Behavioral, Peripheral, and Central Neural Correlates of Augmented Reality Guidance of Manual Tasks

Alejandro L. Callara ⓘ, *Member, IEEE*, Gianluca Rho ⓘ, Sara Condino ⓘ, Vincenzo Ferrari ⓘ, *Member, IEEE*, Enzo Pasquale Scilingo ⓘ, *Senior Member, IEEE*, and Alberto Greco ⓘ, *Senior Member, IEEE*

Abstract—Objective: The use of commercially available optical-see-through (OST) head-mounted displays (HMDs) in their own peripersonal space leads the user to experience two perception conflicts that deteriorate their performance in precision manual tasks: the vergence-accommodation conflict (VAC) and the focus rivalry. In this work, we aim characterizing for the first time the psychophysiological response associated with user's incorrect focus cues during the execution of an augmented reality (AR)-guided manual task with the Microsoft HoloLens OST-HMD. Methods: 21 subjects underwent to a "connecting-the-dots" experiment with and without the use of AR, and in both binocular and monocular conditions. For each condition, we quantified the changes in autonomic nervous system (ANS) activity of subjects by analyzing the electrodermal activity (EDA) and heart rate variability. Moreover, we analyzed the neural central correlates by means of power measures of brain activity and multivariate autoregressive measures of brain connectivity extracted from the electroencephalogram (EEG). Results: No statistically significant differences of ANS correlates were observed among tasks, although all EDA-related features varied between rest and task conditions. Conversely, significant differences among conditions were present in terms of EEG-power variations in the μ (8–13) Hz and β (13–30) Hz bands. In addition, significant changes in the causal interactions of a brain network involved in motor movement and eye-hand coordination comprising the precentral gyrus, the precuneus, and the fusiform gyrus were observed. Conclusion: The physiological plausibility of our results suggest promising future applicability to investigate more complex scenarios, such as AR-guided surgery.

I. INTRODUCTION

AUGMENTED reality (AR) devices implementing optical-see-through (OST) technology allow overlaying of computer-generated imagery on the real-world egocentric view of the user. The real view and virtual content are merged together by rendering the latter on a two-dimensional (2-D) microdisplay. Specifically collimating lenses, placed between the microdisplay and the optical combiner, focus the 2-D virtual image so that it appears at a predefined and comfortable viewing distance on a virtual image plane (i.e., the focal plane of the display) [1]. The implementation of such technology on wearable devices can provide the user with a hand-free setup, which is particularly useful for performing manual tasks of different kinds [2].

However, this 2-D–three-dimensional (3-D) fusion can lead to a perceptual conflict between the 2-D virtual content on the surface of projection and 3-D real-world [2]. More specifically, both vergence-accommodation conflict (VAC) and focus rivalry (FR) phenomena can occur. VAC is a visual phenomenon that arises when the brain receives conflicting cues during binocular vision. Particularly, in the OST paradigm, the virtual image is focused at a fixed depth away from the eyes whereas real-world objects are not, resulting in conflicting information within the vergence-accommodation feedback loop [3]. This difference in the focal distances of virtual and real objects leads also to FR, whose effect is to constrain the subject to selectively focus on a single cue (i.e., the virtual or the real content) [4]. Indeed, human visual system does not allow to focus on more than one focal plane at a time. As a result, when a target is at a distance from the fixation point of the eyes, but outside the optical extension of the human depth-of-field, it will be perceived as blurred by the viewer. All these issues are even more emphasized when using commercial OST head-mounted displays (HMDs) to guide manual tasks [2]. Indeed, the focal distance of such devices is typically around 2–3 m therefore applications in which the real content is located within the user's peripersonal space can lead to VAC- and FR-related discomfort due to the distance gap between the 2-D virtual image and the real object.

A recent study has quantified the effects of VAC and FR in a simple "connecting-the-dots" task performed in an AR
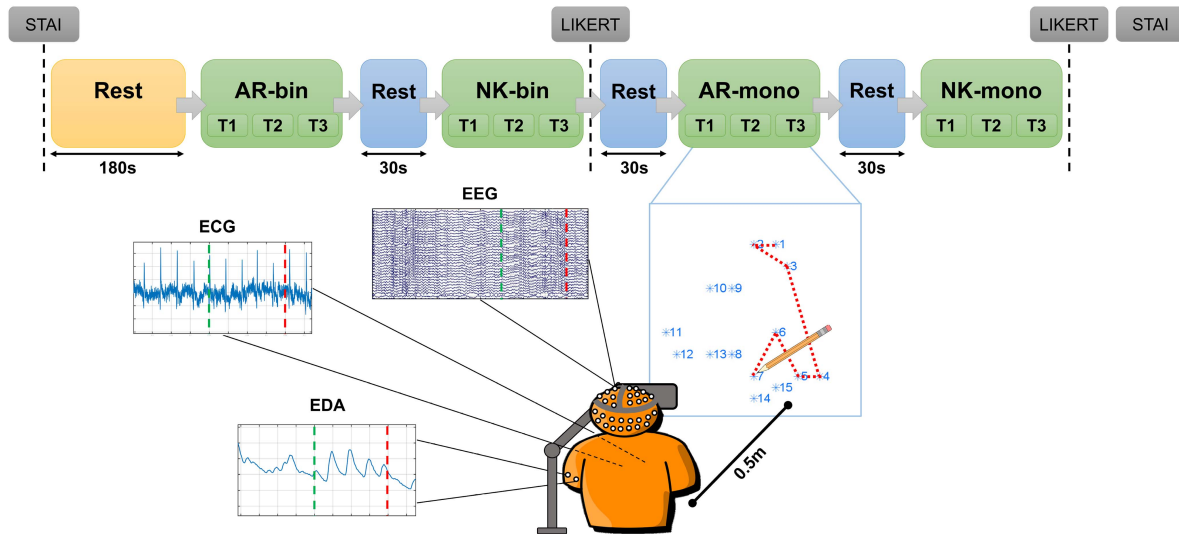
Fig. 1. Experimental setup and exemplary of experimental timeline. Subjects were equipped with EEG, ECG, and EDA sensors, and with a 1st Generation Microsoft HoloLens OST-HMD AR device. During each experimental condition subjects performed three trials of a "connecting-the-dots" task followed by 30 s of rest. Markers of pencil press (green dashed lines) and release (red dashed lines) were collected with an *ad hoc* touch capacitive sensor and time-locked to the recorded signals. Conditions were pseudorandomized across subjects. At the end of each EYE-modality block subjects filled-in a LIKERT-type questionnaire. At the beginning and at the end of the experiments subjects filled in the STAI questionnaire. Subjects performed the experiment by drawing lines on a paper placed on a vertical support at a distance of 0.5 m. The AR device was ensured on the head of the subjects through a metal frame support.

environment [2]. Interestingly, they observed that both VAC and FR during the binocular vision, as well as FR during the monocular vision, deteriorated task performance. VAC and FR effects are usually assessed in the scientific literature by using questionnaires [2], [5], [6]. However, collecting information through self-reporting has several limitations and can be either consciously or unconsciously biased. In this context, more objective measures of visual discomfort and fatigue can be obtained from physiological signals, such as electro-oculogram (e.g., by detecting blink rates [7]), eye-tracking (e.g., by analyzing eye movements [8], [9]), and electroencephalography (EEG) (e.g., by exploiting changes in EEG power [10]). Indeed, physiological signals are less prone to subjective bias compared to self-assessed reports. Our preliminary study has investigated mental workload during an analogous "connecting-the-dots" task by analyzing the EEG frontal-alpha-asymmetry index [11]. Preliminary results suggested that performing the task using AR can be more demanding than performing it with natural vision. However, a fully detailed characterization of the physiological changes induced by AR is still missing. Indeed, to properly characterize a user's psychophysiological state, distinct measures of different bodily responses need to be considered, ranging from autonomic nervous system (ANS) correlates to brain activity and connectivity measures.

To this aim, here, we provide for the first time a multimodal characterization of the psychophysiological response during a "connecting-the-dots" task performed with and without AR, and in both binocular and monocular vision. Specifically, we monitor changes during the AR task in subjects' autonomic response, by focusing on electrodermal activity (EDA) and heart rate variability (HRV) signals, and we analyze the neural central correlates of AR use by considering several measures derived from the EEG. EDA and HRV are two of the most commonly used ANS correlates to infer a subject's psychophysiological state under several controlled conditions such as stress, fear and many other [12], [13], [14]. Moreover, in this study, we analyze EEG to quantify changes in power at the scalp level and, for the first time, to characterize the brain sources participating to the task completion and the changes in the connectivity among such sources based on the different experimental conditions. Connectivity is evaluated by means of multivariate autoregressive (MVAR) models applied to independent component (IC) timecourses [15], [16], [17]. We also introduce another novelty by analyzing physiological responses in an event-related fashion. Indeed, the "connecting-the-dots" task is by its nature an event-related experiment made of several *press* and *release* events. Hence, all methods are applied to study the physiological response immediately after press events and contextual to the line drawing. The efficacy of the stimulation protocol is assessed by analyzing behavioral measures and subjective ratings of users.

II. MATERIALS AND METHODS

A. Subjects

The study was conducted according with the guidelines of the Declaration of Helsinki, and approved by the Bioethics Committee of the University of Pisa Review No. 14/2019, May 3rd, 2019. All subjects gave their informed consent to take part to the study.

Twenty-one healthy volunteers (age 27.5 ± 4.5, four females, all right-handed) were enrolled to participate in the study. All subjects had normal or corrected-to-normal visual acuity and limited previous experience with AR devices.

B. Experimental Protocol

The experimental setup and an example of the experimental timeline are reported in Fig. 1. After a resting period of 180 s, subjects performed a "connecting-the-dots" precision task under four different experimental conditions as follows.

1) Binocular AR-guided task (hereinafter AR-bin).
2) Monocular AR-guided task (hereinafter AR-mono).
3) Binocular naked-eye task (hereinafter NK-bin).
4) Monocular naked-eye task (hereinafter NK-mono).

These were purposely designed to study the influence of both AR and eye (hereinafter called EYE) modalities on the subject's performance. Each condition consisted of three different trials of equal complexity. This latter was controlled by designing paths of equivalent total length and an equal number of segments for each trial. Conditions were interleaved with 30 s of rest.

During tasks performed with AR, random dots were projected on paper using an HMD 1st Generation Microsoft HoloLens device (see Section II-C for details), whereas, for naked-eye tasks, dots were simply printed on paper. The paper was placed on vertical support at a common distance of ~ 0.5 m from the eyes of the subjects inducing visual discomfort and fatigue due to the mismatch with the focal distance of the glasses (i.e., ~ 2 m). This was purposely done to simulate those situations in which the AR glasses are used under suboptimal conditions, for instance in AR-aided surgery tasks. During the experiment, subjects were asked to sit on a comfortable chair and connected the dots by using a pencil. Moreover, subjects were asked to detach the pencil from the paper at the end of each drawn segment. The AR device was mounted on a metal frame support (and not on the subject's head) to limit the potential movements of the glasses during the experiment (see Fig. 2).

To identify the time points at which the subject started drawing each segment (i.e., pressed with the pencil on the paper), we designed and built an in-house touch capacitive sensor. The sensor was interfaced with the EEG triggering system through an Arduino MKR WIFI 1010 microcontroller and an *ad hoc* analog frontend circuit. Accordingly, we were able to determine the time points at which subjects pressed and released the pencil. These two events were tagged as *press* and *release*.

For each subject, conditions were pseudorandomized first according to the EYE modality, and then according to the AR modality. Hence, combinations such as (AR-bin, NK-mono, NK-bin, AR-mono) or (NK-bin, AR-mono, AR-bin, NK-mono) were avoided. Intertrial intervals were given by the time required by the experimenter to change this article of each trial. Instead, inter-block intervals were fixed at a duration of 30 s. At the end of each EYE-modality, subjects were asked to fill in a LIKERT-type questionnaire aiming at measuring visual discomfort, fatigue and perception associated with the AR "connecting-the-dots" task [18]. Finally, before and after the experiment, subjects filled in the state-trait anxiety inventory (STAI) questionnaire [19].

C. AR Device

We used the 1st Generation Microsoft HoloLens OST-HMD as the AR device to display virtual dots during the experiment. Hololens features a self-contained computing power, based on
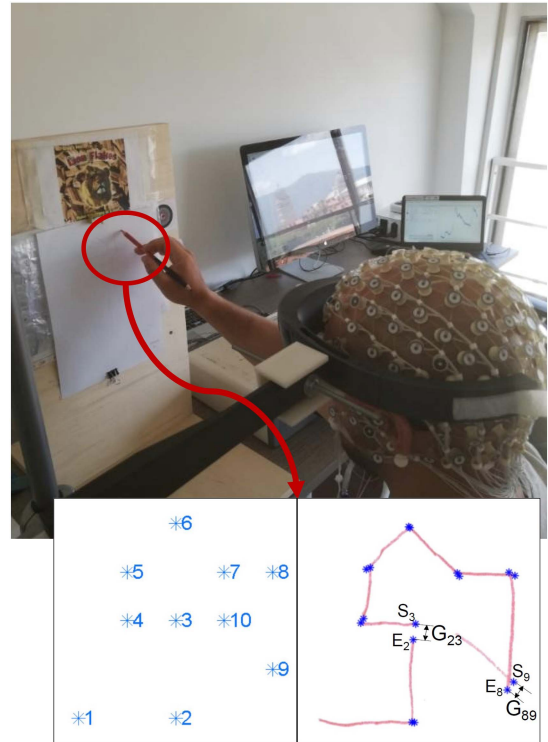
Fig. 2. Exemplary of the experimental setup for one subject and for the gap estimate. The AR device was mounted on a metal frame support limiting movements of the glasses during the experiment. A "connecting-the-dots" task was performed on a paper placed on a vertical support in front of the subject. During NK modality, the dots were printed on a paper, whereas during AR modality, the dots were shown to the subjects through the AR device. An example of the gap (G_ij) for (i = 2, j = 3) and (i = 8, j = 9) is reported in the bottom-right panel.

an undisclosed Intel 32-bit processor, 2 GB of RAM and 64 GB of flash memory, and a custom-built Microsoft Holographic Processing Unit (HPU 1.0) which supports Universal Windows Platform apps. Furthermore, Hololens allows for wireless communication featuring both Wi-Fi 802.11ac and Bluetooth 4.1 LE wireless technology. The sensory system of the device includes: one depth camera, four grayscale tracking cameras, one world-facing photo/video camera (2 MP), one ambient light sensor, one inertial measurement unit to track head movements, and four microphones.

D. Connecting-the-Dots Performance Analysis

For each experimental condition, we analyzed subjects' performance in completing the task by evaluating the gap (G_ij) between the ending and starting points of each pair *(i, j)* of consecutive lines (see Fig. 2). We purposely focused on this measure because of its independence from AR device virtual-to-real calibration errors [2]. Accordingly, we reduced the impact on task performance evaluation of possible distortions of the perceived visual content resulting in the misperception of line lengths. To this aim, we first detected line endpoints automatically by using the Harris corner detector [20] under MATLAB R2017b [21]. Then, for each trial, we estimated the maximum and mean error gaps (i.e., max error and mean error), respectively.

E. Physiological Data Acquisition

1) Peripheral Signals: We acquired EDA and ECG signals during the experiment. The EDA was recorded using a Shimmer 3 GSR+ unit at a sampling rate of 100 Hz. The recording was performed by placing two electrodes on the distal phalanx of the index and middle fingers of the subjects' nondominant hand. Instead, the ECG signal was acquired with the peripheral physiological measurement module of a 128-channel Geodesic EEG System 300 from Electrical Geodesic, Inc. (EGI). The recording was performed at a sampling rate of 500 Hz, with two electrodes positioned below the right and left clavicle, and a reference electrode that corresponded to the Cz channel of the EEG cap.

2) Electroencephalogram: We acquired EEG data using a 128-channel Geodesic EEG System 300 from EGI. We kept electrodes impedances always below 20 kohm for the entire duration of the experiments. Channels were referenced to the Cz electrode during the acquisition. We used a sampling rate of 500 Hz.

F. Physiological Data Analysis

1) ANS Correlates: We extracted several features from the EDA and electrocardiographic (ECG) signals to monitor subjects' changes in ANS activity.

The EDA signal reflects changes in the skin conductance induced by sweat glands' activity. Since sweat glands are under the only control of the sympathetic branch of the ANS, the analysis of EDA is a reliable way to infer SNS dynamics [14]. The EDA signal can be seen as the sum of two components that carry nonoverlapping information: a tonic component, which is a slow-varying component whose spectrum is below 0.05 Hz, and a phasic component, which reflects short-term stimulus-evoked responses. Here, we took advantage of the cvxEDA model to extract the tonic and phasic components of the EDA signal, as well as the sudomotor nerve activity (SMNA) generating phasic responses [22]. Specifically, we downsampled EDA signals to a sampling rate of 50 Hz and normalized them to have zero mean and unit variance before applying the cvxEDA model. Then, starting from the estimated components of tonic, phasic, and SMNA, we derived several features related to SNS activity, namely: the mean and the standard deviation of the tonic (i.e., **Mean Tonic** and **Std Tonic**), the mean of the phasic (i.e., **Mean Phasic**), the number of peaks (i.e., **N peaks**), the maximum peak (i.e., **Max Peak**), and the sum of peaks (i.e., **AmpSum**) of the SMNA, and the power spectrum in the (0.045–0.25) Hz (i.e., **EDASYMP**). More specifically, tonic-related features were estimated in the 35 s after the start of each trial, whereas phasic-related components were time-locked to the *press* events and estimated in the 5 s following the start of each segment drawn. Such a choice of different time ranges for feature extraction was purposely made to capture the different dynamics of tonic and phasic components.

The ECG signal was analyzed with Kubios HRV [23]. Briefly, the RR series were extracted from the ECG using the Pan–Tompkins algorithm [24]. Peak-detection artefacts were removed by applying a cubic spline interpolation method, and

the obtained RR series were resampled to 4 Hz to derive the HRV signal [25]. Starting from the HRV, we extracted several features in the time, frequency, and nonlinear domains, which were representative of ANS dynamics. Different from EDA, HRV features are influenced by both SNS and PNS activity. As for the tonic-related features, we estimated all HRV features in the 35 s after the start of each trial. These were: the mean and the standard deviation of the HRV (i.e., **Mean HRV** and **Std HRV**), the square root of the mean-squared differences of successive normal-to-normal (NN) intervals (i.e., **RMSSD**), the percentage number of pairs of adjacent NN intervals differing by more than 50 ms (i.e., **pNN50**), the power expressed as a percentage of total power in the low-frequency (i.e., **LF**, 0.04–0.15 Hz) and high frequency (i.e., **HF**, 0.15–0.40 Hz) ranges, the ratio of LF to HF power (i.e., **LF/HF ratio**), and the minor (i.e., **SD1**) and major (i.e., **SD2**) axis of the ellipse that best fits the Poincaré plot of RR intervals.

In addition, all features were estimated also during the initial resting condition. In this case, we simply segmented the signals into 35 and 5 s windows to match the estimates on tasks' related conditions.

2) EEG Data Analysis: EEG data were analyzed in terms of power and connectivity measures with EEGLAB and MATLAB custom scripts [26]. To this aim, we first preprocessed EEG signals with a pipeline comprising:

1) downsampling to 125 Hz (after applying a low-pass antialiasing filter);
2) high-pass filtering (filter cutoff 1 Hz);
3) bad channel removal based on correlation criterion [27];
4) ICA decomposition and selection of components related to brain activity [28]; and
5) estimation of the equivalent current dipole associated to each brain-related component [29], [30].

Cleaned signals were segmented into epochs centred around *press* events and ranging from –0.5 to 2 s. A subtractive baseline ranging from –0.5 to 0 s was removed from all epochs. Finally, for each experimental condition, we computed average EEG epochs.

EEG power was estimated for each condition on average EEG epochs. Specifically, we derived the power spectral density (PSD) of each channel in the 2 s following *press* events with the aim of characterizing the brain response during task execution. In particular, PSD estimates were obtained for all the experimental conditions in the $\delta(1$–$4)$, $\theta(4$–$7)$, $\mu(7$–$13)$, $\beta_1(13$–$20)$, $\beta_2(20$–$30)$, and $\gamma(30$–$45)$ Hz frequency bands.

EEG responses were also analyzed in terms of effective brain connectivity by describing the interactions among subject-specific ICA-related EEG sources with MVAR models [31]. Operationally, we estimated MVAR models with a sliding-window approach as implemented in the Source Information Flow Toolbox [32], [33]. Sliding windows were 500 ms long and the window step size was equal to 20 ms. Then, starting from model coefficients, we calculated the renormalized partial directed coherence (RPDC, [34]) in 30-log scaled frequency bins from 1 to 45 Hz. Accordingly, we obtained multivariate causal estimates among sources at specific frequencies of interest. As a result, for each subject, we obtained a spectrotemporal

connectivity matrix RPDC (i, j, t, f) with (i, j) representing networks targets and sources, and (t, f) representing the time-window and the frequency at which the interaction is happening.

Group-level connectivity was obtained by projecting subject-specific RPDC matrices onto a common space across all subjects. To this aim, the dipole locations of ICs were transformed into probabilistic dipole densities smoothed with a 3-D Gaussian kernel with full-width at half maximum at 20 mm [16]. Then, dipole density was segmented into 76 anatomical cortical regions of interest (ROIs) defined by the automated anatomical labeling (AAL) atlas [35]. Afterward, we weighted subject-specific RPDC with dipole densities to obtain a $76x76xtxf$ connectivity matrices. Finally, we thresholded subject-specific weighted connectivity matrices by truncating the Gaussian distribution at a value of three standard deviations and considered a group-level connection only if it was present in at least the 80% of participants [16].

G. Statistical Analysis

We performed different statistical analyses to investigate how AR influenced subjective, behavioral, peripheral, and central physiological measures. For all the statistical comparisons performed, we considered a significance level of $\alpha = 0.05$.

1) STAI Scores: We analyzed the differences in STAI metrics measured before and after the experiment with the Wilcoxon signrank test to control whether performing the experiment negatively affected the anxiety level of subjects.

2) LIKERT-Type Questionnaire: We analyzed differences in subjects' ratings of visual discomfort, fatigue, and perception as measured by the LIKERT-type questionnaire to evaluate whether the EYE modality changed any of these measures while using AR during the "connecting-the-dots" task. Operationally, we tested for group differences with a Wilcoxon sign-rank test on the score of each question followed by false discovery rate (FDR) correction for multiple hypothesis testing [36].

3) Task Performance: We studied how task performance varied among different experimental conditions by analyzing the maximum and mean errors on gaps (see Section II-D) with multiple paired t-tests. More specifically, we compared the max error and mean error between experimental conditions (i.e., AR-bin, AR-mono, NK-bin, NK-mono) but without considering cross-modality comparisons: i.e., AR-bin versus NK-mono and AR-mono versus NK-bin were not included in the analysis. Multiple hypothesis testing was controlled with FDR correction [36].

4) ANS Correlates: We performed an exploratory statistical analysis to assess at the group level how tasks differed in terms of ANS correlates. To this aim, we performed a 1×4 ANOVA to test the null-hypothesis (H0) of no differences among the four conditions (NK-bin, NK-mono, AR-bin, AR-mono) for each EDA and HRV feature. In addition, we analyzed differences in ANS features with respect to the initial resting state by performing multiple paired t-tests (one for each feature) between each experimental condition and the resting condition. Multiple hypothesis testing was controlled using the FDR correction [36].

TABLE I
LIKERT-TYPE QUESTIONNAIRE RESULTS. AFTER FDR-BKY CORRECTION NONE OF THE *P*-VALUES WERE SIGNIFICANT

Item	AR-bin Median (25°-75°)	AR-mono Median (25°-75°)	uncorrected p-value
I perceived VR dots as clear and sharp	3 (3 − 4)	4 (3 − 4)	0.62 (NS)
I perceived VR numbers as clear and sharp	4 (4 − 4)	4 (3 − 4)	0.78 (NS)
I was able to contemporaneously focus at the VR content and real objects	2 (2 − 4)	3 (2 − 4)	0.19 (NS)
I did not perceive any visual discomfort due to blur	2 (2 − 3.25)	4 (3 − 4)	0.01 (NS)
I felt comfortable using this AR guidance modality for the selected task	3 (2 − 4.25)	3 (2 − 4)	1 (NS)
I did not experience visual fatigue	3 (2 − 4)	3 (2 − 4)	1 (NS)
I can trust this AR modality to successfully guide manual tasks	3 (2 − 3.25)	3 (2 − 4)	0.80 (NS)
I am confident of the precision of manual tasks guided by this AR modality	3 (2 − 3)	3 (2 − 4)	0.61 (NS)

5) EEG Neural Correlates: Group differences were also analyzed in terms of EEG correlates of power and connectivity. We performed a 1×4 ANOVA to test the null-hypothesis (H0) of no differences among the PSD values during the four conditions (NK-bin, NK-mono, AR-bin, AR-mono). The statistical significance of the ANOVA was assessed with a Monte Carlo-permutation method (number of permutations = 10 000) followed by cluster correction [29], [37]. Post hoc analyses were carried out by means of paired t-tests followed by Bonferroni correction.

In addition, for each condition, we evaluated significant differences in connectivity by comparing the observed RPDC values after *press* events with their averages during the baseline, as well as by comparing their values between experimental conditions through multiple paired t-tests. The statistical significance of such comparisons was assessed with a Monte Carlo-permutation method (number of permutations = 10 000) followed by cluster correction [29], [37].

III. RESULTS

The analysis conducted on STAI scores measured before and after the experiment was not significant ($p > 0.1$), indicating that the experiment did not induce any change in subjects' anxiety levels. In addition, as reported in Table I, all the performed analyses on LIKERT scores were not significant, indicating that

TABLE II
STATISTICAL ANALYSIS OF ANS FEATURES BETWEEN EACH EXPERIMENTAL CONDITION AND REST

		AR-bin vs Rest	AR-mono vs Rest	NK-bin vs Rest	NK-mono vs Rest
EDA features	Mean Tonic	1.77 ± 1.05 (**S**)	2.08 ± 0.73 (**S**)	1.8 ± 1.04 (**S**)	1.74 ± 0.88 (**S**)
	Std Tonic	0.05 ± 0.06 (**S**)	0.04 ± 0.06 (**S**)	0.05 ± 0.06 (**S**)	0.05 ± 0.07 (**S**)
	Mean Phasic	0.06 ± 0.08 (**S**)	0.05 ± 0.06 (**S**)	0.06 ± 0.07 (**S**)	0.06 ± 0.05 (**S**)
	Peak Max	-2.13 ± 2.02 (**S**)	-2.14 ± 2.03 (**S**)	-2.13 ± 2.03 (**S**)	-2.12 ± 2.01 (**S**)
	N Peaks	1.46 ± 1.59 (**S**)	1.7 ± 1.2 (**S**)	1.36 ± 1.34 (**S**)	1.65 ± 1.69 (**S**)
	AmpSum	-4.09 ± 4.41 (**S**)	-4.09 ± 4.38 (**S**)	-4.08 ± 4.38 (**S**)	-4.05 ± 4.38 (**S**)
	EDASYMP	0.02 ± 0.04 (**S**)	0.02 ± 0.02 (**S**)	0.01 ± 0.03 (**S**)	0.01 ± 0.01 (**S**)
HRV features	Mean HRV	-5.7 ± 19.34 (NS)	-7 ± 24.2 (NS)	-3.97 ± 20.16 (NS)	-6.64 ± 24.42 (NS)
	Std HRV	-1.54 ± 3.46 (NS)	-0.5 ± 4.22 (NS)	1.83 ± 11.56 (NS)	-2.01 ± 2.68 (**S**)
	RMSSD	-0.01 ± 0.03 (NS)	-0.01 ± 0.03 (NS)	0.03 ± 0.12 (NS)	-0.01 ± 0.03 (NS)
	pNN50	-5.38 ± 22.25 (NS)	-7.73 ± 16.15 (**S**)	-3.09 ± 23.11 (NS)	-6.87 ± 17.54 (NS)
	LF	-8.07 ± 21.22 (NS)	-11.77 ± 23.03 (**S**)	-7.9 ± 26.87 (NS)	-13.6 ± 22.46 (**S**)
	HF	2.8 ± 20.67 (NS)	1.18 ± 26.85 (NS)	2.61 ± 26.36 (NS)	2.99 ± 27.14 (NS)
	LF/HF	4.99 ± 24.66 (NS)	-0.43 ± 5.64 (NS)	-0.35 ± 6.77 (NS)	-0.12 ± 8.06 (NS)
	SD1	-0.01 ± 0.02 (NS)	-0.01 ± 0.02 (NS)	0.02 ± 0.09 (NS)	-0.01 ± 0.02 (NS)
	SD2	-0.02 ± 0.05 (NS)	-0.02 ± 0.05 (NS)	0 ± 0.09 (NS)	-0.03 ± 0.04 (**S**)

For each comparison, mean ± std of the difference between each experimental condition and the resting condition are reported, along with statistical significance after FDR-BKY correction (S = significant, NS = nonsignificant).
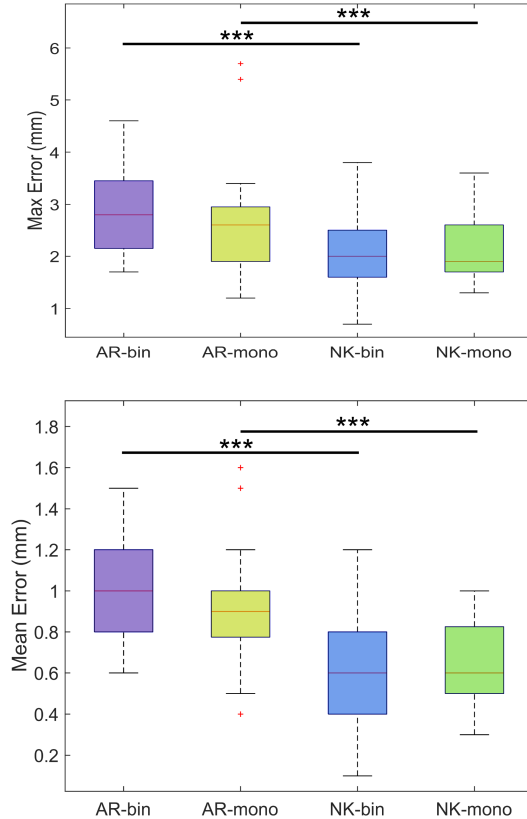

Fig. 3. Top: Max error. Bottom: Mean error. Significant differences are marked with asterisks: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$ (FDR corrected).

Fig. 4. 1×4 ANOVA results on PSD. For each frequency band, the scalp map of the significant p-values of the 1×4 ANOVA test are reported ($p < 0.05$ after cluster correction).

the EYE modality did not influenced visual discomfort, fatigue, and perception.

A. Connecting-the-Dots Task Performance

The results of gap error analysis is reported in Fig. 3. We observed significant differences in the max error between AR-bin and NK-bin conditions, with higher maximum errors for AR-bin. Analogously, significant differences were observed between AR-mono and NK-mono for this feature. Mean error analyses showed as well a significant difference in AR-bin versus
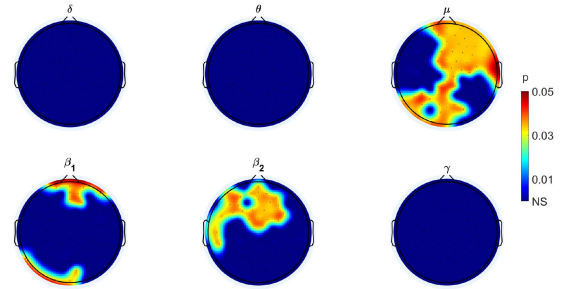
NK-bin and AR-mono versus NK-mono comparisons, with a higher mean error during AR modality. Overall, these results indicated that during AR, perceptual issues arising from FR and VAC deteriorated subjects' performance, confirming previous studies [2].

B. ANS Correlates

In Table II , we report the differences in ANS features between each experimental condition and the initial resting state along with their statistical significance. In particular, for each condition, we observed that all EDA-related features significantly differed from rest. Among these **Mean Tonic**, **Std Tonic**, **Mean Phasic**, **N peaks**, and **EDASYMP** were higher during task compared to rest. Conversely, **Max Peak** and **AmpSum** were lower during the task. Considering HRV features, none of them showed relevant patterns. However, the 1×4 ANOVA did not reveal any significant change in any of the considered features among experimental conditions, indicating that changes in ANS correlates induced by the "connecting-the-dots" task was not affected by AR nor by the EYE modalities. These results indicate that although some ANS features are capable of distinguishing between the task and the resting state, they did not allow to distinguish among conditions during the "connecting-the-dots" task.
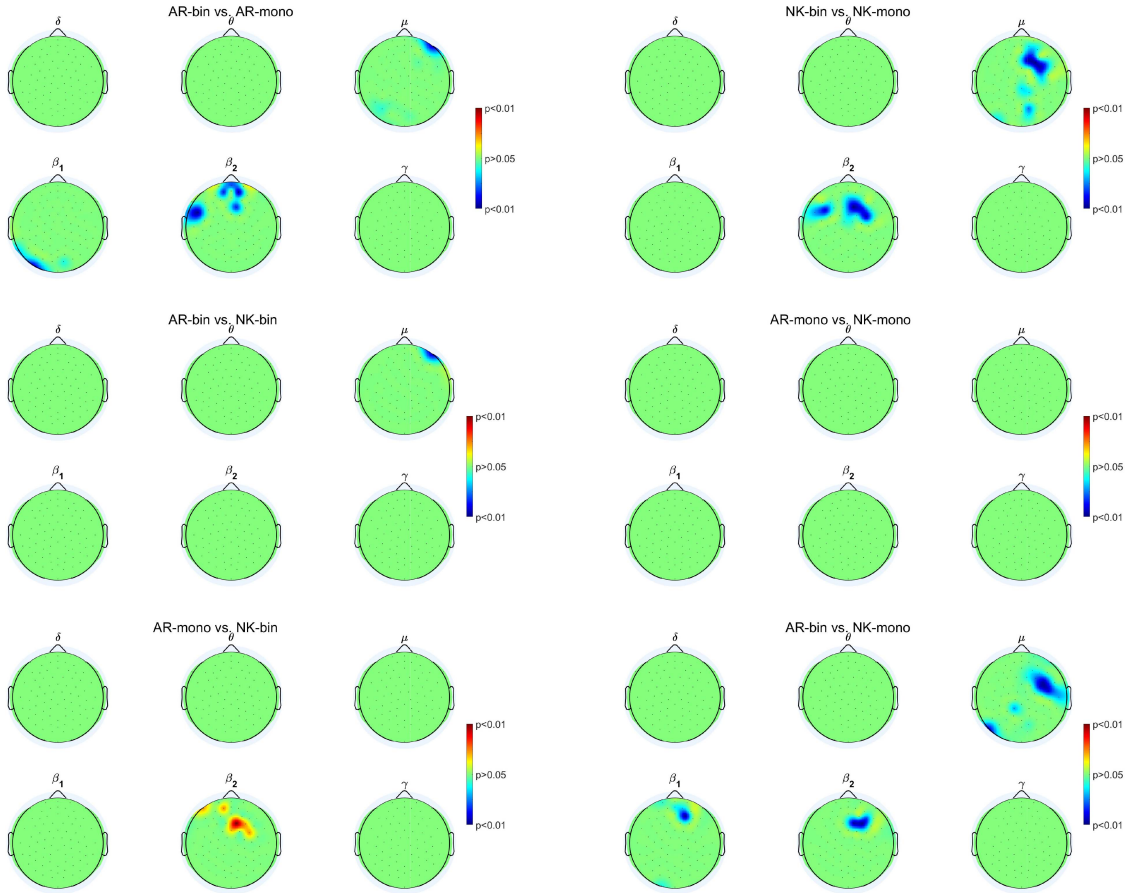
Fig. 5.   PSD post hoc analysis. For each frequency band, the scalp map of significant *p*-values of the post hoc t-tests are reported ($p < 0.05$). *p*-values in red indicate higher power during the condition on the right with respect to the condition on the left, whereas *p*-values reported in blue indicate lower power during the condition on the right with respect to the condition on the left.

C. EEG Power

In Fig. 4, we report the results of the 1×4 ANOVA on PSD estimates. Significant differences among conditions were observed at specific scalp sites and for specific frequency bands. The more widespread differences were at right frontal, middle-central, and left occipital regions in the μ band. In addition, there were significant differences in the frontocentral region in the β_2 band. Less widespread differences were present also in the β_1. Finally, we did not observe any significant change in PSD in the δ, θ, and γ bands.

Post hoc analyses are reported in Fig. 5. During binocular tasks, PSD was lower compared to monocular ones. This happened both for AR-bin versus AR-mono and for NK-bin versus NK-mono comparisons. However, differences were more widespread and lateralized when subjects were using AR. Specifically, this was mainly observed for frontal μ and posterior β_1 activity. Differences between conditions were less widespread when considering the AR modality contrast (i.e., AR-bin versus NK-bin and AR-mono versus NK-mono). In particular, few localized and lateralized differences in the AR-bin versus NK-bin comparison, with lower PSD values during AR-bin. Notably, we did not observe any significant difference between AR-mono and NK-mono conditions. Finally,

we observed some differences in the AR-mono versus NK-bin and AR-bin versus NK-mono contrasts, which were however of less interest.

D. EEG Connectivity

The results of the connectivity analysis are reported in Fig. 6. Among all the brain regions on which RPDC was projected only few of them had network edges that were significant. These nodes were the precentral gyrus (PCG), the precuneus (PCu), the fusiform gyrus (FG), and regions tagged as upper basal (UB) according to the AAL [16], [35].

We observed significant differences for the following comparisons: AR-bin versus AR-mono, NK-bin versus NK-mono, AR-bin versus NK-bin, AR-mono versus NK-mono. In the AR-bin versus AR-mono comparison there was only a significant decrease in connectivity from PCG $\rightarrow$ PCu during AR-bin. This same connection was observed to significantly differ in NK-bin versus NK-mono, AR-bin versus NK-bin and AR-mono versus NK-mono comparisons, showing higher values in the first condition of each comparison. However, also other connections significantly differed for these comparisons. Among these, a decrease in the reciprocal connections between UB and FG in the AR-bin versus NK-bin and AR-mono versus NK-mono during
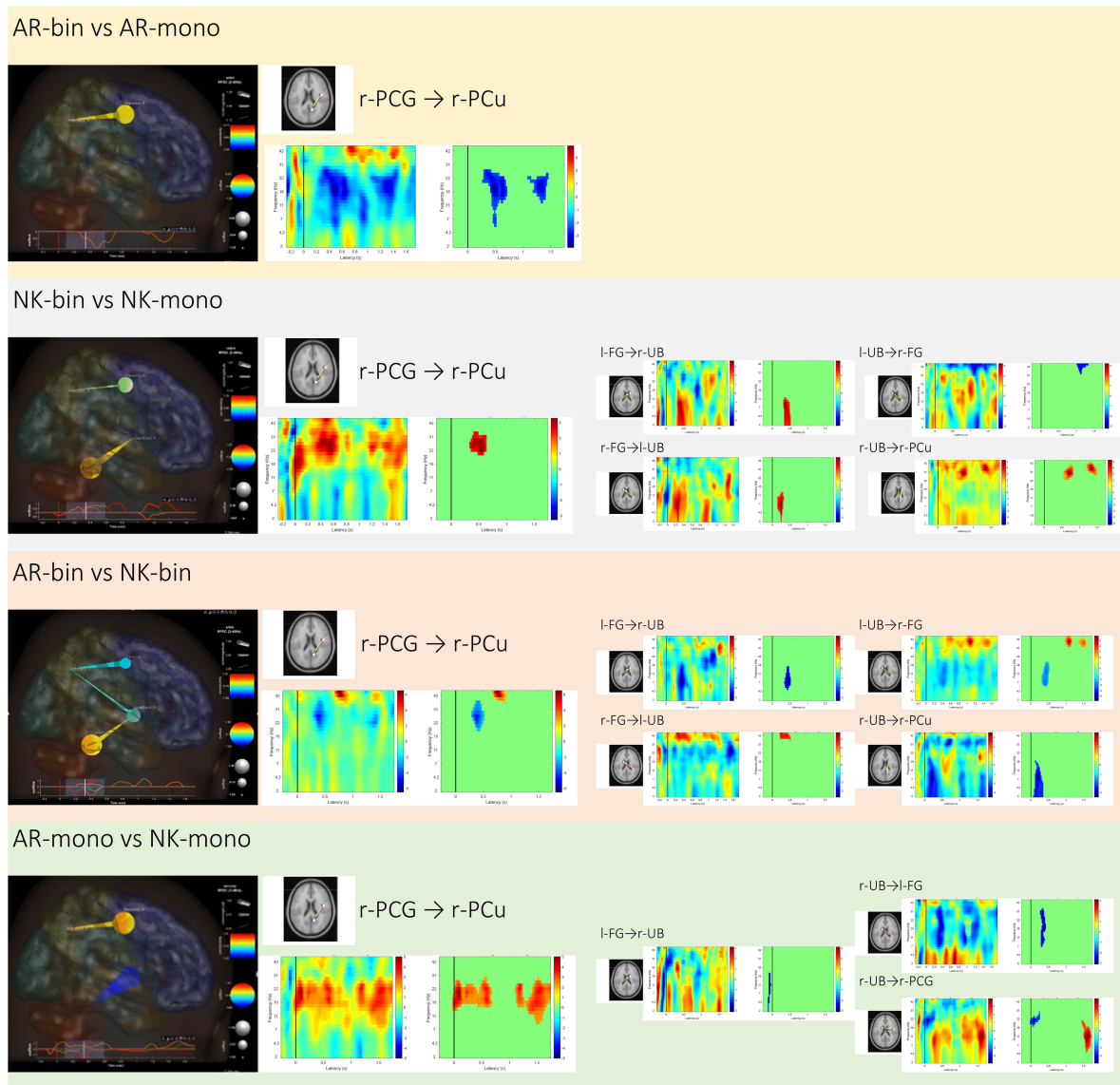
Fig. 6. Connectivity analysis. From top to bottom: AR-bin versus AR-mono, NK-bin versus NK-mono, AR-bin versus NK-bin, and AR-mono versus NK-mono. For each contrast, we report the difference in connectivity between the experimental condition on the left and the one on the right. On left of the image, a 3-D rendering of the differences in connectivity after ~300 ms from press events is reported. On the right, for each significant difference within the epoch, we report target-source *(i, j)* spectrotemporal matrices of RPDC differences between conditions, along with the time-frequency points in which they were statistically significant ($p < 0.05$ after cluster correction).

AR condition was observed. Finally, there were some connections which differed only in specific comparisons. Namely, the PCu → UB decreased during AR-mono compared to NK-mono. Conversely, this connection increased during NK-bin compared to NK-mono.

IV. DISCUSSION

In this work, we analyzed the behavioral and physiological correlates of AR during a "connecting-the-dots" task. We provide a first multimodal analysis of peripheral and central neural correlates of AR use during a precision motor coordination task. Our results integrate with previous reports-based exclusively on behavioral and error measures of task performance and suggest that using AR induces modifications in brain activity and connectivity compared to performing the same task without AR. Our results also highlight that EEG power and connectivity

measures seem to be more suited for characterizing the observed physiological responses compared to ANS correlates. Overall, our work provides useful insights for the study and design of applications of AR in more sophisticated and challenging tasks.

The analysis of subjective reports of visual fatigue, discomfort and perception, measured with the LIKERT-type questionnaire, did not show any significant difference between AR binocular and monocular modalities. This result corroborates the previous study reporting no differences in perceived discomfort between EYE modalities [2] and suggests that the observed differences in physiological measures in AR-bin versus AR-mono comparisons are not due to such factor.

The experimental study was designed to investigate the effects of the "FR" and VAC together (binocular tests) and separately (monocular tests), during an AR-guided manual task. More particularly, a simple task, which does not require a superimposition

of the virtual scenario with a physical counterpart, was designed for the study to limit the influence of (virtual-to-real) registration errors, which can negatively affect user performance. Our results show that subjects are more likely to be inaccurate during the AR-guided task. Accordingly, we can hypothesize that such inaccuracies may depend on the perceptual issues (focal rivalry and VAC in AR-bin, and focal rivalry in AR-mono) that impair the contemporary view of the real world and the virtual target. Indeed, the current generation of consumer-level OST HMDs is only capable of presenting the virtual content at a single fixed focal distance that is generally more than 2 m. Therefore, during manual tasks, the user eye cannot keep in focus both the virtual and real content at the same time. We believe that the use of OST HMDs with a focal distance greater than the length of the arm can be practical when the user can switch fixation points and alternately focus on real-world and VR information (e.g., textual information or simple icons), such as a driver alternately looking at the road and the tachometer in the cockpit. However, we believe that these perceptual issues may adversely affect user performance whenever the AR-guided manual task requires the user to simultaneously focus on virtual and real-world information (in our specific case, virtual dots, pen tip, and trajectory drawn, respectively). The results obtained suggest that although there is a growing interest in using commercial OST-HMDs to guide high-precision manual tasks (such as surgical ones), attention must be paid to the current limitations of the available technology, which is not designed for the peripersonal space.

The analysis performed on ANS correlates allowed distinguishing between each experimental condition and the initial resting state. Interestingly, all EDA features showed significant changes, suggesting that, regardless of the experimental condition, the motor task promoted an increase in sympathetic arousal. Conversely, variations of HRV features were less evident and not homogeneous across experimental conditions. However, none of the considered ANS features (both EDA and HRV) allowed for distinguishing among the experimental conditions. As in other studies, the complexity (e.g., hand–eye coordination) and voluntariness of the tasks induce acute sympathetic reactions that masked possible other differences in the task modality [38], causing a sort of saturation effect.

EEG measures showed a higher variation and specificity in the comparison among conditions compared to ANS correlates. Power and connectivity metrics showed changes based on both AR and EYE modalities that were physiologically plausible in terms of brain areas and frequency bands involved. In particular, the majority of changes in both types of metrics occurred in the μ and β bands: the two frequency bands primarily involved in motor execution [39]. In addition, the scalp maps of PSD analyses, as well as the dipole locations of network nodes involved in significant causalities were in regions known to be involved in motor coordination and execution [39], [40], [41].

For both AR-bin versus AR-mono and NK-bin versus NK-mono analyses, PSD was lower in the μ band for the right-centrofrontal regions for the AR-bin and NK-bin conditions, respectively. This behavior is in line with the idea that the more demanding a task is, the more evident the suppression of the μ band becomes. In this view, we can assume that the binocular conditions were more demanding than the monocular ones. Analogously, suppression in the β_1 and β_2, which is typically associated with voluntary movements, happened at centrofrontal regions. In such a case, however, we did not observe any particular lateralization. Finally, the more widespread differences in PSD scalp maps arose in the AR-bin versus NK-mono contrast, possibly indicating that these two conditions were the ones that differed more in terms of task demand.

Although the AR-bin versus NK-bin comparison showed less widespread differences compared to the above-mentioned ones, lower PSD values in the μ and β_1 bands were observed during AR-bin. Interestingly, the AR-mono versus NK-mono comparison did not show any significant change in PSD, possibly suggesting a limited influence of FR on PSD differences. Conversely, we speculate that the differences in AR-bin versus NK-bin could be due to the VAC, which arises only during binocular vision.

Similarly to PSD analyses, connectivity profiles highlighted a major involvement of μ and β bands, although some changes were observed also in γ and θ bands. Among the 76 ROIs considered in the analysis, only a few of them were involved in significant causal interactions. Notably, three out of four of these regions were the PCG, PCu, and the FG: a set of brain regions involved in different aspects of motor task execution. The PCG includes the primary motor cortex, participating in the control of voluntary motor movement [40]. In addition, among the many tasks carried out by the PCu, its corticocortical projections to the lateral parietal areas and premotor cortex seem to play a pivotal role in the visual guidance of hand movements, such as hand–eye coordination [42]. In this view, the observed changes in PCu → PCG and PCG → PCu causalities can be interpreted as an expression of the necessity of different eye–hand coordination between conditions. Finally, we observed the involvement of the FG, a region responsible for high-visual processing [43]. In this view, we speculate that the changes in the causalities of this region reflected the different needs of visual processing of each condition.

Of note, our results rely on a multimodal approach (i.e., the use multiple measures and signals) aiming at identifying the effect of AR FR and VACs on different behavioral and physiological measures, and on robust statistical testing to prove the validity of our hypothesis. Yet, a limitation of our work is given by the relatively small group of subjects participating in the experiments (n = 21). Future works may consider a higher number of participants to draw more robust conclusions. In this light, also solutions based on machine learning could help at finding specific features of interest characterizing FR and VAC during the use of AR.

V. CONCLUSION

This work investigated the psychophysiological response to VAC and FR perceptual conflicts arising from the use of commercial AR HMDs in manual tasks. Our study corroborates previous studies focusing on performance metrics during a "connecting the dots task" in both binocular and monocular vision, and extends them with new insights in terms of ANS and central neural responses elicited by the task. Our results suggest

physiological measures that could be used to test innovative HMD-ARs specifically designed for peripersonal space, with the goal of achieving correct perceptual augmentation minimizing visual discomfort. In fact, our physiological results complement those of the questionnaires highlighting differences in operating conditions of which the user may be unaware (indeed these are not captured by questionnaires), assuming a key role in the validation and testing of new devices.

Such encouraging and physiological-plausible results push toward future studies focusing on the risks and benefits of using AR technology in high workload demanding applications, as for instance AR-guided surgery.

References

[1] J. P. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence: Teleoperators Virtual Environments*, vol. 9, pp. 287–309, 2000.

[2] S. Condino, M. Carbone, R. Piazza, M. Ferrari, and V. Ferrari, "Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 411–419, Feb. 2020.

[3] G. Kramida, "Resolving the vergence-accommodation conflict in head-mounted displays," *IEEE Trans. Visual. Comput. Graph.*, vol. 22, no. 7, pp. 1912–1931, Jul. 2016.

[4] K. Oshima et al., "SharpView: Improved clarity of defocused content on optical see-through head-mounted displays," in *Proc. IEEE Symp. 3D User Interfaces*, 2016, pp. 173–181.

[5] E. Peli, "The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display," *Vis. Res.*, vol. 38, no. 13, pp. 2053–2066, 1998.

[6] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.*, vol. 8, no. 3, pp. 33–33, 2008.

[7] D. Kim, S. Choi, S. Park, and K. Sohn, "Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks," in *Proc. 17th Int. Conf. Digit. Signal Process.*, 2011, pp. 1–6.

[8] Y. Wang, G. Zhai, S. Chen, X. Min, Z. Gao, and X. Song, "Assessment of eye fatigue caused by head-mounted displays using eye-tracking," *Biomed. Eng. online*, vol. 18, no. 1, pp. 1–19, 2019.

[9] H.-J. Lin, L.-W. Chou, K.-M. Chang, J.-F. Wang, S.-H. Chen, and R. Hendradi, "Visual fatigue estimation by eye tracker with regression analysis," *J. Sensors*, vol. 2022, pp. 1–7, 2022.

[10] C. Chen et al., "Assessment visual fatigue of watching 3DTV using EEG power spectral parameters," *Displays*, vol. 35, no. 5, pp. 266–272, 2014.

[11] G. Rho et al., "A preliminary quantitative eeg study on augmented reality guidance of manual tasks," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, 2020, pp. 1–5.

[12] A. Greco et al., "Acute stress state classification based on electrodermal activity modeling," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 788–799, Jan.–Mar. 2023.

[13] L. Petrescu et al., "Machine learning methods for fear classification based on physiological features," *Sensors*, vol. 21, no. 13, 2021, Art. no. 4519.

[14] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, no. 2, Art. no. 479, 2020.

[15] A. L. Callara et al., "Ld-EEG effective brain connectivity in patients with Cheyne-Stokes respiration," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1216–1225, May 2020.

[16] S. K. Loo et al., "Neural activation and connectivity during cued eye blinks in Chronic Tic Disorders," *NeuroImage: Clin.*, vol. 24, 2019, Art. no. 101956.

[17] A. L. Callara, A. Greco, J. Frasnelli, G. Rho, N. Vanello, and E. P. Scilingo, "Cortical network and connectivity underlying hedonic olfactory perception," *J. Neural Eng.*, vol. 18, no. 5, 2021, Art. no. 056050.

[18] T. Nemoto and D. Beglar, "Likert-scale questionnaires," in *JALT 2013 Conf. Proc.*, pp. 1–8, 2014.

[19] C. D. Spielberger, F. Gonzalez-Reigosa, A. Martinez-Urrutia, L. F. Natalicio, and D. S. Natalicio, "The state-trait anxiety inventory," *Revista Interamericana de Psicologia/Interamerican J. Psychol.*, vol. 5, no. 3 & 4, pp. 145–158, 1971.

[20] C. Harris et al., "A combined corner and edge detector," in *Alvey Vis. Conf.*, 1988, pp. 10–5244.

[21] The MathWorks, Inc., MATLAB version: 7.10.0 (R2017b), 2017. [Online]. Available: https://www.mathworks.com

[22] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016.

[23] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen, "Kubios HRV–heart rate variability analysis software," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 210–220, 2014.

[24] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[25] M. Malik, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the European society of cardiology and the north american society for pacing and electrophysiology," *Ann. Noninvasive Electrocardiol.*, vol. 1, no. 2, pp. 151–181, 1996.

[26] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. methods*, vol. 134, no. 1, pp. 9–21, 2004.

[27] T. Mullen et al., "Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 2184–2187.

[28] S. Makeig, A. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," *Adv. Neural Inf. Process. Syst.*, vol. 8, pp. 145–151, 1995.

[29] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Comput. Intell. Neurosci.*, vol. 2011, 2011, Art. no. 156869.

[30] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent EEG sources are dipolar," *PLoS one*, vol. 7, no. 2, 2012, Art. no. e30135.

[31] A. Schlögl and G. Supp, "Analyzing event-related EEG data with multivariate autoregressive parameters," *Prog. Brain Res.*, vol. 159, pp. 135–147, 2006.

[32] T. Mullen, "Source information flow toolbox (SIFT)," *Swartz CenterComput Neurosci*, pp. 1–69, 2010.

[33] A. Delorme et al., "EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG processing," *Comput. Intell. Neurosci.*, vol. 2011, 2011, Art. no. 130714.

[34] B. Schelter, J. Timmer, and M. Eichler, "Assessing the strength of directed influences among neural signals using renormalized partial directed coherence," *J. Neurosci. methods*, vol. 179, no. 1, pp. 121–130, 2009.

[35] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *Neuroimage*, vol. 206, 2020, Art. no. 116189.

[36] Y. Benjamini, A. M. Krieger, and D. Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, vol. 93, no. 3, pp. 491–507, 2006.

[37] E. T. Bullmore, J. Suckling, S. Overmeyer, S. Rabe-Hesketh, E. Taylor, and M. J. Brammer, "Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, no. 1, pp. 32–42, Jan. 1999.

[38] M. Nardelli et al., "Recognizing AR-guided manual tasks through autonomic nervous system correlates: A preliminary study," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, 2020, pp. 1–6.

[39] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, 2000.

[40] L. Banker and P. Tadi, "Neuroanatomy, postcentral gyrus," in *StatPearls*. Treasure Island FL, USA: StatPearls Publishing, 2023.

[41] N. Wenderoth, F. Debaere, S. Sunaert, and S. P. Swinnen, "The role of anterior cingulate cortex and precuneus in the coordination of motor behaviour," *Eur. J. Neurosci.*, vol. 22, no. 1, pp. 235–246, 2005.

[42] A. E. Cavanna and M. R. Trimble, "The precuneus: A review of its functional anatomy and behavioural correlates," *Brain*, vol. 129, no. 3, pp. 564–583, 2006.

[43] K. S. Weiner and K. Zilles, "The anatomical and functional specialization of the fusiform gyrus," *Neuropsychologia*, vol. 83, pp. 48–62, 2016.