Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Chiossi, Francesco; Turgut, Yagiz; Welsch, Robin; Mayer, Sven

Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality

Published in:
Proceedings of the ACM on Human-Computer Interaction

DOI:
10.1145/3604243

Published: 12/09/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Chiossi, F., Turgut, Y., Welsch, R., & Mayer, S. (2023). Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, *7*(MHCI), Article 196. https://doi.org/10.1145/3604243

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality

FRANCESCO CHIOSSI, LMU Munich, Germany
YAGIZ TURGUT, LMU Munich, Germany
ROBIN WELSCH, Aalto University, Finland
SVEN MAYER, LMU Munich, Germany

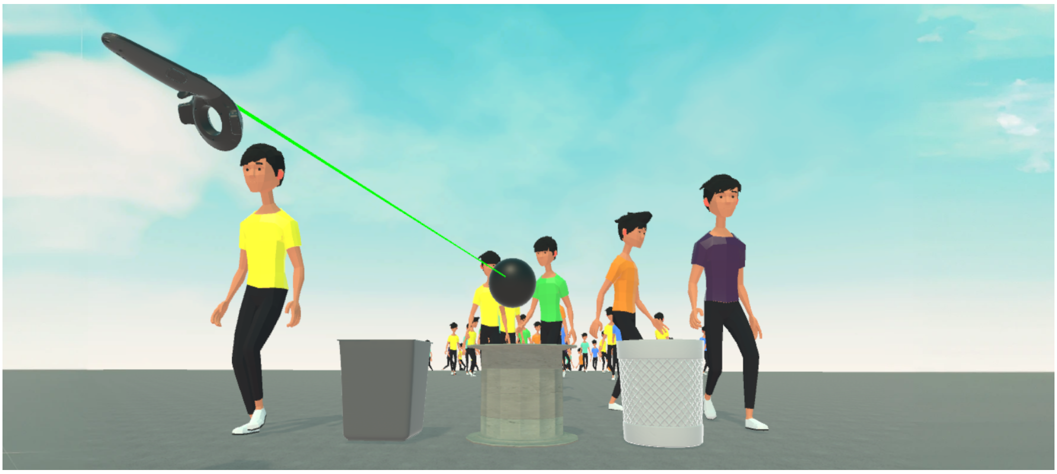


Fig. 1. VR view capture of a single trial of the VR N-back (N-Back=2) while NPCs distractors are approaching the user. The VR N-Back task required participants to drop a sphere into the appropriate bucket. If the sphere matched the color of the previous sphere, participants placed it in the appropriate bucket. Otherwise, the sphere should be placed into the left bucket.

Biocybernetic loops encompass users' state detection and system adaptation based on physiological signals. Current adaptive systems limit the adaptation to task features such as task difficulty or multitasking demands. However, virtual reality allows the manipulation of task-irrelevant elements in the environment. We present a physiologically adaptive system that adjusts the virtual environment based on physiological arousal, i.e., electrodermal activity. We conducted a user study with our adaptive system in social virtual reality to verify improved performance. Here, participants completed an n-back task, and we adapted the visual complexity of the environment by changing the number of non-player characters. Our results show that an adaptive virtual reality can control users' comfort, performance, and workload by adapting the visual complexity based on physiological arousal. Thus, our physiologically adaptive system improves task performance and perceived

Authors' addresses: Francesco Chiossi, LMU Munich, Germany, francesco.chiossi@um.ifi.lmu.de; Yagiz Turgut, LMU Munich, Germany, y.turgut@campus.lmu.de; Robin Welsch, Aalto University, Helsinki, 02150, Finland, robin.welsch@aalto.fi; Sven Mayer, LMU Munich, Munich, 80337, Germany, info@sven-mayer.com.

workload. Finally, we embed our findings in physiological computing and discuss applications in various scenarios.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: Physiological Computing, Adaptive Systems, Virtual Reality, Working Memory, Visual Complexity, Electrodermal Activity

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

1 INTRODUCTION

Physiological computing employs physiological data as system inputs in real-time for user-state detection and system adaptation [37]; for example, inferring increased arousal from changes in skin conductance level [101] or electroencephalographic (EEG) alpha frequency for attentional demands [32]. Here, the motivational intensity model [125] states that cognitive engagement increases in proportion to task demand. However, when the task difficulty is overwhelming, the engagement drops as users withdraw from the task. Physiologically-adaptive systems using the motivational intensity model focused on the adaptation of the primary task demands using functional near-infrared spectroscopy (fNIRS) [1, 39], electroencephalographic (EEG) alpha and theta frequencies [36] and electrodermal activity (EDA) [21]. Dey et al. [32] adapted the primary task difficulty using EEG alpha oscillations by altering the number and properties of distractors in the environment. Similarly, Chiossi et al. [21] adapted a secondary task, a visual detection task, by reducing the number of visual distractors based on EDA. Here, adaptation naturally comes at the cost of task-performance let it be in a primary or secondary task. However, related work has shown that task-irrelevant and visual features affect task performance [56, 99, 105], such as visual complexity [98] in virtual environments. As such, we argue that a more efficient approach is to adapt the visual complexity of the virtual environment and not the task itself to improve performance and UX in Virtual Reality (VR).

Visual complexity is defined as the amount of detail, clutter, and objects in an environment [98]. This fidelity, on the one hand, makes VR experiences engaging; however, on the other hand, studies have found that increased visual complexity in VR environments can lead to delayed reaction times [102], increased physiological arousal, and cognitive overload [80, 102, 104]. Thus, increasing visual complexity in VR environments can overwhelm the user, hinder their goals and decrease their task performance [109]. Therefore, task-irrelevant visual features can divert users' attention from the primary task, which impairs task engagement and performance [9], especially in situations when there is a high visual overlap between a target and distracting information [97]. This effect of visual complexity on performance has been shown both from social stimuli, such as faces [127], and low-level visual distractors [77, 87]. With this in mind, an adaptive VR system should optimize visual complexity to a level that keeps the user engaged but, at the same time, does not overwhelm the user.

In this work, we investigate the feasibility of adapting the visual complexity of the environment to improve the performance of the main task. Overcoming the limitation of prior work [21] who need an initial setup recording; we designed and implemented a continuously physiologically-adaptive system to enhance task performance. Moreover, in contrast to prior work [96] that limited the adaptation to features of the primary task or added a secondary task, we adapted the visual complexity of the VR environment only. Building on prior work [21], we adapt the visual complexity of the surrounding VR environment based on the variation of physiological arousal measured by the tonic component of EDA, i.e., Skin Conductance Level (SCL). We evaluated our physiologically-adaptive system in a user study with 20 participants. As the main task, we used a cognitively demanding 2-back task while we manipulated visual complexity by non-player characters (NPCs) passing by the participant. Finally, the physiologically-adaptive system adapts the number of NPCs passing by to reduce or increase the visual complexity of the environment.

Our results show that an adaptive system that optimizes for visual complexity based on physiological arousal modeled after the motivational intensity model increased behavioral performance and improved perceived workload as compared to an adaptation algorithm that used a reversed motivational intensity model [125]. In this work, we make the following contributions: we investigated the effect of visual complexity levels over the attentional load over physiological arousal, behavioral performance and subjective workload (i) and evaluated the feasibility of real-time skin conductance analysis to infer changes in visual complexity as shown by the associated increase in electrodermal tonic activity (ii). Furthermore, we adapted in an online manner the visual complexity as compared to an adaptive control algorithm, and we found increased working memory task performance and diminished perceived workload (iii). Lastly, we provide a freely available dataset with the co-registration of electrocardiogram and electroencephalogram signals (iii). We further discuss mobile domain applications for physiologically-adaptive VR systems based on physiological arousal detection, such as digital content blending and wayfinding, and Social VR based on the results. Finally, we highlight implications for adaptive algorithm designs for arousal detection based on EDA and future hybrid or sensor fusion approaches.

2 RELATED WORK

In the following, we give an overview of physiological computing in VR and summarize the relevance of visual complexity and its effect on task performance in VR environments.

2.1 Physiological Computing and Virtual Reality

The cybernetics approach had broad goals for using psychophysiological data, which ranged from developing new control channels to task adaptation in response to changes in workload [103]. To the present day, one of the most favorable settings for physiological computing is VR [49]. VR allows for online manipulation and adaptation of visualizations, virtual objects, and interactions that can resemble the real world as much as possible and cannot be realized in all physical spaces. Adding physiological sensing to VR can augment its ability to monitor overt human behavior and adapt accordingly.

Based on a physiological input and given the desired goal, physiologically adaptive VR systems can detect users' state based on user input and drive interaction features toward a (shared) goal. This combination of implicit physiological monitoring and VR environment adaptation can be defined as a closed-loop model [38]. Closed loop models provide the conceptual framework for adaptive systems designed to personalize software, visualizations, or interaction in real-time in a personalized manner [22]. In human-computer interaction contexts, adaptive VR systems dynamically adjust the system's parameters to the current task or user. The adaptive controller in a closed-loop system acts as a dynamic mechanism, responding to changing inputs under a specific standard or goal. The ultimate purpose has been to improve task performance, already since Pope et al. [103] seminal work in which a biocybernetic loop was designed to sustain a high level of engagement with task performance.

So far, adaptive VR systems followed one approach: dynamic task difficulty adjustments based on different physiological user inputs to support users' performance in various settings. This is because tailoring difficulty might be a preferable outcome in gaming [36], adaptive training for cognitive load estimation [32], in single [126] and multitasking settings [21]. However, such adaptive systems only managed to index demand and motivation for adaptation purposes [32, 36], support performance in the form of reaction time but not accuracy [126] or, ultimately, decrease perceived workload and improved user experience [21].

Although there have been significant steps forward in integrating biocybernetics loops based on physiological computing in VR, the ultimate goal of supporting task performance needs further research. In addition, there is space for adapting the virtual environment to support task performance without directly manipulating task features or demands. Thus, we focus on adapting task-irrelevant features of the environment to investigate their effect on performance and adaptation. Therefore, in the next section, we focus on the relevance of task-irrelevant visual features, i.e., visual complexity in VR environments and its effect on behavioral performance.

2.2 Visual Complexity and Task Performance in Virtual Reality

Visual realism is of central relevance for VR systems in the context of visual stimuli display [15]. VR systems aim to provide users with realism as well as real-time updates of visual stimuli. Ideally, this fidelity of visual stimulation would allow achieving a comparable to real and improved VR experience [48]. These can be especially relevant for training simulations or knowledge work to ensure proper engagement in VR, improved user experience, and translation to real-world performance in the same task [48]. However, when we increase the level of detail, realism, and amount of virtual elements, we inherently contribute to the amount of visual complexity present in the VR environment [98]. Thus, we have to consider that our perceptual and attentional processing capacities can only be allocated to a small amount of the information present in a visual scene, and thus complex visual scenes increase mental effort in visual search [33] and executive tasks [119], such as working memory tasks [29].

Working memory (WM) is involved in comprehension, reasoning, planning, and learning [14]. Studies on the effects of visual distraction on visual WM demonstrate how vulnerable our task performance is [2, 46, 89]. Thus, when we are cognitively engaged in a task in VR, visual distractors lead to increased response times [102], discomfort [59], increased physiological arousal, and cognitive overload [80, 102, 104]. In contrast, if a VR environment is overly automated and low in fidelity, users may become disengaged and drowsy.

WM overload is generally inserted in a framework where all its functionalities (store, update, retrieve and information manipulation) are degraded by the processing demands of the task [66, 94]. This relationship is in line with the motivational intensity model, which describes a unitary perspective where relationship task engagement and task demands are linked on a behavioral and physiological level. When task demands are fixed, and participants are confident of achieving a successful level of performance, the relationship between mental effort and task demand is proportional [106]. However, if demand rises and success is unlikely, effort investment decreases and perceived workload increases. Therefore, working memory capacity is associated with three outcomes: impaired performance, an increased perceived workload, and reduced engagement.

