

ACCEPTED MANUSCRIPT

Neural Sources of Prediction Errors Detect Unrealistic VR Interactions

To cite this article before publication: Lukas Gehrke *et al* 2022 *J. Neural Eng.* in press <https://doi.org/10.1088/1741-2552/ac69bc>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2022 IOP Publishing Ltd.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere. As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Neural Sources of Prediction Errors Detect Unrealistic VR Interactions

Lukas Gehrke¹, Pedro Lopes² PhD, Marius Klug¹, Sezen Akman¹, Klaus Gramann¹ PhD

¹Biological Psychology and Neuroergonomics, Department of Psychology and Ergonomics, TU Berlin, Germany

²University of Chicago, USA

E-mail: lukas.gehrke@tu-berlin.de

Abstract.

Objective Neural interfaces hold significant promise to implicitly track user experience. Their application in VR/AR simulations is especially favorable as it allows user assessment without breaking the immersive experience. In VR, designing immersion is one key challenge. Subjective questionnaires are the established metrics to assess the effectiveness of immersive VR simulations. However, administering such questionnaires requires breaking the immersive experience they are supposed to assess.

Approach We present a complimentary metric based on a ERPs. For the metric to be robust, the neural signal employed must be reliable. Hence, it is beneficial to target the neural signal's cortical origin directly, efficiently separating signal from noise. To test this new complementary metric, we designed a reach-to-tap paradigm in VR to probe EEG and movement adaptation to visuo-haptic glitches. Our working hypothesis was, that these glitches, or violations of the predicted action outcome, may indicate a disrupted user experience.

Main Results Using prediction error negativity features, we classified VR glitches with 77% accuracy. We localized the EEG sources driving the classification and found midline cingulate EEG sources and a distributed network of parieto-occipital EEG sources to enable the classification success.

Significance Prediction error signatures from these sources reflect violations of user's predictions during interaction with AR/VR, promising a robust and targeted marker for adaptive user interfaces.

Keywords: EEG, Virtual Reality, BCI, Neural Interface Technology, Post-error Slowing, Prediction Error, Predictive Coding Submitted to: *J. Neural Eng.*

1. Introduction

One of the key challenges in virtual reality (VR) is to create a user experience that mimics the natural, real-world, experience as closely as possible. The overarching goal

Neural Sources Detecting Unrealistic VR Interactions

in designing immersive experiences is that users “treat what they perceive as real” and, as a consequence, feel present in the virtual world [1]. This requires a high degree of visual and haptic synchronization, for example for successful tele-operated surgeries, VR simulations and experiences. To achieve such haptic realism, many (consumer) devices are available on the market today.

Yet, to assess the effectiveness of VR simulations, the most established metrics rely on the users subjective interpretation of unspecific, yet standardized, questions [2, 3]. Unfortunately, answering these immersion questionnaires requires to break the users immersion to collect data about the previous interaction [4].

One way to overcome this limitation is to elicit a body illusion in VR using avatars. The idea is, that when the body illusion is effective, users will identify strongly with their avatar which in turn evidences an effective VR simulation [5]. Using the famed rubber-hand illusion, the proprioceptive displacement estimates of the real arm position towards the avatar’s arm position are a robust feature of effective VR simulations [6]. However, for accurate diagnostics of a given simulation and the immersion channels and hardware in use, a *continuous* labeling of the user experience is desirable. Psychometric tests enrich the labels accuracy. But, with increasing frequency of such tests, users may experience severe distraction. For example, in order to quantify the contributions of vibrotactile feedback to the user’s experience when grabbing a virtual hammer, immediate observation is required.

We and others have previously proposed the use of the frontal ‘prediction error’ negativity (PEN) as a feature for fast, real-time, detection of VR system errors which may, in turn, cause a loss in the sense of physical immersion [7, 8, 9]. Based on the idea that the brain has evolved to optimize motor behavior by detecting sensory mismatches, these studies promoted the usage of PENs to label a perceived loss in physical immersion, potentially impacting presence experience thereby providing *continuous* diagnostics about user experience.

To feel present in an environment, users need to establish a dynamic and precise interaction with their surroundings. This allows users to infer the causal structures in the (virtual) world and develop strategies to deal with uncertainties in their dynamic environment [10]. Today, the brain is frequently conceived of as a model of its environment, in the constant game of predicting the causes of its available sensory data [11, 12, 13]. In this predictive coding conception, probabilistic analyses of previous experiences drive inferences about which actions and perceptual events are causally related. This is inherently tied to the bodys capacity to act on the environment, rendering the action-perception cycle of cognition into an embodied process [14]. When all movement-related sensory data (i.e., sensorimotor data) are consistent with the predicted outcome of an action, the action is regarded as successful. However, when a discrepancy between the predicted and the actual sensorimotor data is detected, a prediction error occurs, and attention will be directed to this discrepancy to correct an erroneous action in real-time [15]. Therefore, the fast and accurate detection of such discrepancies is crucial to perform precise interactions, in the real as well as in virtual

worlds.

The underlying mechanisms and neural foundations of predictive coding have been extensively studied, see for example [16, 11, 17]. The frontal mismatch negativity paradigm (MMN, a type of event-related potential, also known as ERP) using stationary experimental setups has often been employed to probe the predictive brain hypothesis, see [18] for a review. [19] show that the best fitting explanation of MMN activity are computations of a Bayes-optimal generative model, i.e., prediction errors. Recently, [20] demonstrated a passive brain-computer interface (BCI) relying on the frontal MMN generated by prediction errors. In their work, the brain-computer interface decoded a user's intended cursor movement direction on a 6x6 grid. The system regularly probed the user by observing the EEG response to random cursor movements. How severely the random dot movement violated the user's intention was directly reflected in anterior cingulate (ACC) EEG activity.

However, such stationary EEG protocols that require a user to passively observe the presented stimuli largely neglect the embodied cognitive aspects of goal-directed behavior. As a consequence, the cortical activity patterns underlying predictive embodied processes during goal-directed movement are not fully established. How these electrocortical features reflect a perceived loss in physical immersion when interacting with VR/AR is yet to be understood.

In this paper, we address (1) whether the frontal MMN, originating in ACC, is sufficiently robust to reflect visuo-tactile prediction errors in *naturalistic* interaction with virtual worlds, and (2) whether behavioral adaptation, post-error slowing, follows the visuo-tactile prediction errors.

Recently, the Mobile Brain/Body Imaging (MoBI) paradigm has opened new possibilities to investigate multimodal predictors of user behavior and experience [21, 22, 23, 24]. We leveraged MoBI to record synchronous EEG and motion capture data during an interactive VR experience in which we purposefully introduced visuo-haptic mismatches (please see [7] for details). In the current work, we classified trials into two categories: following predicted VR feedback (match) and following visuo-tactile VR glitches (mismatch). We hypothesized a high classification accuracy employing PENS for this two-class separation. Crucially, we hypothesized that the classification would strongly rely on (anterior) midline cingulate EEG source activity [20, 25]. Furthermore, we hypothesized motor behavior to slow down following mismatch trials.

2. Materials & Methods

The overarching idea of our work is to calibrate a classifier that can be applied online to provide information about the realism underlying the interaction with objects in VR. To this end, we designed a study in which participants performed a 3D reach-to-tap task in VR. Our task was inspired by [9]. As a participant reached out to tap an object, they were presented with three sensory feedback modalities (a visual only baseline, visuo-tactile, and visuo-tactile with force-feedback). However, to provoke

Neural Sources Detecting Unrealistic VR Interactions

participants into processing an unrealistic VR interaction, we sometimes provided the feedback prematurely.

In a previous paper, we reported the results of 10 participants experiencing the force-feedback condition and provided a general description of the PEN in increasing levels of haptic immersion [7]. There, we reported a strength modulation of PEN depending on the haptic channels available for interaction.

In the current paper, we report data of a significantly larger sample and excluded trials of the visuo-tactile with force-feedback condition. They were only collected for a subset of the participants (10 out of 19) and were always presented following the counterbalanced conditions of visual only baseline and visuo-tactile. Therefore, the force-feedback condition did not impact the visual only and visuo-tactile contrast. In order to improve real-time classification, we leveraged our ERP-based classification system to localize the network of EEG sources underlying the (linear) separation of unrealistic VR interactions through PEN.

Further, we hypothesized that following the unrealistic situations participants movements are slowed down, indicating a more cautious behavioral approach to the next trial. Therefore, we investigated whether the movement feature ‘tap time’ changed following unrealistic VR interactions.

In the present work, we leveraged this PEN for classification as well as source localization while also modeling visuo-tactile mismatches using the movement feature ‘tap time’.

2.1. Apparatus

The experimental setup, depicted in figure 2a, comprised: (1) a VR headset and a wrist-mounted wearable VIVE tracker, (2) one vibrotactile actuator worn on the fingertip, and (3) a 64-channel EEG system. A medically-compliant EMS device connected via two electrodes was worn on the forearm by a subset of participants, see exclusion statement for this data above.

(1) VR and hand tracking. We used an HTC Vive headset (HTC Corporation, Taoyuan, Taiwan) with the Vive Deluxe Audio Strap and custom EEG cap spacers ‡ to ensure a good fit and less discomfort due to the EEG cap. We used a Vive Tracker, attached to the participant’s wrist, to track their right hand.

(2) Vibrotactile feedback. We used a vibration motor (Model 308-100 from Precision Microdrives), which generates 0.8g at 200Hz. This motor measures 8mm in diameter, making it ideal for the fingertip. The vibration feedback was driven at 70mA by a 2N7000 MOSFET, which was connected to an Arduino output pin at 3V.

(3) EEG Setup. EEG data was recorded from 64 actively amplified electrodes using BrainAmp DC amplifiers from BrainProducts. Electrodes were placed according to the extended 10% system [26]. After fitting the cap, all electrodes were filled with conductive gel to ensure proper conductivity.

‡ <https://grabcad.com/library/adapter-for-vr-eeeg-setups-1>

Neural Sources Detecting Unrealistic VR Interactions

5

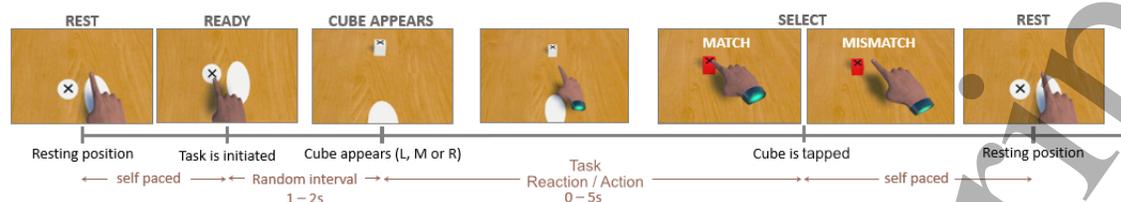


Figure 1. Interaction flow depicting one trial in our 3D reach-to-tap task.

2.2. Task

Participants performed a 3D reach-to-tap task in VR designed with Unity Software (Unity Technologies, San Francisco, USA). The interaction flow of our task, depicted in Figure 1, was as follows: (1) participants moved their hands from the *resting position* to the *ready position*, to indicate they were ready to start the next trial; (2) participants waited for a new target to appear (the time of a new target spawning was randomized between 1-2 s); (3) then, the target (a cube) would appear in one of three possible positions (center, left, right), all equidistant from the participant's *ready position*. A black cross on the top of the cube indicated the location participants were instructed to tap; (4) then, participants completed the task by moving and tapping the target with their index finger. Tapping success was, at least, indicated by a color change of the cube, see below for a detailed explanation of the feedback conditions. (5) After a target was tapped, participants moved back to the *resting position*. Here, they could take a break before the next trial.

To maximize EEG data quality, participants were instructed to remain in a calm upright seated position while carrying out the reaching movement. Further, they were instructed to be precise and keep a comfortable pace. However, no feedback was given on the accuracy and speed of their task completion.

2.3. Interface conditions

Participants performed the task in two additive feedback conditions:

(1) **Visual-only (Visual)**: When participants tapped the cube, it changed its color from white to red (visual feedback.)

(2) **Visual with vibro-tactile (Vibro)**: When participants tapped the cube in the Vibro condition, they received a 100 ms vibro-tactile stimulus with the color change (Visual + vibro-tactile feedback).

In this paper, our key focus was the calibration and source localization of a system detecting unrealistic VR interactions such as visual glitches or visuo-haptic synchronization errors. To maximize statistical power for the focus of our investigation we pooled trials from the two interface conditions. In order to build a stimulus-agnostic classification system detecting unrealistic system behavior, we chose to subject the pooled data to the cross-validation- and localization method.

Neural Sources Detecting Unrealistic VR Interactions

6

2.4. Introducing Visuo-Haptic Mismatches

To allow us to compare the event-related EEG and movement signatures in a realistic vs. unrealistic interaction, we presented participants with two different classes of trials: **match trials (C)** (75% of the trials) and **mismatch trials (M)** (25%). This procedure elicits a prediction mismatch signal in 25% of the trials similar to previous designs investigating the impact of target probabilities [27].

In the **matching** trials, the feedback stimuli were presented upon tapping the object, exactly when participants expected them to occur based on the available visual information (finger touching the target in the virtual environment). In contrast, in the **mismatch** trials, the feedback stimuli were triggered prematurely, which was accomplished by enlarging the invisible radius of tap detection (collision volume around the cube object) by 350%. While in the match trials, we used a collision detection volume of the exact size of the VR cube, in the mismatch trials, we used a larger sphere for collision detection. Our enlargement of the collision detecting volume was based on the study design by Singh et al. [9], in which they showed that VR users can detect a visual mismatch at around 200% of offset from the target. In our pilot tests, we decided to extend the offset to 350% to make the mismatch more obvious so as to provoke more pronounced prediction errors.

We used a match-to-mismatch ratio of 75%-25% of the total trials by modeling our study after previous studies, which also ensured that participants were faced with a detectable unrealistic behavior of the virtual environment [28, 29, 30]. For these unrealistic trials to occur, the participants must first be able to create a stable model of how the VR world operates, thus the VR world cannot behave at a random 50%-50% match-mismatch ratio.

Finally, these match vs. mismatch trials were presented in five pseudo-randomly generated sequences. Following each mismatch trial, the next trial was always a match trial. To reduce the predictability of when the next mismatch trial would occur, the number of consecutive match trials was pseudo-randomized between 1 and 5.

2.5. Experimental design

The experiment consisted of five phases: (1) a setup phase; (2) a calibration phase; (3) a short training phase; (4) the task itself, in all three possible interface conditions, each followed by a subset of items from the IPQ questionnaire (G1, REAL2, SP4 and INV1) [2] and the NASA-TLX [31]. Lastly (5) participants were asked about their experience in the VR and which condition they enjoyed the most.

For training purposes, we asked participants to wear the HTC VIVE VR headset for a maximum of 24 practice trials. Overall, the EEG fitting, calibration, and practice trials took around 30 minutes.

Next, we recorded a within-subjects design with 300 trials for each the Visual and Vibro feedback condition. The order of the Visual and Vibro conditions was randomized across participants. We chose to present the two interface conditions in a blocked

Neural Sources Detecting Unrealistic VR Interactions

7

design. This was done to emphasize the influence of the additional haptic channel while attenuating higher order interactions, such as a prediction error about the upcoming interface condition.

2.6. Dataset

2.6.1. Participants 20 participants (12 female, mean age = 26.7 (sd = 3.6)) were recruited through an online tool provided by the Department of Psychology and Ergonomics and through local listings. Participants were right-handed, had normal or corrected to normal vision and had no experience with VR with vibro-tactile feedback at the fingertip. Participants were compensated with 10 Euros per hour or 1 study participation hour (course credit). Participants were informed of the nature of the experiment, recording and anonymization procedures and signed a consent form approved by the local ethics committee of the Department of Psychology and Ergonomics at the TU Berlin (Ethics approval: GR_10_20180603). Data of the first subject had to be removed from further analyses due to data recording error.

2.6.2. Recordings: Motion Capture and EEG EEG was recorded using 64 active Ag/AgCl electrodes placed according to the extended international 1020 system [26]. The electrode at position FP2 was detached from the cap and placed under the left eye to provide additional information about eye movements (EOG). Impedance was kept under $5k\Omega$ where possible and the EEG was sampled at 500 Hz and amplified using BrainAmp DC amplifiers (Brainproducts GmbH, Gilching, Germany). Hand and head movements were sampled at 90 Hz when coming out of the HTC Vive processing cascade. EEG, motion capture and an experiment marker stream were recorded and synchronized using `labstreaminglayer` §.

2.6.3. Reproducing Results and Data Availability Data, experimental protocol, analyses code including scripts for a reproduction of the presented results and earlier publications are accessible from a comprehensive repository hosted at open science foundation (OSF) ||. BIDS formatted data is hosted on openneuro [32].

2.7. Processing

2.7.1. Behavioral Adaptation Following VR Glitches Motion capture data was filtered with a 6Hz low-pass filter and re-sampled to match the EEG sample rate using MoBILAB routines for concurrent analyses [33]. Subsequently the first derivative was computed and velocity was extracted.

We computed ‘tap time’, the time elapsed between the start of the reaching movement following object spawn and the end of that movement, using the hand velocity time series. The reach onset was detected on the hand velocity time series by moving

§ <https://github.com/sccn/labstreaminglayer>

|| <https://osf.io/x7hnm/>

Neural Sources Detecting Unrealistic VR Interactions

backwards from the velocity peak of the reach movement and selecting the first sample where the velocity fell below 0.05 m/s. The end of the reach was determined as the first sign reversal of the movement change in z-direction, the primary reach direction, following the start of the reach. As such, tap time was detected on the continuous time series and not on the experimental event. This would have been problematic since the premature appearance of the mismatch feedback event would have artificially created an effect.

To assess behavioral adaptation, we modeled the *rate of change* in ‘tap time’ with a linear model. To this end, we computed the difference in ‘tap time’ between subsequent trials and report this *rate of change* as a response to the experimental manipulation [34]. We reported tap time instead of reaction time since participants were not primed, nor did they receive any reward for fast and accurate trial completion. The model ‘*change in tap time* \sim *trial change*’ was fitted using Matlab’s ‘fitlm’ function and assessed using ