Society of Applied Neuroscience (SAN) special issue.)
- Bottani, E., & Vignali, G. (2019). Augmented reality technology in the manufacturing industry: A review of the last decade. *Iise Transactions*, 51(3), 284–310.
- Bowman, D. A. (1999). Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application. Georgia Institute of Technology.
- Brasier, E., Chapuis, O., Ferey, N., Vezien, J., & Appert, C. (2020). Arpads: Mid-air indirect input for augmented reality. In 2020 ieee international symposium on mixed and augmented reality (ismar) (p. 332-343).

Buckman, J. M. (2005). Chief fire officer's desk reference. Jones & Bartlett Learning.

- Bunce, S. C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K., & Onaral, B. (2011). Implementation of fnirs for monitoring levels of expertise and mental workload. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition. directing the future of adaptive systems* (pp. 13–22). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dey, A., Chatburn, A., & Billinghurst, M. (2019). Exploration of an eeg-based cognitively adaptive training system in virtual reality. In 2019 ieee conference on virtual reality and 3d user interfaces (vr) (pp. 220–226).
- Doswell, J. T., & Skinner, A. (2014). Augmenting human cognition with adaptive augmented reality. In Foundations of augmented cognition. advancing human performance and decisionmaking through adaptive systems: 8th international conference, ac 2014, held as part of hci international 2014, heraklion, crete, greece, june 22-27, 2014. proceedings 8 (pp. 104–113).
- Dubb, J., & Boas, D. (2016). Homer 2 toolbox, v2.1. https://www.nitrc.org/frs/shownotes.php?release_id = 3156.
- Dwivedi, S., Hayes, J., Pedron, I., Kang, J., Brenner, L. J., Barnes, C., ... others (2022). Comparing the efficacy of ar-based training with video-based training. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 66, pp. 1862–1866).
- Evarts, B., & Molis, J. L. (2018). United states firefighter injuries 2017 (Tech. Rep.). National Fire Protection Association.
- Fuentes-García, J. P., Villafaina, S., Collado-Mateo, D., De la Vega, R., Olivares, P. R., & Clemente-Suárez, V. J. (2019). Differences between high vs. low performance chess players in heart rate variability during chess problems. *Frontiers in psychology*, 10, 409.
- Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical issues in ergonomics science*, 4(1-2), 113–131.
- Ghilardi, M.-F., Ghez, C., Dhawan, V., Moeller, J., Mentis, M., Nakamura, T., ... Eidelberg, D. (2000). Patterns of regional brain activation associated with different forms of motor learning. *Brain Research*, 871(1), 127-145. Retrieved from https://www.sciencedirect.com/science/article/pii/S0006899300023659
- Gibson, J. J. (2014). The ecological approach to visual perception: classic edition. Psychology press.
- Gitelman, D. R., Parrish, T. B., Friston, K. J., & Mesulam, M.-M. (2002). Functional anatomy of visual search: regional segregations within the frontal eye fields and effective connectivity of the superior colliculus. *Neuroimage*, 15(4), 970–982.
- Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of motor behavior*, 36(2), 212–224.
- Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological review*, 109(2), 401.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020, September). Array programming with NumPy. *Nature*, 585(7825), 357–362. Retrieved from https://doi.org/10.1038/s41586-020-2649-2
- Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, pp. 904–908).
- Hayes, J., Dwivedi, S., Karthikeyan, R., Abujelala, M., Kang, J., Ye, Y., ... Mehta, R. K. (2022). Identifying early predictors of learning in vr-based drone training. In *Proceedings* of the human factors and ergonomics society annual meeting (Vol. 66, pp. 1872–1876).
- Henderson, S. J., & Feiner, S. K. (2011). Augmented reality in the psychomotor phase of a procedural task. In 2011 10th ieee international symposium on mixed and augmented reality (p. 191-200).
- Herndon, K. P., Van Dam, A., & Gleicher, M. (1994). The challenges of 3d interaction: a chi'94 workshop. ACM SIGCHI Bulletin, 26(4), 36–43.
- Hoshi, Y., Kobayashi, N., & Tamura, M. (2001). Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. *Journal of applied* physiology, 90(5), 1657–1662.

- Huang, G., Qian, X., Wang, T., Patel, F., Sreeram, M., Cao, Y., ... Quinn, A. J. (2021). Adaptutar: An adaptive tutoring system for machine tasks in augmented reality. In *Proceedings* of the 2021 chi conference on human factors in computing systems (pp. 1–15).
- Huang, H.-M., Huang, T.-C., & Cheng, C.-Y. (2021). Reality matters? exploring a tangible user interface for augmented-reality-based fire education. Universal Access in the Information Society, 1–13.
- Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied optics*, 48(10), D280–D298.
- Iizuka, T., Ohiwa, N., Atomi, T., Shimizu, M., & Atomi, Y. (2020). Morning heart rate variability as an indication of fatigue status in badminton players during a training camp. *Sports*, 8(11), 147.
- Jahani, S., Setarehdan, S. K., Boas, D. A., & Yücel, M. A. (2018). Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and savitzky-golay filtering. *Neurophotonics*, 5(1), 015003–015003.
- Jensen, E. (2005). Teaching with the brain in mind. ASCD.
- Karthikeyan, R., Smoot, M. R., & Mehta, R. K. (2021). Anodal tdcs augments and preserves working memory beyond time-on-task deficits. *Scientific reports*, 11(1), 19134.
- Kast, V., & Leukel, C. (2016). Motor experts care about consistency and are reluctant to change motor outcome. *Plos one*, 11(8), e0161798.
- Kelley, C. R. (1969). What is adaptive training? Human Factors, 11(6), 547–556.
- Kesedžić, I., Šarlija, M., Božek, J., Popović, S., & Ćosić, K. (2021). Classification of cognitive load based on neurophysiological features from functional near-infrared spectroscopy and electrocardiography signals on n-back task. *IEEE Sensors Journal*, 21 (13), 14131-14140.
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in psychology*, 8, 1997.
- Kopper, R., Bowman, D. A., Silva, M. G., & McMahan, R. P. (2010). A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10), 603–615.
- LaViola, J., Williamson, B., Brooks, C., Veazanchin, S., Sottilare, R., & Garrity, P. (2015). Using augmented reality to tutor military tasks in the wild. In Proceedings of the interservice/industry training simulation & education conference, orlando, florida.
- Looser, J., Billinghurst, M., Grasset, R., & Cockburn, A. (2007). An evaluation of virtual lenses for object selection in augmented reality. In *Proceedings of the 5th international* conference on computer graphics and interactive techniques in australia and southeast asia (pp. 203–210).
- Ma, M., Fallavollita, P., Seelbach, I., Von Der Heide, A. M., Euler, E., Waschke, J., & Navab, N. (2016). Personalized augmented reality for anatomy education. *Clinical Anatomy*, 29(4), 446–453.
- Masood, T., & Egger, J. (2019). Augmented reality in support of industry 4.0—implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing*, 58, 181– 195.
- Mehta, R., Moats, J., Karthikeyan, R., Gabbard, J., Srinivasan, D., Du, E., ... Fernandes, R. (2022). Human-centered intelligent training for emergency responders. AI Magazine, 43(1), 83–92.
- Melo, H. M., Nascimento, L. M., & Takase, E. (2017). Mental fatigue and heart rate variability (hrv): The time-on-task effect. *Psychology & Neuroscience*, 10(4), 428.
- Micheletti, C. (1996). A survey of medical neuroscience. Neurophysiologie Clinique / Clinical Neurophysiology, 26(6), 450-451.
- Militello, L., Sushereba, C., Hernandez, O., & Patterson, E. S. (2019). Augmented reality adaptive training principles. In *Proceedings of the international symposium on human* factors and ergonomics in health care (Vol. 8, pp. 72–75).
- Młynarska, E., Greene, D., & Cunningham, P. (2016). Time series clustering of moodle activity

data. In 24th irish conference on artificial intelligence and cognitive science (aics'16), university college dublin, dublin, ireland, 20-21 september 2016.

- Mohanty, R. R., & Krishnamurthy, V. R. (2021). Kinesthetic metaphors for precise spatial manipulation: A study of object rotation. *Journal of Computing and Information Science* in Engineering, 21(2).
- Mukherjee, S., Yadav, R., Yung, I., Zajdel, D. P., & Oken, B. S. (2011). Sensitivity to mental effort and test–retest reliability of heart rate variability measures in healthy seniors. *Clinical Neurophysiology*, 122(10), 2059–2066.
- Nizam, S. M., Abidin, R. Z., Hashim, N. C., Lam, M. C., Arshad, H., & Majid, N. (2018). A review of multimodal interaction technique in augmented reality environment. *Int. J. Adv. Sci. Eng. Inf. Technol*, 8(4-2), 1460.
- Nunes, I. L., Lucas, R., Simões-Marques, M., & Correia, N. (2018). Augmented reality in support of disaster response. In Advances in human factors and systems interaction: Proceedings of the ahfe 2017 international conference on human factors and systems interaction, july 17-21, 2017, the westin bonaventure hotel, los angeles, california, usa 8 (pp. 155–167).
- O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112, 28–39.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peretz, C., Korczyn, A. D., Shatil, E., Aharonson, V., Birnboim, S., & Giladi, N. (2011). Computer-based, personalized cognitive training versus classical computer games: a randomized double-blind prospective trial of cognitive stimulation. *Neuroepidemiology*, 36(2), 91–99.
- Rhee, J., & Mehta, R. K. (2018). Functional connectivity during handgrip motor fatigue in older adults is obesity and sex-specific. Frontiers in human neuroscience, 12, 455.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664–681.
- Seidler, R. D., Noll, D. C., & Thiers, G. (2004). Feedforward and feedback processes in motor control. Neuroimage, 22(4), 1775–1783.
- Shafeeq, A., & Hareesha, K. (2012). Dynamic clustering of data with modified k-means algorithm. In *Proceedings of the 2012 conference on information and computer networks* (pp. 221–225).
- Sher, V., Hatala, M., & Gašević, D. (2020). Analyzing the consistency in within-activity learning patterns in blended learning. In *Proceedings of the tenth international conference* on learning analytics & knowledge (pp. 1–10).
- Shi, Y., Zhu, Y., Mehta, R. K., & Du, J. (2020). A neurophysiological approach to assess training outcome under stress: A virtual reality experiment of industrial shutdown maintenance using functional near-infrared spectroscopy (fnirs). Advanced Engineering Informatics, 46, 101153.
- Simon, S. R., Meunier, M., Piettre, L., Berardi, A. M., Segebarth, C. M., & Boussaoud, D. (2002). Spatial attention and memory versus motor preparation: premotor cortex involvement as revealed by fmri. *Journal of neurophysiology*, 88(4), 2047–2057.
- Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. International Journal of Human-Computer Interaction, 38(9), 801–824.
- Steed, A., Pan, Y., Zisch, F., & Steptoe, W. (2016). The impact of a self-avatar on cognitive load in immersive virtual reality. In 2016 ieee virtual reality (vr) (pp. 67–76).
- Suriya-Prakash, M., John-Preetham, G., & Sharma, R. (2015). Is heart rate variability related to cognitive performance in visuospatial working memory? *PeerJ PrePrints*, 3, e1377v1.
- Tanaka, A., Craighead, J., Taylor, G., & Sottilare, R. (2019). Adaptive learning technology for ar training: Possibilities and challenges. In Adaptive instructional systems: First international conference, ais 2019, held as part of the 21st hei international conference, heii 2019,

orlando, fl, usa, july 26-31, 2019, proceedings 21 (pp. 142-150).

- Tao, Y., Coltey, E., Wang, T., Alonso, M., Shyu, M.-L., Chen, S.-C., ... Vassigh, S. (2020). Confidence estimation using machine learning in immersive learning environments. In 2020 ieee conference on multimedia information processing and retrieval (mipr) (pp. 247–252).
- Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., ... Park, S. (2017). Classification of a driver's cognitive workload levels using artificial neural network on ecg signals. *Applied* ergonomics, 59, 326–332.
- Tomasi, D., Wang, G.-J., & Volkow, N. D. (2013). Energetic cost of brain functional connectivity. Proceedings of the National Academy of Sciences, 110(33), 13642–13647.
- Tran, Y., Wijesuriya, N., Tarvainen, M., Karjalainen, P., & Craig, A. (2009). The relationship between spectral changes in heart rate variability and fatigue. *Journal of Psychophysiology*, 23(3), 143–151.
- Tyagi, O., Hopko, S., Kang, J., Shi, Y., Du, J., & Mehta, R. K. (2021). Modeling brain dynamics during virtual reality-based emergency response learning under stress. *Human* factors, 00187208211054894.
- Vávra, P., Roman, J., Zonča, P., Ihnát, P., Němec, M., Kumar, J., ... El-Gendi, A. (2017). Recent development of augmented reality in surgery: a review. *Journal of healthcare engi*neering, 2017.
- Vleugels, L. W., Swinnen, S. P., & Hardwick, R. M. (2020). Skill acquisition is enhanced by reducing trial-to-trial repetition. *Journal of Neurophysiology*, 123(4), 1460–1471.
- Walia, P., Fu, Y., Norfleet, J., Schwaitzberg, S. D., Intes, X., De, S., ... Dutta, A. (2022). Error related fnirs-eeg microstate analysis during a complex surgical motor task. In 2022 44th annual international conference of the ieee engineering in medicine & biology society (embc) (pp. 941–944).
- Westerfield, G., Mitrovic, A., & Billinghurst, M. (2015). Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education*, 25, 157–172.
- Wolf, J., Lohmeyer, Q., Holz, C., & Meboldt, M. (2021). Gaze comes in handy: Predicting and preventing erroneous hand actions in ar-supported manual tasks. In 2021 ieee international symposium on mixed and augmented reality (ismar) (pp. 166–175).
- Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature neuroscience*, 17(2), 312–321.
- Xu, W., Liang, H.-N., He, A., & Wang, Z. (2019). Pointing and selection methods for text entry in augmented reality head mounted displays. In 2019 ieee international symposium on mixed and augmented reality (ismar) (pp. 279–288).
- Ye, Y., Shi, Y., Xia, P., Kang, J., Tyagi, O., Mehta, R. K., & Du, J. (2022). Cognitive characteristics in firefighter wayfinding tasks: An eye-tracking analysis. *Advanced Engineering Informatics*, 53, 101668.
- Zhou, J., & Bhat, S. (2021). Modeling consistency using engagement patterns in online courses. In Lak21: 11th international learning analytics and knowledge conference (pp. 226–236).
- Zhu, Y., Rodriguez-Paras, C., Rhee, J., & Mehta, R. K. (2020). Methodological approaches and recommendations for functional near-infrared spectroscopy applications in hf/e research. *Human factors*, 62(4), 613–642.

		Poking		Raycasting	
		p-value	η^2	p-value	η^2
High vs. Low	Training	< 0.0001	0.96	0.001	1
Performers	Evaluation	0.04	0.54	0.031	0.58

 Table 1. Significant differences in median completion times between high- and low-performing groups during

 Training and Evaluation tasks for Phase-II.

		Poking	g M(SD)	Raycasting M(SD)		
		Training	Evaluation	Training	Evaluation	
$\mathrm{HR/V}$	Mean HR (beats/minute)	83.33(9.38)	83.17(8.43)	87.58(9.15)	84.35(14.22)	
	LF/HF ratio	24.89(9.12)	27.37(15.39)	22.9(9.04)	26.43(12.32)	
	RMSSD (ms)	$5.24^{*}(3.57)$	4.84(5.58)	7.19(6.07)	4.63(3.58)	
	LF (Hz)	805.6 (714.4)	1100.77 (1987.71)	1234.55(1925.52)	699.91 (661.63)	
	HF (Hz)	190.52(159.33)	467.19(1011.52)	178.34(130.89)	241.68(293.67)	
$\begin{array}{c} \textbf{UES} \\ (\min \ (1) - \max \ (5)) \end{array}$	Focused attention	2.84(0.76)	3.17(0.94)	2.84(0.76)	3.17(0.94)	
	Perceived usability	3.73(0.62)	3.93(0.87)	3.73(0.62)	3.93(0.87)	
	Aesthetics	2.73(0.52)	2.58(0.76)	2.73(0.52)	2.58(0.76)	
	Reward factor	2.27(0.62)	2.03(0.89)	2.27(0.62)	2.03(0.89)	
$\begin{array}{c} \textbf{CLT} \\ (\min \ (1) - \max \ (7)) \end{array}$	Intrinsic load	2.6(0.83)	2.55(1.08)	2.6(0.83)	2.55(1.05)	
	Germane load	4.92(1.27)	4.78(1.85)	4.92(1.27)	4.67(1.85)	
	Extraneous load	2.88(0.79)	2.69(0.57)	2.88(0.79)	2.7(0.55)	

 Table 2.
 Mean (SD) values during Training and Evaluation of heart rate variability (HRV), User Engagement

 Scale (UES), and Cognitive Load Theory (CLT) responses across all users for Phase-II. * denotes significant

 differences between low- and high-performing groups.



Figure 1. Phase I Study Workflow: Twenty-seven participants started with a (a) Training Task comprising of 8 trials of sequentially selecting buttons using the poking and raycasting interactions. (b) The bi-variate performance metric was used to compute the time (overall median completion time) and consistency (RMSE) values for each user. (c) Using the data from the twenty-seven participants, we trained adaptations models for both the interaction types (only poking show in figure) using the k-means clustering algorithm. We obtained clusters for low (white circles) and high (black circles) performing users.



Figure 2. Training results of the poking (left) and raycasting (right) adaptation models using the k-means clustering algorithm. Twenty-seven participants from Phase-I study were clustered into low (white circles) and high (black circles) performing groups, with the cluster centroids marked in red diamond markers. Nine and fifteen participants were labeled as low performers in the poking and raycasting interactions respectively.



Figure 3. a) Devices used during the Phase II study, where participants performed *selection* tasks in augmented reality using different interaction techniques. The devices included are the: functional Near-Infrared Spectroscopy device (NIRSport2, NIRx Medical Technologies LLC,NY,USA) to study functional connectivity and neural activation(top), an Augmented Reality HMD (HoloLens2, Microsoft, USA) to perform the AR-based task (middle) and, an Actiheart Device (Actiheart5, CamNTech Inc.,UK) that recorded electrocardiogram signals to quantify physiological load (bottom) b) fNIRS probe montage highlighting the sensors (pink circle) and detectors (green circle) are connected by solid black lines indicating the channels



Figure 4. Phase II Study: Participants were equipped with bioinstrumentation and were administered a 3minute baseline. Next, they completed a training task, consisting of six trials each of poking and raycasting interactions in AR. The adaptation model was run on their temporal performance data to classify them either as low or high performers. Low performers completed three additional trials of poking and/or raycasting interactions (based on the model determination). All participants then completed an evaluation task consisting of three trials in both interaction types. Subjective surveys were administered after each interaction type and after each type of task.



Figure 5. Comparison of time variables across low (white) and high performing (black) participants during the training and evaluation tasks for poking (left) and raycasting (right) interactions (mean(SE)) for Phase-II. In both interaction types, high performers were significantly quicker than low performers in the training task (p < 0.05). Results from the linear mixed-effects model showed significant task and group (low vs. high) main effects (p < 0.05) as well as significant interaction effect (p < 0.05) for both the poking and raycasting interactions.



Figure 6. (a-f) Brain Activation (mean(SE)) of low (white columns) and high (black columns) performers during poking interactions across different regions of interest (orange highlights), * denotes significant differences between groups (p < 0.05); While brain activation levels were comparable between groups during the training phase, low performers exhibited higher activation in frontal eye field (FEF/CG) region during the evaluation phase. (g) Group differences in functional connectivity strengths during training (top) and evaluation (bottom) phases (orange lines denote significantly (p < 0.05 stronger connections in high performers compared to low performers). High performers exhibited greater functional connectivity between brain regions than low performers during training, indicating higher neural efficiency. The adaptation trials enhanced functional connectivity patterns.



Figure 7. (a-f) Brain Activation (mean(SE)) of low (white) and high (black) performers during ray casting interactions across different regions of interest (orange highlights), * denotes significant differences between groups (p < 0.05); brain activation levels for the right motor cortex was higher in the high performers, however low performers exhibited higher activation in frontal eye field (FEF/CG) and supplementary motor (SMA) regions during the evaluation phase. (g) Group differences in functional connectivity strengths during training (top) and evaluation (bottom) phases (orange lines denote significantly (p < 0.05 stronger connections in high performers compared to low performers). High performers exhibited greater functional connectivity between brain regions than low performers during training, indicating higher neural efficiency. While the adaptation trials improved functional connectivity in low performers during evaluation, they still exhibited weaker connections than high performers during evaluation.



Figure 8. Mean (SE) ratings on the NASA TLX subscores for Poking (top row) and Raycasting (bottom row) for low (white columns) and high (black columns) performers during the training and evaluation phases of the Phase-II study; * denotes significant differences between groups (p < 0.05). While perceived workload across all subscores were comparable between low and high performers during training and evaluation phases of poking (top), low performers reported greater perceptions of mental demand, frustration, and poorer performance than high performers during training on the raycasting interactions (bottom). The adaptation trials successfully mitigated the perceptions of these demands between groups during evaluation.

Figure Captions:

- Figure 1 Phase I Study Workflow: Twenty-seven participants started with a (a) Training Task comprising of 8 trials of sequentially selecting buttons using the poking and raycasting interactions. (b) The bi-variate performance metric was used to compute the time (overall median completion time) and consistency (RMSE) values for each user. (c) Using the data from the twenty-seven participants, we trained adaptations models for both the interaction types (only poking show in figure) using the k-means clustering algorithm. We obtained clusters for low (white circles) and high (black circles) performing users.
- Figure 2 Training results of the poking (left) and raycasting (right) adaptation models using the *k*-means clustering algorithm. Twenty-seven participants from Phase-I study were clustered into low (white circles) and high (black circles) performing groups, with the cluster centroids marked in red diamond markers. Nine and fifteen participants were labeled as low performers in the poking and raycasting interactions respectively.
- Figure 3 a) Devices used during the Phase II study, where participants performed selection tasks in augmented reality using different interaction techniques. The devices included are the: functional Near-Infrared Spectroscopy device (NIRSport2, NIRx Medical Technologies LLC,NY,USA) to study functional connectivity and neural activation(top), an Augmented Reality HMD (HoloLens2, Microsoft, USA) to perform the AR-based task (middle) and, an Actiheart Device (Actiheart5, CamNTech Inc.,UK) that recorded electrocardiogram signals to quantify physiological load (bottom) b) fNIRS probe montage highlighting the sensors (pink circle) and detectors (green circle) are connected by solid black lines indicating the channels
- Figure 4 Phase II Study: Participants were equipped with bioinstrumentation and were administered a 3-minute baseline. Next, they completed a training task, consisting of six trials each of poking and raycasting interactions in AR. The adaptation model was run on their temporal performance data to classify them either as low or high performers. Low performers completed three additional trials of poking and/or raycasting interactions (based on the model determination). All participants then completed an evaluation task consisting of three trials in both interaction types. Subjective surveys were administered after each interaction type and after each type of task.
- Figure 5 Comparison of time variables across low (white) and high performing (black) participants during the training and evaluation tasks for poking (left) and raycasting (right) interactions (mean(SE)) for Phase-II. In both interaction types, high performers were significantly quicker than low performers in the training task (p < 0.05). Low performers were significantly quicker in the evaluation after receiving additional training. Results from the linear mixed-effects model showed significant task and group (low vs. high) main effects (p < 0.05) as well as significant interaction effect (p < 0.05) for both the poking and raycasting interactions.
- Figure 6 (a-f) Brain Activation (mean(SE)) of low (white columns) and high (black columns) performers during poking interactions across different regions of interest (orange highlights), * denotes significant differences between groups (p < 0.05); While brain activation levels were comparable between groups during the training phase, low performers exhibited higher activation in frontal eye field (FEF/CG) region during the evaluation phase. (g) Group differences in

functional connectivity strengths during training (top) and evaluation (bottom) phases (orange lines denote significantly (p < 0.05 stronger connections in high performers compared to low performers). High performers exhibited greater functional connectivity between brain regions than low performers during training, indicating higher neural efficiency. The adaptation trials enhanced functional connectivity in low performers during evaluation, with both groups exhibiting nearly comparable connectivity patterns.

- Figure 7 (a-f) Brain Activation (mean(SE)) of low (white) and high (black) performers during raycasting interactions across different regions of interest (orange highlights), * denotes significant differences between groups (p < 0.05); brain activation levels for the right motor cortex was higher in the high performers, however low performers exhibited higher activation in frontal eye field (FEF/CG) and supplementary motor (SMA) regions during the evaluation phase. (g) Group differences in functional connectivity strengths during training (top) and evaluation (bottom) phases (orange lines denote significantly (p < 0.05 stronger connections in high performers compared to low performers). High performers exhibited greater functional connectivity between brain regions than low performers during training, indicating higher neural efficiency. While the adaptation trials improved functional connectivity in low performers during evaluation, they still exhibited weaker connections than high performers during evaluation.
- Figure 8 Mean (SE) ratings on the NASA TLX subscores for Poking (top row) and Raycasting (bottom row) for low (white columns) and high (black columns) performers during the training and evaluation phases of the Phase-II study; * denotes significant differences between groups (p < 0.05). While perceived workload across all subscores were comparable between low and high performers during training and evaluation phases of poking (top), low performers reported greater perceptions of mental demand, frustration, and poorer performance than high performers during training on the raycasting interactions (bottom). The adaptation trials successfully mitigated the perceptions of these demands between groups during evaluation.

Notes on contributors:

Shantanu Vyas is a Ph.D. student in the J. Mike Walker '66 Department of Mechanical Engineering at Texas A&M University. His research interests include human-computer interaction, augmented and virtual reality, applied AI and generative design.

Shivangi Dwivedi is a master's student at Texas A&M University in the Industrial & Systems Engineering Department. Her research interests are machine learning, signal processing, human computer interactions, wearables, usability, and accessibility.

Lindsey J. Brenner is a Project Manager for the NeuroErgonomics Laboratory at Texas A&M University. Her research interests include human factors and the integration of emerging technologies in the workforce.

Isabella Pedron is a Chemical Engineer alumnus from Texas A&M University. Her research interests include addressing healthcare issues through implementation of human factors and human augmentation technologies.

Dr. Joseph L. Gabbard is a Professor of Human Factors at Virginia Tech's Grado Department of Industrial & Systems Engineering. His research focuses on the connections between user interface design and human performance, and specifically the development of techniques to design and evaluate AR and VR user interfaces.

Dr. Vinayak R. Krishnamurthy is an Associate Professor in the J. Mike Walker '66 Department of Mechanical Engineering and Computer Science and Engineering by Affiliation at Texas A&M University. His research is at the intersection of geometric & topological modeling, human-computer interaction, and product & engineering design.

Dr. Ranjana K. Mehta is Professor in the Department of Industrial and Systems Engineering at the University of Wisconsin Madison. Her research focuses on understanding, monitoring, and augmenting mind-motor-machine interactions using brain-behavior approaches in high-risk environments).