Specifically, physiological arousal correlated with increased engagement, as is typically shown in studies that employed EDA recordings. For example, Fairclough and Venables [40] reported increased SCL with lower task engagement and higher stress in high-demand, multi-component tasks. Similarly, in a target detection task in immersive VR, an increased perceptual load elicited an increased mean SCL [83]. Thus, a physiologically-adaptive VR system that dynamically adjusts surrounding visual complexity based on physiological arousal can aid the user in processing information and increase task performance.

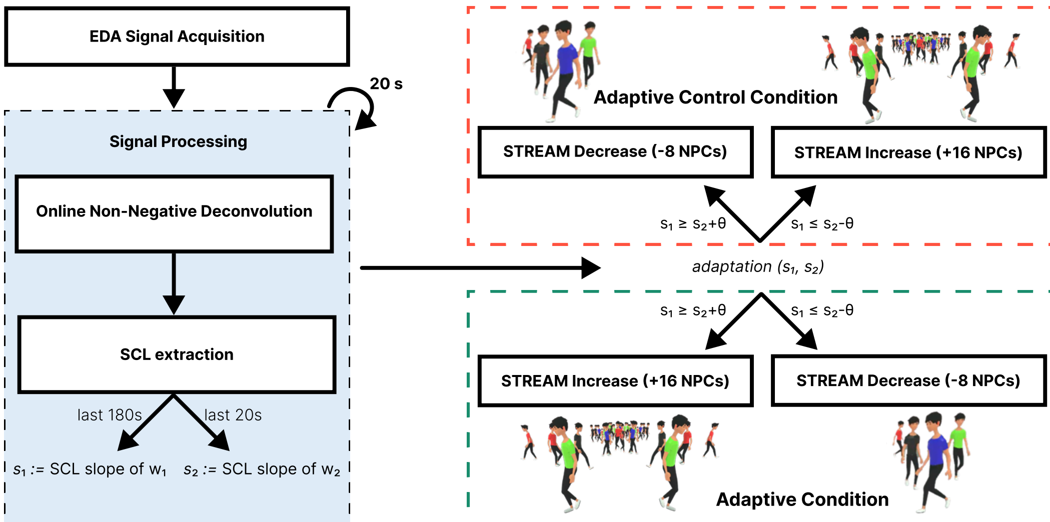


Fig. 2. Components of the two adaptive conditions. In both conditions, the Stream adapts according to changes in the slope between the tonic averages s_1 and s_2. Here, s_1 represents the slope between first 20sec of w_1 and in the last 20sec of w_1, while s_1 is the slope between first 20sec of W_1 and in the last 20sec of w_2. In the adaptive condition, if the slope of the SCL in w_1 is smaller than in the one of w_2 ($s_1 < s_2$), then 8 Not Playable Characters(NPCs) are removed from the scene. Alternately, in the case $s_1 > s_2$, 16 NPCs are added. The motivational intensity model inspires adaptive conditions and aims to support task engagement. The adaptive control condition follows the opposite logic.

3 ARCHITECTURE OF THE PHYSIOLOGICALLY-ADAPTIVE VR SYSTEM

The closest work to our system is by Chiossi et al. [21]; they based their adaptation on raw EDA data for task difficulty adaptation. Specifically, they adapted the secondary task difficulty of a visual detection task. Therefore, they computed the EDA slope of a 180sec baseline recording before the study. With this, they adapted the difficulty based on the difference between the EDA baseline slope and the online EDA slope computed every 20sec. A major drawback of the work by Chiossi et al. [21] is that they tend to compare to an old baseline, e.g., after one hour of online adaptation, the baseline recording is also one hour old. Thus, they are not accounting for effects like EDA saturation and drifts [50, 61]. In this work, we overcome this limitation by eliminating the baseline recording and, thus, only including the latest EDA data, i.e., we adjust the amount of NPCs comparing the EDA signal to the averaged EDA signal acquired in the previous 20sec. This choice allows for increased reliability of the signal on which the adaptation is based on and better deployment outside of the lab. The architecture of adaptive and control adaptive conditions is depicted in Figure 2.

First, we preprocessed the raw EDA signal by removing low-frequency noise with a fourth-order Butterworth filter with a 3 Hz high-pass cutoff. Second, we used non-negative deconvolution analysis to decompose the signal and extract its tonic component, i.e., SCL [8].

Second, we apply a rolling window approach for the adaptation. We use a long-term window for low-frequency changes in SCL of window length w_1 (in our study, w_1 is 180sec long). Additionally, we use a short window for high-frequency changes with window length w_2 (in our study, w_2 is 30sec long). Resulting in the three points t_0, t_{-1}, and t_{-2}, see Figure 2. We calculate the SCL level at all three points t by averaging the SCL values from t to $t - \epsilon$. Here, ϵ allows averaging over some

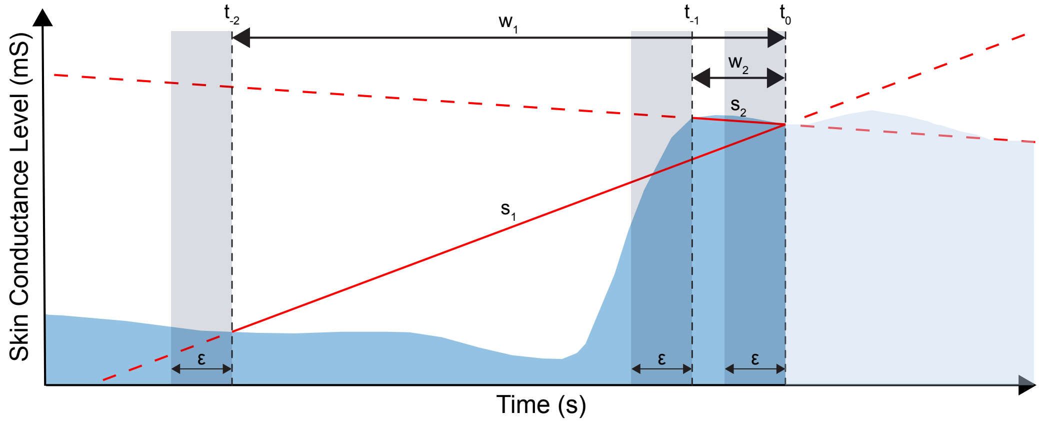


Fig. 3. An overview of how the adaptation is computed. In this case s_2 shows an increased slope as compared to s_1. This result would lead to a decrease in Stream in the adaptive condition or an increase of Stream in the control adaptive condition.

data before the points to stabilize the value. Thus, in the following, the SCL level at t_x is the mean value of $t_x - \epsilon$ until t_x, defined as $SCL(t_x)$.

Thirds, we compute the slopes of the change in SCL in the two windows, i.e., s_1, and s_2. We calculate s_1 from the average tonic value of t_{-2} and t_0, $s_1 = (SCL(t_0) - SCL(t_{-2}))/w_1$. Moreover, we calculate s_2 the same way but using t_{-1}, t_0, and w_2.

Finally, we compare the low-frequency slope s_1 to the high-frequency slope s_2. We compare their difference and drive the adaptation as the following:

$$adaptation(s_1, s_2) = \begin{cases} \texttt{increase} & \text{if } s_1 \leq s_2 - \theta \\ \texttt{decrease} & \text{if } s_1 \geq s_2 + \theta \end{cases} \quad (1)$$

Here, the threshold parameter θ allows the adaptation to occur within a certain variance, preventing quick and unstable adaptations. Lastly, Equation 1 is only used for adaptation every 20s, as depicted in Figure 2.

4 USER STUDY

We state the following research questions, informed by related work:

RQ1: Can online dynamic adjustments of visual complexity based on a peripheral physiological measure of arousal support task performance?

RQ2: Does adaptation based on the motivational intensity model support task performance?

The present experiment was designed and conducted to evaluate if a physiologically-adaptive system, based on SCL, supports the user's behavioral performance as compared to not adaptive ones and an adaptive control condition. As a primary task, we chose the established N-Back task [58] as adapted from [21] in an immersive VR environment. This task required a constant information update stored in working memory for each trial and constant attention to spheres presented, and maintenance of previously presented information.

4.1 Design

To examine differences in performance, perceived workload, and UX, we performed a within-subjects study for the system's adaptability factor (Physiologically-Adaptive system vs. Reverse adaptive system vs. Not Adaptive system). The order of all conditions was randomized across participants. The experiment encompasses seven blocks, of which five have a stable Stream, and two have an adaptive Stream. Here, Stream is defined as the amount of non-player characters (NPCs) entering the VR scene per minute, see Figure 4. The "adaptive" experimental condition involved the manipulation of the visual complexity (i.e., the Stream of NPCs) through changes in the participant's tonic EDA level as measured by Skin Conductance Level (SCL) [43]. We evaluated four aspects of the system: (i) N-Back task performance, (ii) average SCL, (iii) overall perceived workload, subjective engagement (in-game Game Experience Questionnaire), and UX (ad-hoc survey) across conditions. Specifically, as a control condition to investigate RQ2, we implemented a reverse adaptation that follows a reversed algorithm based on the motivational intensity model [125] that increases visual complexity and physiological arousal. In the remaining five non-adaptive conditions, participants were stimulated with fixed Stream of 24, 110, 191, 270, and 347 of NPCs entering the scene per minute. Participants were not aware whether they would experience the adaptive condition, control adaptive condition, or one of the non-adaptive conditions to avoid biases in performance and subjective ratings.

4.2 Physiological Data Recording

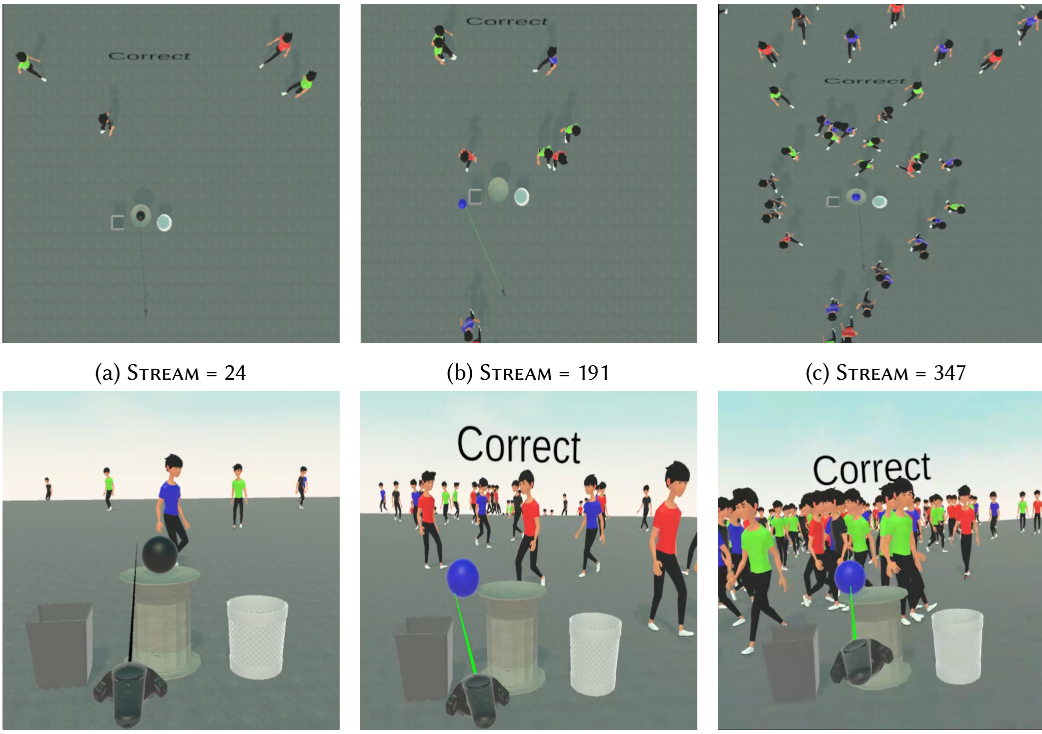
We followed the recent guidelines and framework for the human-computer interaction [6] for EDA data recording. EDA was recorded using standard Ag/AgCl electrodes (7 mm surface diameter) placed on the distal surfaces of the middle phalanges of the index and middle fingers of the participant's non-dominant hand. Data collection started with an electrolyte solution application (0.5% $CaCl_2$) over acquisition sites to ensure proper hydration and minimize individual differences' effects. After the application, the participant waited 10 minutes before the electrodes were attached to the participant's phalanx utilizing double-sided adhesive collars. For data acquisition and amplification, we used a LiveAmp from BrainProducts GmbH amplifier combined with a Sensor and Trigger Extension for a GSR-Module, which exploited the exosomatic recording principle with direct current (DC) with a constant voltage of 0.5 V. Sampling rate was set at 250 Hz. For offline data preprocessing, we followed the same pipeline as for the adaptive and adaptive control conditions via the Neurokit toolbox [79]. After non-negative deconvolution analysis, we derived two metrics of physiological arousal: the average amplitude of Non-Specific Skin Conductance Responses (nsSCRs) and the average tonic SCL. We identified nsSCRs peaks using a .05 mS threshold value, following the recommendations by the Society for Psychophysiological Research [43].

4.3 Apparatus

We implemented the VR environment and tasks using Unity 3D (Version 2020.1.8f1). We acquired three physiological measurements: EDA signal using a GSR Module (BrainProducts GmbH, Germany, 250 Hz), ECG via PolarH10 chest strap (Polar, Finland, 130 Hz), and EEG signal (DSI-VR 300, Wearable Sensing, San Diego, CA, 300 Hz). For the scope of this paper, we will only report and analyze results on EDA. Physiological data were streamed within the Unity VR environment within the Lab Streaming Layer (LSL) framework[1] to the acquisition PC (Intel Core i7 with 3.00GHz, 32GB RAM).

The number of NPCs in the adaptive condition was adjusted using Equation 1. After multiple testing runs, we set w_1 to be 180sec long and w_2 30sec long. Further, we determined ϵ to be 20sec. In detail, sixteen NPCs were added to the VR scene if the s_2 slope was smaller than the s_1 slope. On the other hand, eight NPCs were removed if s_2 slope was greater than s_1 slope, which is indexing increased arousal state. Those adjustment values represent 10% and 5% of the starting Stream value for adaptation, respectively. Several test sessions were employed to determine these settings

[1] https://github.com/labstreaminglayer/



(a) Stream = 24 (b) Stream = 191 (c) Stream = 347

(d) Stream = 24 (e) Stream = 191 (f) Stream = 347

Fig. 4. VR capture of a single trial of the VR n-back from a birds-eye perspective in the first row, and first-person view in the second row. In (a) is depicted the condition with low visual complexity with Stream = 24. In (b) is depicted the condition of moderate visual complexity with Stream = 191 and lastly in (c), the highest visual complexity with Stream = 347.

empirically. Regarding the control adaptive condition, we used the opposite algorithm. Therefore, if s_2 was smaller than s_1 we increased the Stream of 16 NPCs, while if it was bigger than s_1 we decreased of 8 NPCs, see Figure 2. This design is inspired by the previous work of Chiossi et al. [21], which employed a +4/-2 for task-relevance elements adaptation. Here, we employed a +16/-8 design to account for task-irrelevance of the distractors based on previous work [19]. Regardless of the adaptive or control adaptive condition, or the physiological activity of the participants, NPCs were spawned and removed from the VR environment within a range of 24-347.

To extract the tonic component via non-negative deconvolution analysis, we streamed EDA raw data via a Transmission Control Protocol (TCP) /Internet Protocol (IP) client to the TCP/IP server developed by Python network programming. This implementation allowed us to pass forward and backward data between LSL and the VR Unity environment. To ensure that the time of the VR Unity scene is synchronized with the time of the bridge server's operating system, we synchronized both systems with a Network Time Protocol (NTP) service. We preprocessed online EDA raw data using the Neurokit Python Toolbox [79] within the Python Client-Server bridge service.

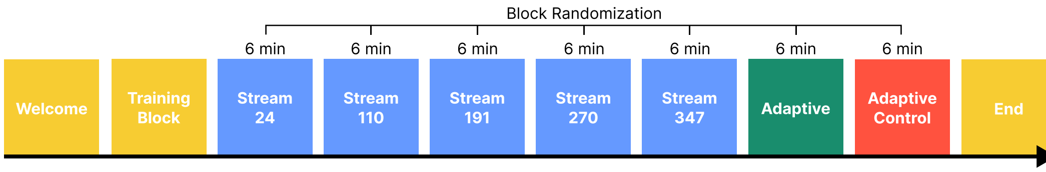


Fig. 5. The experiment encompassed seven different blocks for data recording (blue, green, and red). Participants filled in NASA-TLX [54], GEQ subscales, and ad-hoc UX surveys between these blocks. Participants started the experiment with a training phase and then experienced the seven blocks in randomized order.

4.4 Task

The task was adapted from the N-Back in the study by Chiossi et al. [21]. Participants were immersed in a neutral VR environment and presented with a marble-like pillar and two buckets placed on the left and the right, respectively. Over the pillar, spheres were spawned in four possible colors (green, red, blue, and black), following [88]. The color sequence was randomly generated. Participants had to grab spheres with an HTC VIVE controller and drop them into the correct buckets. If the sphere matched the color of the sphere presented two steps before, the sphere has to be placed in the right bucket. If the sphere's color did not match the sphere's color two steps before, the sphere had to be put on the left bucket. Each sphere's appearance was signaled by a tone (800 Hz). Then, in order to avoid making an error, the participant had to pick up the sphere within 4sec. New spheres would appear when the current sphere was placed into one of the two buckets or after 4sec. Here, participants received accuracy feedback every 20 spheres and were instructed to maintain 90% performance.

4.5 Procedure

Upon arrival, we informed participants of the study procedure and answered any open questions, followed by signing the informed consent form. Next, we attached the participants' EDA sensor, EEG-VR headset, and ECG chest strap. We collected ECG data positioned over the xiphoid process of the sternum beneath the chest muscles. Participants performed a VR visual working memory task, i.e., the N-Back task (N=2). The experimental procedure encompassed a training phase and the execution of the seven experimental blocks in randomized order. The experimental procedure is depicted in Figure 5.

The study began with a *training phase* to familiarize with the VR environment. Here, participants practiced with the 2-back task until they reached an accuracy level of at least 90% within a sequence of 80 spheres. The *experimental phase* consisted of seven experimental blocks, lasting six minutes each. Between the blocks, we asked participants to complete the raw NASA-TLX [54], the Game Experience Questionnaire (in-Game Core Module) [57], and two system's UX ad-hoc surveys on a 7-point Likert Scale 1) "I would like to use the system in the future," 2) "The flow of the Not-Playable Characters was appropriate." The experiment lasted for around one hour, and the participant received monetary compensation of 15 Euros.

4.6 Participants

Twenty participants (M_{age} = 26.05, SD_{age} = 3.62; *male* = 15, *female* = 5) took part in our study. We recruited the participants using our institutional mailing lists and social networks and using convenient sampling. We did not recruit participants that experienced intense physical activity or consume any caffeine or nicotine in the 3-hour pre-study period, according to Babaei et al. [6]. None of the participants reported a history of neurological, psychological, or psychiatric symptoms.

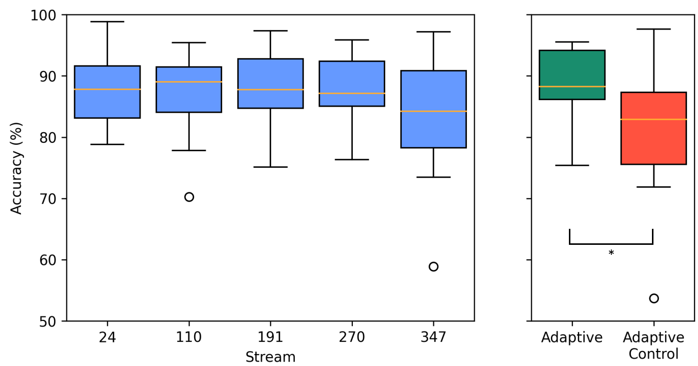


Fig. 6. Box-plots for accuracy in the N-back task as a function of Stream in the stable and for adaptive conditions, respectively depicted on the right and on the left . Here, we did not find any significant differences across Stable Stream conditions. In the adaptive conditions, the adaptive control showed significantly decreased accuracy on an average of 7.59% less than the adaptive condition.

However, due to technical difficulties, after EDA signal inspection, we removed 3 participants, leading to $N = 17$. According to the fast-track conditions of the local institutional ethics board, the study qualified as fast-track, i.e., participants are not subject to any risk (e.g., deception, stress beyond normal levels, recording of sensitive information).

5 RESULTS

We analyzed indicators of physiological arousal, performance, and subjective experience across the adaptive and non-adaptive conditions. We first briefly present the results for our system's non-adaptive levels of Stream and then investigate whether the adaptive systems can support performance, increase user experience, and relatively result in a lower level of perceived workload. Normality of residuals for Analysis of Variance (ANOVA) models was checked via D'Agostino normality test [28] as Shapiro-Wilk test is too sensitive for n >50. Upon normality testing, we use one-way repeated measures ANOVAs (RM ANOVA) for parametric analysis or a Friedman's rank sum test for not-normally distributed data. Furthermore, for post-hoc comparisons, we use Conover's tests [26]. In this section, we perform statistical testing across all conditions i.e., stable Stream and adaptive Stream conditions.

5.1 Stable Stream Conditions

5.1.1 Behavioral Performance. We conducted a Friedman test, as the D'Agostino normality test showed that performance data are not normally distributed ($\chi = 9.65$, $p < .05$). We computed as accuracy the proportion of times the sphere was placed in the correct bucket overall. This analysis revealed that the accuracy in the N-Back task was not significantly influenced by Stream ($\chi = 2.49$, $p = .64$), see Table 1.

5.1.2 Electrodermal Activity Results.

Skin Conductance Level. Due to a D'Agostino tested normality assumption satisfaction ($\chi = 2.93$, $p = .23$), a RM ANOVA showed that average SCL was significantly influenced by Stream ($\chi = 2.49$, $p < .05$). Pairwise comparisons via a Conover test with Holm's correction showed that the highest condition of Stream (347) significantly decreased SCL as compared to the lowest condition of

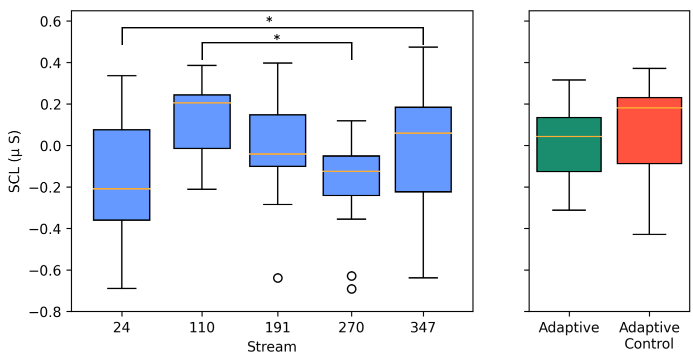


Fig. 7. Box-plots for Skin Conductance Level (SCL) as a function of STREAM. Results for stable manipulation of STREAM are shown on the left while adaptive and adaptive control conditions are depicted on the right. We found that STREAM (347) significantly increased the SCL as compared to STREAM (24) and that the second lowest condition of STREAM (110) was significantly higher than STREAM (270).

STREAM (24), while STREAM 110 showed increased SCL as compared to the condition where STREAM was equal to 270 NPCs entering the scene per minute (all $p =< .005$).

nsSCRs. Since the normality assumption was violated ($\chi = 15.78$, $p < .05$), we conducted a Friedman test indicating that the average amplitude of nsSCRs is significantly influenced by STREAM ($\chi = 8.37$, $p = .07$). A Conover's post-hoc test did not show any significant differences between pairs.

5.1.3 *Subjective Results.*

Perceived Workload. The raw NASA-TLX scores did not meet the normality assumption (D'Agostino: $\chi = 64.74$, $p < .001$). We, thus, performed a Friedman test, reporting a significant effect ($\chi = 15.65$, $p < .001$). The Conover's post-hoc test revealed that the stable condition with the highest STREAM (347) showed significantly increased workload as compared to the lowest condition of STREAM (24) and to STREAM (190) (all $p =< .005$). Results are depicted in Figure 9.

GEQ. **Competence.** Since GEQ-Competence subscale ratings were not normally distributed (D'Agostino: $\chi = 50.28$, $p < .001$) we performed a Friedman test that detected a main effect of STREAM ($\chi = 11.39$, $p < .05$). Conover's post-hoc revealead only one significant difference in the

Table 1. Results of repeated-measures ANOVA or Friedman test on Behavioral Performance, subjective questionnaires and physiological arousal measures across conditions of stable STREAM.

	Stream 24		**Stream 110**		**Stream 191**		**Stream 270**		**Stream 347**		**ANOVA / Friedman**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***p***	***F / χ***
N-Back Accuracy [%]	88.14	6.22	87.12	87.53	88.21	5.90	87.33	5.52	83.51	9.55	.64	2.49
Raw NASA TLX	42.94	13.49	48.87	17.64	46.57	17.60	50.0	16.63	55.59	17.98	**<.01**	15.6
SCL	−.177	.32	.14	.18	−.002	.25	−.17	.322	−.02	.32	**<.01**	15.24
nsSCRs	−.002	.003	−.001	.002	.00	.004	−.002	.003	−.002	.004	.07	8.37
GEQ−Competence	2.12	.86	2.38	.93	2.47	.7	2.0	0.90	2.18	.88	**<.05**	.07
GEQ−Pos. Affection	1.50	1.03	1.24	1.17	1.41	1.15	1.29	1.19	1.65	1.34	.81	1.54
GEQ−Immersion	.94	.92	.62	.88	.85	.98	1.13	1.29	1.27	1.65	.46	3.59
Stream Appropriate	3.53	2.07	3.12	2.15	2.94	1.91	3.35	3.35	3.53	2.18	.44	.44
Desire To Use	4.18	2.07	3.82	2.27	3.77	2.11	4.06	2.05	4.12	2.37	.57	2.91

pairwise comparison between. Specifically, participant felt less competent in condition Stream (270) as compared to condition of decreased Stream (191). For the the descriptive statistics (*M* & *SD*) see Table 1 and Figure 10.

Positive Affection. GEQ-Positive Affection ratings were not normally distributed ($\chi = 14.46$, $p < .001$). Thus, we used a Friedman which did not show any effect of Stream ($\chi = 1.53$, $p = .81$).

Immersion. Similarly, the GEQ-Immersion subscale was not normally distributed ($\chi = 35.66$, $p < .001$) and did not show any significant effect of Stream for the Friedman test ($\chi = 3.59$, $p = .46$)

User Experience. For both UX ad-hoc surveys, no significant differences were found employing a Friedman test. Both Stream ($\chi = 12.68$, $p < .005$), and Desire to Use ($\chi = 31.14$, $p < .001$) deviated from the normal distribution. We summarized the results in Table 1.

5.2 Adaptive Stream Conditions

To further inspect the effect on physiological arousal, performance, and subjective experience of the two adaptive systems, we conducted t-tests or Wilcoxon tests depending on the normality assumption tested on residuals. Results are summarized in Table 2. On average, Stream for the adaptive condition lead to an average of 241.88 NPCs per minute (*SD*=30.31), while in the Adaptive Control condition, Stream stabilized on average of 216 NPCs (*SD*=32.82).

5.2.1 Behavioral Results. Given a violation of the normality assumption ($W = .91$, $p < .05$), a Wilcoxon test detected a significant difference in the behavioral accuracy for the N-Back, showing increased performance in the adaptive condition ($v = 135$, $p < .005$).

5.2.2 Electrodermal Activity Results. **Skin Conductance Level.** Due to a Shapiro-Wilk not significant results ($W = .96$, $p = .71$), a t-test showed that average SCL was not influenced by Stream ($t = -0.79$, $p = 0.44$).

nsSCRs. Similarly to SCL, also the distribution of nsSCRs was normally distributed ($W = 0.97$, $p = .84$), therefore a t-test showed that the amplitude of nsSCRs, indexing sympathetic physiological arousal, was impacted by Stream as indicated by an an increased in the adaptive control condition ($t = -2.33$, $p < 0.05$), see Table 2.

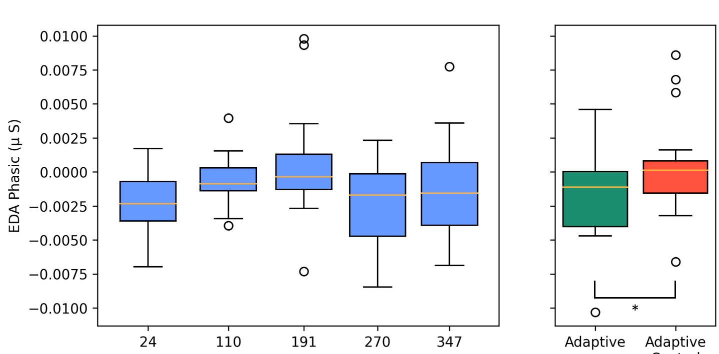


Fig. 8. Box-plots for mean amplitude of non-specific Skin Conductance Responses (nsSCRs) as a function of Stream. Here, we show the stable conditions on the left while the adaptive conditions are on the right. No significant differences were detected in pairwise comparison across stable conditions. In the adaptive conditions, the adaptive control condition showed significantly increased mean nsSCRs amplitude as compared to the adaptive condition .

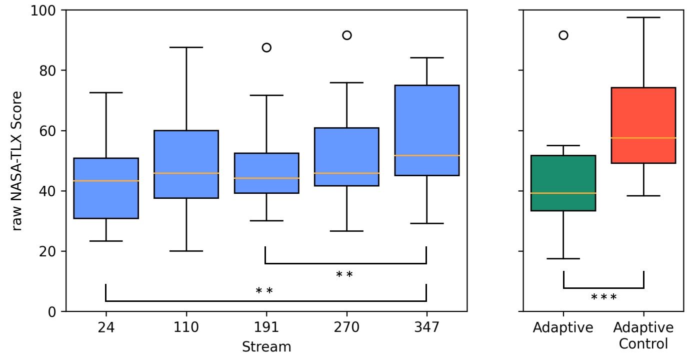


Fig. 9. Box-plots for raw NASA-TLX scores as a function of Stream. Stable Stream manipulation is shown on the left, and the adaptive conditions on the right. Those results mirrored the one for N-Back task accuracy. Here, participants reported significantly increased perceived workload in the condition of highest Stream compared to the lowest condition of Stream. We report similar results for the adaptive control condition that showed increased perceived workload compared to the adaptive condition.

5.2.3 Subjective Results. **Perceived Workload** The control adaptive condition significantly increased subjective workload, given a normal distribution ($W = .96$, $p = .79$), as shown by a t-test ($t = -6.18$, $p < .001$), see Table 2 and Figure 9.

GEQ Subscales. We did not find any significant results across all three GEQ subscales, i.e., Competence, Positive Affection, and Immersion; see Table 2 and Figure 10.

User Experience. The results for the ad-hoc UX surveys were similar to those of the GEQ subscales. Both items were non-normally distributed. However, no significant differences were reported between the two adaptive conditions; see also Table 2.

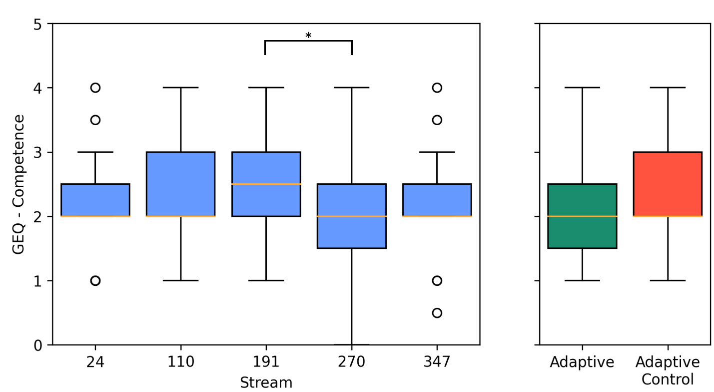


Fig. 10. Box-plots for GEQ-Competence scores as a function of Stream. Stable Stream manipulation is shown on the left and the adaptive conditions on the right. We did not find any significant differences across GEQ-Competence ratings.

Table 2. Results of Pairwise comparison for behavioral performance, subjective questionnaires and physiological arousal measures between the two adaptive conditions.

	Adaptive		Adaptive Control		*t*-Test/Wilcoxon		
	M	***SD***	***M***	***SD***	***t/V***	***p***	***d/r***
N-Back Accuracy [%]	88.56	5.96	80.97	10.07	135	**<.05**	.63
Raw NASA-TLX	42.16	16.34	62.55	17.99	−6.17	**<.001**	−0.37
SCL	.015	.17	.072	.24	−.79	.44	−.30
nsSCRs	−.002	.002	<.001	.004	−2.33	**<.05**	−.36
GEQ—Competence	2.15	.95	2.29	.89	17	.55	.12
GEQ—Positive Affection	1.41	1.03	1.50	1.40	23.5	.95	−.34
GEQ—Immersion	.67	.96	1.09	1.25	9	.22	−.50
Stream Appropriate	3.23	2.04	3.47	2.32	8.5	.75	−.14
Desire To Use	3.41	2.21	3.53	2.29	21	.71	−.33

5.3 Summary of Results

We first investigated the effect of different stable levels of visual complexity, i.e., Stream of NPCs, while participants were engaged in a WM Visual N-Back task over a series of dependent variables: behavioral accuracy, physiological arousal (SCL and nSCRs amplitude) and subjective workload, engagement, and UX. Second, we compared Adaptive Stream Conditions, where we adapted the visual complexity based on variations of SCL.

5.3.1 Stable Stream Conditions. We analyzed the effect of different levels of visual complexity, i.e., the amount of NPCs entering the VR scene per minute (Stream). We found no significant effect of Stream over behavioral accuracy in stable conditions. Regarding physiological arousal, we found that SCL was decreased in higher Stream conditions (347 and 270 NPCs per minute) as compared to the lower ones (24 and 110 NPCs per minute). Finally, participants reported significantly higher perceived workload in the condition with the highest Stream (347 NPCs) compared to all the other conditions. When investigating perceived engagement, we found that participants reported feeling less competent (GEQ - Competence) when the Stream decreased from 191 to 270 NPCs per minute. No further significant differences were detected.

5.3.2 Adaptive Stream Conditions. Based on the MIM model [125], we designed two physiologically-adaptive systems with two opposite architectures, i.e., adaptive and adaptive control systems. Overall, the adaptive system resulted in improved behavioral accuracy, decreased physiological arousal as shown by decreased nSCRs amplitude, and decreased perceived workload. No specific significant differences were detected in perceived engagement and UX between the two systems. Our results revealed how the adaptation of visual complexity based on physiological arousal, following the MIM model, results in improved WM performance, as compared to a reversed adaptation.

6 DISCUSSION

We presented a physiologically adaptive VR system that employed electrodermal activity to perform dynamic visual complexity adjustments to enhance task performance. We evaluated the effect of visual complexity, in the form of NPCs, on task performance, SCL, perceived workload, engagement, and UX. In the study, participants performed an N-Back task recruiting working memory resources. Here, we discuss our results regarding the outcome of our adaptive algorithms for modeling visual complexity and its effect on EDA. Then, highlight applications for arousal detection and implications of physiologically-adaptive systems.

6.1 Mapping Visual Complexity to Electrodermal Activity

Evaluating autonomous, closed-loop control systems raises issues for physiologically-adaptive systems design. The relationship between psychophysiological data and triggering adaptive responses at the interface necessitates careful design. One stage of this process is the generation of valid input measures and effectively categorising psychophysiological data in real time.

Our results indicate that we could model the visual complexity of the surrounding VR environment user in the N-back task using SCL. Participants who interacted with the adaptive system performed better than non-adaptive. In the reverse adaptation, we saw an inefficient user model, which led to a potentially harmful state of overload. In both of our adaptive conditions, participants' performance was influenced by the Stream adaptation at a behavioral, physiological, and subjective level. We found reciprocal effects on performance levels (see Figure 6), significant differences in the measure of physiological arousal, i.e., nsSCRs (see Figure 8) and lastly, in perceived workload (see Figure 9).

6.1.1 Impact of Visual Complexity on Task Performance. In line with the findings of previous work on the effect of visual complexity [3, 35, 104], we report that visual WM performance is decreased with an increasing amount of distractors surrounding the user when adapting the visual complexity based on arousal, following the MIM model. However, we did not find any main effect of Stream in the stable conditions. These results might be due to the high individual variation and sensibility to the variation of Stream, pointing even more towards the necessity to individualize and adapt visual complexity to optimize task performance. Participants significantly performed worse when the amount of NPCs was modulated following an increase in physiological arousal, as shown in the adaptive control condition. Furthermore, we imposed capacity limits on visual WM memory when increasing the visual information load and the number of distracting objects. Finally, we add to results on social crowdedness [30, 75] and the distracting and perceptual demands that an increasing amount of human, tough virtual, place on their observers [68, 69, 76, 114].

6.1.2 Impact of Visual Complexity on Physiological Arousal. In the adaptive control condition, once our physiologically-adaptive function detected an increase in arousal, we increased the visual complexity, adding visual distractors to the surroundings. We found a main effect of Stream in the stable condition, showing how visual complexity can impact users' tonic component of arousal. However, we did not find an effect on the phasic component as this component might have encountered habituation effects over time, giving the continuous visual stimulation [95]. On the other side, we found an increased amplitude of the nsSCRs in the adaptive control condition. nsSCRs underlie the tonic stress produced during a sustained stimulation period, which resembles our experimental setting as participants were continuously exposed to variations of Stream of NPCs. These results can be explained given the bottom-up, involuntary orienting of attention that is driven by novelty detection and inevitably coupled with an automatic shift of attention [91, 111, 115] of attentional control caused by the distracting Stream of NPCs.

Furthermore, amplitude of nsSCRs has been associated with increased workload [11] in the N-back task [41, 90], in biofeedback [100] and in adaptive automation settings [13]. Therefore, increasing the visual complexity every 20sec might have progressively forced participants to reallocate attentional resources, related to the phasic component of electrodermal activity [52, 107]. This claim is also supported by the decrease in task accuracy.

Although SCL, slower than nsSCRs variations [50], failed to discriminate between the two adaptive algorithms, we can still confirm that the general physiological arousal was affected by Stream manipulations as shown in the Stable Stream conditions results. Therefore, we were able to model the user state sufficiently enough in a real-time environment to either decrease or improve performance by modulating distracting features of the environment based on variations of physiological arousal. The subjective ratings of perceived workload also support this claim. The condition of the highest visual complexity corresponded to the highest ratings in the NASA-TLX and the adaptive control condition. This is in line with previous work that showed either fluctuation of SCL [21, 34, 63] or phasic components [41] aligned with subjective measures of workload. Finally, we report a series of not significant differences over GEQ subscales. Even though this might be counterintuitive, we might argue that in our case the duration of each task (6 minutes) was not long enough to elicit a change in subjective engagement scores, as a high-demanding task that showed decreased engagement after steps of 20 minutes [40]. Future work should investigate the effect of time on task over subjective engagement and on adaptation to verify if our system can replicate previous work on this relationship.

6.1.3 Relationship Between Visual Complexity, Mental Workload, and Physiological Arousal. The results from NASA-TLX suggest that our manipulation of visual complexity through variations in the number of NPCs did not map onto systematic variation in perceived workload. To reiterate, we found significant differences between the highest (Stream = 347) and intermediate (Stream = 191) and lowest levels of visual complexity (Stream = 24). In contrast, no significant difference was found between the remaining intermediate levels. While there was no one-to-one mapping of perceived workload and number of NPCs, our findings thus still support the overall trend that increasing visual complexity leads to a higher perceived workload. Furthermore, the post hoc tests indicated that the relationship between visual complexity and perceived workload might not be strictly linear, which suggests that more sophisticated workload manipulation strategies could be explored in future research. The nonlinearity in the relationship between visual complexity and perceived workload could be attributed to factors such as individual differences in cognitive processing abilities [112] or attentional resource allocation [42, 81], which may interact with the level of visual complexity to influence workload perception in a complex and dynamic manner. It is possible that the variations in visual complexity did not affect everyone in the same way. Any variability in the overload point may be individually important for individuals. Therefore, user-dependent models [4, 31] are needed to evaluate visual complexity which further speaks towards user-adaptive environments.

When examining physiological arousal, we found that SCL does not evolve continuously with the number of NPCs, i.e., visual complexity. SCL follows a Bateman distribution and does not have a linear behavior [113]. The Bateman distribution is a biexponential function that describes the relationship between the intensity of a stimulus and the magnitude of a physiological response [7]. The Bateman distribution is characterized by a sigmoidal shape, with a steeper slope at low levels of arousal and a flatter slope at high levels of arousal. This means that small changes in the intensity of a stimulus can elicit a larger physiological response at low levels of arousal, while a larger shift in stimulus intensity is required to elicit the same response at high levels of arousal [113]. Therefore, increasing the number of NPCs may not necessarily lead to a proportional increase in SCL, as the response may saturate at higher arousal levels which must be considered when using physiological arousal as input for adaptive systems.

6.2 Physiological Computing VR Applications for Arousal Detection

Our physiologically adaptive system is independent of the simulation software package and thus can be applied to any VR scenario. The data are preprocessed over TCP/IP within the LSL framework and returned to the Unity environment. Stripling et al. [117] emphasizes the need for a "robust user-friendly closed-loop training environment" for producing automated methods for adaptive manipulations in complex simulations. Our work attempts to do so using peripheral physiological data. Our adaptive algorithm and paradigm can be applied to diverse mobile VR [17] and augmented reality (AR) settings.

First, considering the reported effect of visual complexity on nsSCRs, performance, and perceived workload, we can foresee how adapting visual complexity can improve Social VR scenarios or visualizations. This is not limited to VR and can be enlarged to blended mixed reality [10], where digital content complexity can be adapted to support users in productivity settings [63, 86]. With recent advancements in wireless HMDs, users can now immerse themselves in virtual reality beyond the limited screen size of a laptop or mobile phone, providing a larger virtual workspace for productive activities [71]. Our work contributes to improving these mobile VR and MR scenarios [72] by exploring the relationship between visual complexity, workload, and the contribution of adaptive systems.

Second, physiological measures of arousal are frequently used in therapy and rehabilitation applications for evaluation purposes. For example, when considering exposure therapy in a therapeutic VR environment, we could leverage stimuli that evoke anxiety responses, i.e., spiders and heights, as previously shown in literature [24, 92, 124]. While VR may provide a more controlled experience, mobile AR scenarios may offer the opportunity for exposure therapy and rehabilitation to occur in everyday situations, such as on a bus or train. Thus, adapting content that causes discomfort and arousal responses in everyday context might help understand and better predict at which stage of the therapeutic process the patient is situated and better drive the treatment.

Third, a related feeling of discomfort in VR environments is cybersickness. Cybersickness is defined as a set of adverse symptoms caused by the mismatches in visual and vestibular information both in VR and AR [116] and impacts EDA components [47, 51]. As visual sensory mismatches [84] and field of view [74] are VR components that might impact cybersickness, arousal-detection and adaptation might help in preventing this unpleasant effect and could be integrated into existing systems [82, 120, 123]. For example, cybersickness might occur when being engaged in a mobile VR scenario due to conflicting self-motion and visual stimulation with the natural movement (e.g. such as playing games or watching videos while travelling on a train or bus) [85, 129]. By detecting cybersickness in mobile applications, developers can adjust the content, reduce the severity of the stimuli, or adapt the motion to minimize the occurrence of cybersickness.

Fourth, when considering navigation and wayfinding in VR environments, adapting visual complexity based on physiological arousal might improve navigation by changing the environment to guide the user. The idea is that physiological arousal can be used as a proxy for cognitive load [21, 64]. When a user is experiencing a high cognitive load, for example, when they are lost or trying to navigate a complex environment, their physiological arousal levels may increase. Indeed, Lee et al. [70] investigated different visualizations to support navigation, and showed fading out virtual avatars is a viable alternative for users. Thus, when walking and searching for a specific direction, our system might contribute to fading out task-irrelevant bystanders and facilitating wayfinding dynamically.

Finally, our physiologically-adaptive systems might find applications in Social VR settings involving multi-agent interactions. NPCs display [44], proxemics [75, 108] and, quality of interactions with virtual agents are critical to the usability [118] in Social VR scenarios. Here again, physiological arousal as indexed by EDA features can quantify emotional discomfort [122], violation of proxemic space [55, 75], and the influence of virtual crowds [23, 93]. Therefore, arousal detection and adaptation might predict and prevent negative emotional valence associated with virtual agents, adapt the proxemic space of such agents, and adjust features of large virtual crowds [128].

6.3 Limitations and Future Work

While our study has provided valuable insights into the efficacy and applicability of the physiologically adaptive systems for visual complexity adaptation in VR, several limitations need to be addressed to enhance the validity and generalizability of our results and improve the architecture of future adaptive systems.

We discuss different approaches for EDA online preprocessing and how multimodal physiological input, e.g., ECG and EEG, can be implemented in adaptive systems to gain a more comprehensive understanding of the relationship between physiological responses and visual complexity. The generalizability of our results would also benefit from testing the feasibility of machine learning algorithms applied to EDA data and incorporating placebo or randomized conditions in the experimental design. Finally, we reflect on how ethical and privacy considerations should be implemented when designing physiologically adaptive systems.

6.3.1 Approaches for Online EDA Preprocessing and Machine Learning for Arousal Prediction. Even though we chose and implemented an established standardized method for EDA preprocessing, different preprocessing steps can be evaluated for adaptation purposes. For example, unusual steep rises might arise in ambulatory settings and impair adaptation. Here we used a high-pass filter that might compensate for artifacts and distort the entire EDA time series recording, considering high values. Therefore either adaptive filtering or thresholding might improve the quality of the physiological input to the adaptive system. For example, Wavelet transforms employ a Gaussian mixture distribution to model tonic and phasic components of EDA. Chen et al. [16] compared this approach to previous approaches and showed higher performance in artifact reduction.

Second, generalizability to the largest population is a crucial goal, and EDA suffers from significant interindividual variability. People with higher SCL frequently have more SCRs, and larger amplitude nsSCRs [12, 121]. Thus, using an a priori threshold for detecting nsSCRs might lead to either an inflated chance of false positives or negatives. Fixed thresholding might lead to more detected nsSCRs, and a higher nsSCRs frequency in populations with higher SCL as peaks with a low amplitude will be considered as nsSCRs. Again, this outcome is especially relevant for adaptation purposes, as if the adaptation impacts the arousal detection and VR adaptation, leading to malfunctioning adaptations. Hence, adaptive thresholding can improve sensitivity to detect phasic changes in EDA for users with high SCL variability. Kleckner et al. [62] showed the feasibility of such an approach, using a novel fixed plus adaptive threshold proportional to the SCL.

Third, another possible approach is to investigate the application of machine learning algorithms to adapt the visual complexity, predict physiological arousal increase, and counteract the detrimental effect on task performance. For example, naive Bayes and Support Vector Machines successfully classified high and low performers in a VR Stroop Task based on behavioral data [5]. Therefore testing EDA features classification jointly with behavioral measures might lead to improved algorithm accuracies.

6.3.2 Multimodal Adaptive Systems and Placebo Control Condition. First, integration with other physiological inputs, such as EEG or ECG, would allow us to account for and monitor different cognitive or emotional states that might be impacted by visual complexity or other adaptations. In this regard, it is worth mentioning how the EEG Alpha band is associated with an increase in arousal activation [19, 73, 110], flow experience [27, 67] and internal attention states [25, 78] which are of central interest for adaptive systems aiming at supporting task performance and engagement. Another example that involved NPCs is the work by Keynan et al. [60], which demonstrated how virtual human avatars, responsive to EEG signals, can guide neurofeedback training for stress resilience. Hybrid Brain-Computer Interfaces [45] and sensor fusion approaches should be further explored, opening wide applications and allowing for deeper and fine-grained detection of user states.

Second, a potential methodological improvement that could be implemented in future studies is comparing stable and adaptive streams. This could be achieved by including two additional conditions, a randomly changing STREAM as a "placebo" condition and no STREAM as a control condition. When studying the efficacy of an adaptive system, it is essential to compare it to a placebo condition to ensure that the observed effects are not simply due to placebo effects [65]. However, in our work, participants were unaware of which condition they were experiencing, minimizing any possible placebo effects. The control condition, on the other hand, would allow us to assess the impact of the stream itself on the outcome measures without the confounding influence of the adaptive STREAM.

6.3.3 Ethical and Privacy Considerations in Physiological Computing. Physiological computing systems, designed to tap into private psychophysiological events for dynamic MR interactions, raise ethical concerns regarding informed consent, manipulation of users' states, and privacy [37]. Users must be fully aware of how physiological data are collected, used, and shared, specifically when the data are employed for model training and validation. To prioritize ethical design, researchers should inform participants which physiological state the system is optimizing for and allow users to return to a neutral affective state if the final state is perceived as undesirable [53]. Note, however, that verbal descriptions of system functionality might come with the problem of placebo effects in evaluation [65].

Physiologically-adaptive systems are grounded on symmetrical interaction between users and systems. Here, privacy concerns on data usage and protection might arise. The individual should retain formal and legal ownership of psychophysiological data, and any third party should receive access to such information only with user approval [53]. This is specifically relevant, as psychophysiological data might underlie specific cognitive or affective states and could be used for secondary medical diagnostic purposes. To mitigate these concerns, a privacy-by-design approach should be used, embedding privacy considerations into every stage of the design and development process [18, 37]. This includes conducting privacy impact assessments, implementing privacy-enhancing technologies, and collecting privacy-preserving data in every implementation stage of physiologically adaptive systems.

7 CONCLUSION

In this paper, we showed the feasibility of designing a physiologically-adaptive system with the potential to optimize the visual complexity of a VR scene while supporting task performance. Dynamic visual complexity adjustments based on physiological arousal allowed modulation of distracting elements and supported task performance without needing to adapt main task features or explicit input. Our work contributes to the effect of visual complexity on attentional load, physiological arousal, behavioral performance and subjective workload. Our results have important implications for the design of physiologically-adaptive systems, particularly those aimed at mobile applications, such as digital content blending and wayfinding, and social VR. Additionally, we highlight the potential for hybrid or sensor fusion approaches to improve arousal detection and adaptive algorithms. We believe that the data collection and processing methods we implemented online and the collected multimodal dataset will be helpful for various physiologically-adaptive systems applications.

8 OPEN SCIENCE

We encourage readers to reproduce and extend our results and analysis methods. Therefore our experimental setup, collected datasets, and analysis scripts are available on the Open Science Framework [20] (https://osf.io/axvfy/).

ACKNOWLEDGMENTS

Francesco Chiossi and Sven Mayer were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 251654672-TRR 161.

REFERENCES

[1] Daniel Afergan, Evan M. Peck, Erin T. Solovey, Andrew Jenkins, Samuel W. Hincks, Eli T. Brown, Remco Chang, and Robert J.K. Jacob. 2014. Dynamic Difficulty Using Brain Metrics of Workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3797–3806. https://doi.org/10.1145/2556288.2557230

[2] Richard J Allen, Alan D Baddeley, and Graham J Hitch. 2017. Executive and perceptual distraction in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* 43, 9 (2017), 1677. https://doi.org/10.1037/xhp0000413

[3] George A Alvarez and Patrick Cavanagh. 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science* 15, 2 (2004), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

[4] Lena M Andreessen, Peter Gerjets, Detmar Meurers, and Thorsten O Zander. 2021. Toward neuroadaptive support technologies for improving digital reading: a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. *User Modeling and User-Adapted Interaction* 31 (2021), 75–104. https://doi.org/10.1007/s11257-020-09273-5

[5] Justin Asbee, Kimberly Kelly, Timothy McMahan, and Thomas D Parsons. 2023. Machine learning classification analysis for an adaptive virtual reality Stroop task. *Virtual Reality* (2023), 1–17. https://doi.org/10.1007/s10055-022-00744-1

[6] Ebrahim Babaei, Benjamin Tag, Tilman Dingler, and Eduardo Velloso. 2021. A Critique of Electrodermal Activity Practices at CHI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 177, 14 pages. https://doi.org/10.1145/3411764.3445370

[7] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190, 1 (2010), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

[8] Mathias Benedek and Christian Kaernbach. 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47, 4 (2010), 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x

[9] Yulong Bian, Chenglei Yang, Chao Zhou, Juan Liu, Wei Gai, Xiangxu Meng, Feng Tian, and Chia Shen. 2018. Exploring the Weak Association between Flow Experience and Performance in Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173975

[10] Felix Born and Maic Masuch. 2017. Increasing presence in a mixed reality application by integrating a real time tracked full body representation. In *International Conference on Advances in Computer Entertainment*. Springer, 46–60. https://doi.org/10.1007/978-3-319-76270-8_4

[11] Wolfram Boucsein and Richard W Backs. 2000. Engineering psychophysiology as a discipline: Historical and theoretical aspects. *Engineering psychophysiology: Issues and applications* (2000), 3–30.

[12] Wolfram Boucsein, Ruediger Baltissen, and Manfred Euler. 1984. Dependence of skin conductance reactions and skin resistance reactions upon previous level. *Psychophysiology* 21, 2 (1984), 212–218. https://doi.org/10.1111/j.1469-8986.1984.tb00207.x

[13] Wolfram Boucsein, Andrea Haarmann, and Florian Schaefer. 2007. Combining skin conductance and heart rate variability for adaptive automation during simulated IFR flight. In *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 639–647. https://doi.org/10.1007/978-3-540-73331-7_70

[14] Peter Carruthers. 2013. Evolution of working memory. *Proceedings of the National Academy of Sciences* 110, supplement_2 (2013), 10371–10378. https://doi.org/10.1073/pnas.1301195110

[15] Alan Chalmers and Andrej Ferko. 2008. Levels of realism: From virtual reality to real virtuality. In *Proceedings of the 24th Spring Conference on Computer Graphics*. 19–25. https://doi.org/10.1145/1921264.1921272

[16] Weixuan Chen, Natasha Jaques, Sara Taylor, Akane Sano, Szymon Fedor, and Rosalind W Picard. 2015. Wavelet-based motion artifact removal for electrodermal activity. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, New York, NY, USA, 6223–6226. https://doi.org/10.1109/EMBC.2015.7319814

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

[18] Francesco Chiossi and Sven Mayer. 2023. How Can Mixed Reality Benefit From Physiologically-Adaptive Systems? Challenges and Opportunities for Human Factors Applications. *Proceedings of Workshop on the Future of Computational Approaches for Understanding and Adapting User Interfaces* (2023). https://doi.org/10.48550/arXiv.2303.17978

[19] Francesco Chiossi, Changkun Ou, and Sven Mayer. 2023. Exploring Physiological Correlates of Visual Complexity Adaptation: Insights from EDA, ECG, and EEG Data for Adaptation Evaluation in VR Adaptive Systems. In *ACM Conference on Human-Computer Interaction (CHI '23 Extended Abstracts)*. Association for Computing Machinery, Hamburg, Germany. https://doi.org/10.1145/3544549.3585624

[20] Francesco Chiossi, Yagiz Turgut, Robin Welsch, and Sven Mayer. 2023. Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality. https://doi.org/10.17605/OSF.IO/AXVFY

[21] Francesco Chiossi, Robin Welsch, Steeven Villa, Lewis Chuang, and Sven Mayer. 2022. Virtual Reality Adaptation Using Electrodermal Activity to Support the User Experience. *Big Data and Cognitive Computing* 6, 2 (2022), 55. https://doi.org/10.3390/bdcc6020055

[22] Francesco Chiossi, Johannes Zagermann, Jakob Karolus, Nils Rodrigues, Priscilla Balestrucci, Daniel Weiskopf, Benedikt Ehinger, Tiare Feuchtner, Harald Reiterer, Lewis L. Chuang, Marc Ernst, Andreas Bulling, Sven Mayer, and Albrecht Schmidt. 2022. Adapting visualizations and interfaces to the user. *it - Information Technology* (2022). https://doi.org/10.1515/itit-2022-0035

[23] Chris Christou, Kyriakos Herakleous, Aimilia Tzanavari, and Charalambos Poullis. 2015. Psychophysiological responses to virtual crowds: implications for wearable computing. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, New York, NY, USA, 35–41. https://doi.org/10.1109/ACII.2015.7344548

[24] Taylor W Cleworth, Brian C Horslen, and Mark G Carpenter. 2012. Influence of real and virtual heights on standing balance. *Gait & posture* 36, 2 (2012), 172–176. https://doi.org/10.1016/j.gaitpost.2012.02.010

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

[26] William Jay Conover. 1999. *Practical nonparametric statistics*. Vol. 350. John Wiley & Sons. https://doi.org/10.2307/2981807

[27] Mihaly Csikszentmihalyi, Sami Abuhamdeh, and Jeanne Nakamura. 2005. Flow. *Handbook of competence and motivation* (2005), 598–608.

[28] Ralph B d'Agostino. 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58, 2 (1971), 341–348. https://doi.org/10.2307/2334522

[29] Jan W de Fockert, Geraint Rees, Christopher D Frith, and Nilli Lavie. 2001. The role of working memory in visual selective attention. *Science* 291, 5509 (2001), 1803–1806. https://doi.org/10.1126/science.1056496

[30] Janet A Desor. 1972. Toward a psychological theory of crowding. *Journal of personality and social psychology* 21, 1 (1972), 79. https://doi.org/10.1037/h0032112

[31] Tilman Deuschel. 2018. On the Influence of Human Factors in Adaptive User Interface Design. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 187–190. https://doi.org/10.1145/3213586.3213587

[32] Arindam Dey, Alex Chatburn, and Mark Billinghurst. 2019. Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE, New York, NY, USA, 220–226. https://doi.org/10.1109/VR.2019.8797840

[33] John Duncan and Glyn W Humphreys. 1989. Visual search and stimulus similarity. *Psychological review* 96, 3 (1989), 433. https://doi.org/10.1037/0033-295X.96.3.433

[34] Mai ElKomy, Yomna Abdelrahman, Markus Funk, Tilman Dingler, Albrecht Schmidt, and Slim Abdennadher. 2017. ABBAS: an adaptive bio-sensors based assistive system. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 2543–2550. https://doi.org/10.1145/3027063.3053179

[35] Hing Yee Eng, Diyu Chen, and Yuhong Jiang. 2005. Visual working memory for simple and complex visual stimuli. *Psychonomic bulletin & review* 12, 6 (2005), 1127–1133. https://doi.org/10.3758/BF03206454

[36] Kate C Ewing, Stephen H Fairclough, and Kiel Gilleade. 2016. Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop. *Frontiers in human neuroscience* 10 (2016), 223. https://doi.org/10.3389/fnhum.2016.00223

[37] Stephen H Fairclough. 2009. Fundamentals of physiological computing. *Interacting with computers* 21, 1-2 (2009), 133–145. https://doi.org/10.1016/j.intcom.2008.10.011

[38] Stephen H Fairclough. 2017. Physiological computing and intelligent adaptation. In *Emotions and affect in human factors and human-computer interaction*. Elsevier, 539–556. https://doi.org/10.1016/B978-0-12-801851-4.00020-3

[39] Stephen H Fairclough, Christopher Burns, and Ute Kreplin. 2018. FNIRS activity in the prefrontal cortex and motivational intensity: impact of working memory load, financial reward, and correlation-based signal improvement. *Neurophotonics* 5, 3 (2018), 035001. https://doi.org/10.1117/1.NPh.5.3.035001

[40] Stephen H Fairclough and Louise Venables. 2006. Prediction of subjective states from psychophysiology: A multivariate approach. *Biological psychology* 71, 1 (2006), 100–110. https://doi.org/10.1016/j.biopsycho.2005.03.007

[41] Ying Xing Feng, Tong Boon Tang, and Eric Tatt Wei Ho. 2021. Phasic Electrodermal Activity Indicates Changes in Workload and Affective States. In *2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*. IEEE, New York, NY, USA, 133–137. https://doi.org/10.1109/ICICyTA53712.2021.9689112

[42] Daniel Fitousi and Michael J Wenger. 2011. Processing capacity under perceptual and cognitive load: a closer look at load theory. *Journal of experimental psychology: human perception and performance* 37, 3 (2011), 781. https://doi.org/10.1037/a0020675

[43] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, Wolfram Boucsein, Don C Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T Roth, Michael E Dawson, and Diane L Filion. 2012. Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 8 (2012), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

[44] Guo Freeman and Divine Maloney. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27. https://doi.org/10.1145/3432938

[45] Doron Friedman. 2015. Brain-computer interfacing and virtual reality. *Handbook of digital games and entertainment technologies* (2015), 1–22. https://doi.org/10.1007/978-981-4560-52-8_2-1

[46] Keisuke Fukuda and Edward K Vogel. 2011. Individual differences in recovery time from attentional capture. *Psychological science* 22, 3 (2011), 361–368. https://doi.org/10.1177/0956797611398493

[47] Alireza Mazloumi Gavgani, Keith V Nesbitt, Karen L Blackmore, and Eugene Nalivaiko. 2017. Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience* 203 (2017), 41–50. https://doi.org/10.1016/j.autneu.2016.12.004

[48] Guilherme Gonçalves, Hugo Coelho, Pedro Monteiro, Miguel Melo, and Maximino Bessa. 2022. Systematic Review of Comparative Studies of the Impact of Realism in Immersive Virtual Experiences. *ACM Comput. Surv.* (apr 2022). https://doi.org/10.1145/3533377

[49] Klaus Gramann, Ryan McKendrick, Carryl Baldwin, Raphaëlle N Roy, Camille Jeunet, Ranjana K Mehta, and Giovanni Vecchiato. 2021. Grand field challenges for cognitive neuroergonomics in the coming decade. , 643969 pages. https://doi.org/10.3389/fnrgo.2021.643969

[50] Alberto Greco, Gaetano Valenza, and Enzo Pasquale Scilingo. 2016. *Modeling for the Analysis of the EDA*. Springer International Publishing, Cham, 19–33. https://doi.org/10.1007/978-3-319-46705-4_2

[51] Jože Guna, Gregor Geršak, Iztok Humar, Maja Krebl, Marko Orel, Huimin Lu, and Matevž Pogačnik. 2020. Virtual reality sickness and challenges behind different technology and content settings. *Mobile Networks and Applications* 25, 4 (2020), 1436–1445. https://doi.org/10.1007/s11036-019-01373-w

[52] Hugo Hammond, Graham Thomas, and Iain D Gilchrist. 2022. Dynamically changing attention in complex visual stimuli. *Journal of Vision* 22, 14 (2022), 3602–3602. https://doi.org/10.1167/jov.22.14.3602

[53] PA Hancock and JL Szalma. 2003. The future of neuroergonomics. *Theoretical Issues in Ergonomics Science* 4, 1-2 (2003), 238–249. https://doi.org/10.1080/1463922021000020927

[54] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

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

[56] Yizhen Huang, Eric Richter, Thilo Kleickmann, Axel Wiepke, and Dirk Richter. 2021. Classroom complexity affects student teachers' behavior in a VR classroom. *Computers & Education* 163 (2021), 104100. https://doi.org/10.1016/j.compedu.2020.104100

[57] Wijnand A IJsselsteijn, Yvonne AW de Kort, and Karolien Poels. 2013. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* 46, 1 (2013).

[58] Susanne M Jaeggi, Martin Buschkuehl, Walter J Perrig, and Beat Meier. 2010. The concurrent validity of the N-back task as a working memory measure. *Memory* 18, 4 (2010), 394–412. https://doi.org/10.1080/09658211003702171

[59] Manolya Kavakli, Iwan Kartiko, John Porte, and Nolwenn Bigoin. 2008. Effects of digital content on motion sickness in immersive virtual environments. In *3rd International Conference on Computer Science & Information Systems, July*. 23–24.

[60] Jackob N Keynan, Avihay Cohen, Gilan Jackont, Nili Green, Noam Goldway, Alexander Davidov, Yehudit Meir-Hasson, Gal Raz, Nathan Intrator, Eyal Fruchter, et al. 2019. Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nature human behaviour* 3, 1 (2019), 63–73. https://doi.org/10.1038/s41562-018-0484-3

[61] Madison Klarkowski, Daniel Johnson, Peta Wyeth, Cody Phillips, and Simon Smith. 2018. Don't Sweat the Small Stuff: The Effect of Challenge-Skill Manipulation on Electrodermal Activity. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 231–242. https://doi.org/10.1145/3242671.3242714

[62] Ian Kleckner, Jolie B Wormwood, Rebecca M Jones, Erika Siegel, Eva Culakova, James Heathers, Lisa Feldman Barrett, Catherine Lord, Karen Quigley, and Matthew Goodwin. 2021. Adaptive thresholding increases ability to detect changes in rate of skin conductance responses to psychologically arousing stimuli. (2021). https://doi.org/10.31234/osf.io/b4agz

[63] Thomas Kosch, Jakob Karolus, Havy Ha, and Albrecht Schmidt. 2019. Your skin resists: exploring electrodermal activity as workload indicator during manual assembly. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. 1–5. https://doi.org/10.1145/3319499.3328230

[64] Thomas Kosch, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. *ACM Comput. Surv.* (jan 2023). https://doi.org/10.1145/3582272

[65] Thomas Kosch, Robin Welsch, Lewis Chuang, and Albrecht Schmidt. 2022. The Placebo Effect of Artificial Intelligence in Human-Computer Interaction. *ACM Trans. Comput.-Hum. Interact.* (mar 2022). https://doi.org/10.1145/3529225

[66] Arthur Kramer and John Spinks. 1991. Capacity views of human information processing. (1991).

[67] Daniel Kramer. 2007. Predictions of performance by EEG and skin conductance. *Indiana undergraduate journal of cognitive science* 2 (2007), 3–13.

[68] Nilli Lavie. 1995. Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance* 21, 3 (1995), 451. https://doi.org/10.1037/0096-1523.21.3.451

[69] Nilli Lavie, Tony Ro, and Charlotte Russell. 2003. The role of perceptual load in processing distractor faces. *Psychological science* 14, 5 (2003), 510–515. https://doi.org/10.1111/1467-9280.03453

[70] Jaewook Lee, Fanjie Jin, Younsoo Kim, and David Lindlbauer. 2022. User Preference for Navigation Instructions in Mixed Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 802–811. https://doi.org/10.1109/VR51125.2022.00102

[71] Jingyi Li, Ceenu George, Andrea Ngao, Kai Holländer, Stefan Mayer, and Andreas Butz. 2021. Rear-seat productivity in virtual reality: Investigating vr interaction in the confined space of a car. *Multimodal Technologies and Interaction* 5, 4 (2021), 15. https://doi.org/10.3390/mti5040015

[72] Jingyi Li, Luca Woik, and Andreas Butz. 2022. Designing Mobile MR Workspaces: Effects of Reality Degree and Spatial Configuration During Passenger Productivity in HMDs. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 181 (sep 2022), 17 pages. https://doi.org/10.1145/3546716

[73] CL Lim, RJ Barry, E Gordon, A Sawant, C Rennie, and C Yiannikas. 1996. The relationship between quantified EEG and skin conductance level. *International journal of psychophysiology* 21, 2-3 (1996), 151–162. https://doi.org/10.1016/0167-8760(95)00049-6

[74] JJ-W Lin, Henry Been-Lirn Duh, Donald E Parker, Habib Abi-Rached, and Thomas A Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings ieee virtual reality 2002*. IEEE, New York, NY, USA, 164–171. https://doi.org/10.1109/VR.2002.996519

[75] Joan Llobera, Bernhard Spanlang, Giulio Ruffini, and Mel Slater. 2010. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP)* 8, 1 (2010), 1–12. https://doi.org/10.1145/1857893.1857896

[76] Elizabeth S Lorenc, Remington Mallett, and Jarrod A Lewis-Peacock. 2021. Distraction in visual working memory: Resistance is not futile. *Trends in cognitive sciences* 25, 3 (2021), 228–239. https://doi.org/10.1016/j.tics.2020.12.004

[77] Svein Magnussen, Mark W Greenlee, Rolf Asplund, and Stein Dyrnes. 1991. Stimulus-specific mechanisms of visual short-term memory. *Vision research* 31, 7-8 (1991), 1213–1219. https://doi.org/10.1016/0042-6989(91)90046-8

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

[79] Dominique Makowski, Tam Pham, Zen J Lau, Jan C Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods*

53, 4 (2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[80] Guido Makransky, Thomas S Terkildsen, and Richard E Mayer. 2019. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction* 60 (2019), 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

[81] Hadas Marciano and Yaffa Yeshurun. 2017. Large inter-individual and intra-individual variability in the effect of perceptual load. *PLoS One* 12, 4 (2017), e0175060. https://doi.org/10.1371/journal.pone.0175060

[82] Nicolas Martin, Nicolas Mathieu, Nico Pallamin, Martin Ragot, and Jean-Marc Diverrez. 2020. Virtual reality sickness detection: An approach based on physiological signals and machine learning. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, New York, NY, USA, 387–399. https://doi.org/10.1109/ISMAR50242.2020.00065

[83] Matteo Marucci, Gianluca Di Flumeri, Gianluca Borghini, Nicolina Sciaraffa, Michele Scandola, Enea Francesco Pavone, Fabio Babiloni, Viviana Betti, and Pietro Aricò. 2021. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports* 11, 1 (2021), 1–15. https://doi.org/10.1038/s41598-021-84196-8

[84] Alireza Mazloumi Gavgani, Deborah M Hodgson, and Eugene Nalivaiko. 2017. Effects of visual flow direction on signs and symptoms of cybersickness. *PloS one* 12, 8 (2017), e0182790. https://doi.org/10.1371/journal.pone.0182790

[85] Michael E McCauley and Thomas J Sharkey. 1992. Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators & Virtual Environments* 1, 3 (1992), 311–318. https://doi.org/10.1162/pres.1992.1.3.311

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

[87] DJ McKeefry, MP Burton, and C Vakrou. 2007. Speed selectivity in visual short term memory for motion. *Vision research* 47, 18 (2007), 2418–2425. https://doi.org/10.1016/j.visres.2007.05.011

[88] Kathryn M McMillan, Angela R Laird, Suzanne T Witt, and M Elizabeth Meyerand. 2007. Self-paced working memory: Validation of verbal variations of the n-back paradigm. *Brain research* 1139 (2007), 133–142. https://doi.org/10.1016/j.brainres.2006.12.058

[89] Fiona McNab and Raymond J Dolan. 2014. Dissociating distractor-filtering at encoding and during maintenance. *Journal of Experimental Psychology: Human Perception and Performance* 40, 3 (2014), 960. https://doi.org/10.1037/a0036013

[90] Bruce Mehler, Bryan Reimer, Joseph F Coughlin, and Jeffery A Dusek. 2009. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation research record* 2138, 1 (2009), 6–12. https://doi.org/10.3141/2138-02

[91] Abdul Momin and Sudip Sanyal. 2019. Analysis of Electrodermal Activity Signal Collected During Visual Attention Oriented Tasks. *IEEE Access* 7 (2019), 88186–88195. https://doi.org/10.1109/ACCESS.2019.2925933

[92] Andreas Mühlberger, Martin J Herrmann, Georg Wiedemann, Heiner Ellgring, and Paul Pauli. 2001. Repeated exposure of flight phobics to flights in virtual reality. *Behaviour research and therapy* 39, 9 (2001), 1033–1050. https://doi.org/10.1016/S0005-7967(00)00076-0

[93] Michael Nelson and Christos Mousas. 2020. A Virtual Reality Framework for Human-Virtual Crowd Interaction Studies. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, New York, NY, USA, 209–213. https://doi.org/10.1109/AIVR50618.2020.00043

[94] Mansoor Niaz and Robert H Logie. 1993. Working memory, mental capacity and science education: towards an understanding of the 'working memory overload hypothesis'. *Oxford Review of Education* 19, 4 (1993), 511–525. https://doi.org/10.1080/0305498930190407

[95] Reiner Nikula. 1991. Psychological correlates of nonspecific skin conductance responses. *Psychophysiology* 28, 1 (1991), 86–90. https://doi.org/10.1111/j.1469-8986.1991.tb03392.x

[96] Domen Novak, Matjaž Mihelj, and Marko Munih. 2012. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with computers* 24, 3 (2012), 154–172. https://doi.org/10.1016/j.intcom.2012.04.003

[97] Klaus Oberauer, Stephan Lewandowsky, Edward Awh, Gordon DA Brown, Andrew Conway, Nelson Cowan, Christopher Donkin, Simon Farrell, Graham J Hitch, Mark J Hurlstone, et al. 2018. Benchmarks for models of short-term and working memory. *Psychological bulletin* 144, 9 (2018), 885. https://doi.org/10.1037/bul0000153

[98] Aude Olivia, Michael L Mack, Mochan Shrestha, and Angela Peeper. 2004. Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 26.

[99] Bettina Olk, Alina Dinu, David J Zielinski, and Regis Kopper. 2018. Measuring visual search and distraction in immersive virtual reality. *Royal Society open science* 5, 5 (2018), 172331. https://doi.org/10.1098/rsos.172331

[100] Avinash Parnandi and Ricardo Gutierrez-Osuna. 2015. A comparative study of game mechanics and control laws for an adaptive physiological game. *Journal on Multimodal User Interfaces* 9, 1 (2015), 31–42. https://doi.org/10.1007/s12193-014-0159-y

[101] Thomas D Parsons and James L Reinebold. 2012. Adaptive virtual environments for neuropsychological assessment in serious games. *IEEE Transactions on Consumer Electronics* 58, 2 (2012), 197–204. https://doi.org/10.1109/TCE.2012.6227413

[102] Andrea C Pierno, Andrea Caria, Scott Glover, and Umberto Castiello. 2005. Effects of increasing visual load on aurally and visually guided target acquisition in a virtual environment. *Applied ergonomics* 36, 3 (2005), 335–343. https://doi.org/10.1016/j.apergo.2004.11.002

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

[104] Eric D Ragan, Doug A Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P McMahan. 2015. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics* 21, 7 (2015), 794–807. https://doi.org/10.1109/VR.2002.996519

[105] Eric Redlinger, Bernhard Glas, and Yang Rong. 2021. Enhanced Cognitive Training using Virtual Reality: Examining a Memory Task Modified for Use in Virtual Environments. In *2021 5th International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 1–8. https://doi.org/10.1145/3480433.3480435

[106] M. Richter, G.H.E. Gendolla, and R.A. Wright. 2016. Three Decades of Research on Motivational Intensity Theory: What We Have Learned About Effort and What We Still Don't Know. Advances in Motivation Science, Vol. 3. Elsevier, 149–186. https://doi.org/10.1016/bs.adms.2016.02.001

[107] Luana L Righi, Gilberto F Xavier, Marcus VC Baldo, and Hamilton Haddad. 2019. Responses to auditory distractors during a class depend on participant's enjoyment: An eye-tracking and skin conductance study. *Psychology & Neuroscience* 12, 4 (2019), 473. https://doi.org/10.1037/pne0000178

[108] Radiah Rivu, Yumeng Zhou, Robin Welsch, Ville Mäkelä, and Florian Alt. 2021. When friends become strangers: Understanding the influence of avatar gender on interpersonal distance in virtual reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 234–250. https://doi.org/10.1007/978-3-030-85607-6_16

[109] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of vision* 7, 2 (2007), 17–17. https://doi.org/10.1167/7.2.17

[110] Jacqueline A Rushby, Robert J Barry, Adam R Clarke, and Mohammad VaezMousavi. 2007. Arousal and activation effects on physiological and behavioral responding during a continuous performance task. (2007).

[111] Tsugunosuke Sakai, Harunya Tamaki, Yosuke Ota, Ryohei Egusa, Shigenori Imagaki, Fusako Kusunoki, Masanori Sugimoto, and Hiroshi Mizoguchi. 2017. Eda-based estimation of visual attention by observation of eye blink frequency. *International Journal on Smart Sensing and Intelligent Systems* 10, 2 (2017), 1–12. https://doi.org/10.21307/ijssis-2017-212

[112] Rajwant Sandhu and Benjamin James Dyson. 2016. Cross-modal perceptual load: the impact of modality and individual differences. *Experimental Brain Research* 234 (2016), 1279–1291. https://doi.org/10.1007/s00221-015-4517-0

[113] RL Schneider. 1987. A mathematical-model of human-skin conductance. In *Psychophysiology*, Vol. 24. SOC PSYCHOPHYSIOL RES 1010 VERMONT AVE NW SUITE 1100, WASHINGTON, DC 20005, 610–610.

[114] Alexander Skulmowski, Steve Nebel, Martin Remmele, and Günter Daniel Rey. 2021. Is a preference for realism really naive after all? A cognitive model of learning with realistic visualizations. *Educational Psychology Review* (2021), 1–27. https://doi.org/10.1007/s10648-021-09638-1

[115] Evgeny N Sokolov, John A Spinks, Risto Näätänen, and Heikki Lyytinen. 2002. *The orienting response in information processing*. Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410601490

[116] Jan-Philipp Stauffert, Florian Niebling, and Marc Erich Latoschik. 2020. Latency and cybersickness: impact, causes, and measures. A review. *Frontiers in Virtual Reality* 1 (2020), 582204. https://doi.org/10.3389/frvir.2020.582204

[117] Roy Stripling, Joseph T. Coyne, Anna Cole, Daniel Afergan, Raymond L. Barnes, Kelly A. Rossi, Leah M. Reeves, and Dylan D. Schmorrow. 2007. Automated SAF Adaptation Tool (ASAT). In *Foundations of Augmented Cognition*, Dylan D. Schmorrow and Leah M. Reeves (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 346–353. https://doi.org/10.1007/978-3-540-73216-7_39

[118] Philipp Sykownik and Maic Masuch. 2020. The experience of social touch in multi-user virtual reality. In *26th ACM symposium on virtual reality software and technology*. 1–11. https://doi.org/10.1145/3385956.3418944

[119] Margaret Toms, Neil Morris, and Peter Foley. 1994. Characteristics of visual interference with visuospatial working memory. *British Journal of Psychology* 85, 1 (1994), 131–144. https://doi.org/10.1111/j.2044-8295.1994.tb02513.x

[120] Samaikya Valluripally, Vaibhav Akashe, Michael Fisher, David Falana, Khaza Anuarul Hoque, and Prasad Calyam. 2021. Rule-based Adaptations to Control Cybersickness in Social Virtual Reality Learning Environments. In *2021 8th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, New York, NY, USA, 350–358.

https://doi.org/10.1109/FiCloud49777.2021.00057

[121] Peter H Venables and Margaret J Christie. 1980. Electrodermal activity. *Techniques in psychophysiology* 54, 3 (1980).

[122] Matias Volante, Sabarish V Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B Daily, and Tracy Fasolino. 2016. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1326–1335. https://doi.org/10.1109/TVCG.2016.2518158

[123] Jialin Wang, Hai-Ning Liang, Diego Monteiro, Wenge Xu, and Jimin Xiao. 2022. Real-time Prediction of Simulator Sickness in Virtual Reality Games. *IEEE Transactions on Games* (2022). https://doi.org/10.1109/TG.2022.3178539

[124] Frank H Wilhelm, Monique C Pfaltz, James J Gross, Iris B Mauss, Sun I Kim, and Brenda K Wiederhold. 2005. Mechanisms of virtual reality exposure therapy: The role of the behavioral activation and behavioral inhibition systems. *Applied psychophysiology and biofeedback* 30, 3 (2005), 271–284. https://doi.org/10.1007/s10484-005-6383-1

[125] Rex A Wright. 2008. Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. *Social and Personality Psychology Compass* 2, 2 (2008), 682–701. https://doi.org/10.1111/j.1751-9004.2008.00093.x

[126] Dongrui Wu, Christopher G Courtney, Brent J Lance, Shrikanth S Narayanan, Michael E Dawson, Kelvin S Oie, and Thomas D Parsons. 2010. Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Transactions on Affective Computing* 1, 2 (2010), 109–118. https://doi.org/10.1109/T-AFFC.2010.12

[127] Jong H Yoon, Clayton E Curtis, and Mark D'Esposito. 2006. Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. *Neuroimage* 29, 4 (2006), 1117–1126. https://doi.org/10.1016/j.neuroimage.2005.08.024

[128] Xuguang Zhang, Xiuxin Yang, Weiguang Zhang, Gongfa Li, and Hui Yu. 2021. Crowd emotion evaluation based on fuzzy inference of arousal and valence. *Neurocomputing* 445 (2021), 194–205. https://doi.org/10.1016/j.neucom.2021.02.047

[129] Magdalena Żukowska, Paweł Buń, Filip Górski, and Beata Starzyńska. 2019. Cyber sickness in industrial virtual reality training. In *Advances in Manufacturing II: Volume 1-Solutions for Industry 4.0*. Springer, 137–149. https://doi.org/10.1007/978-3-030-18715-6_12

View publication stats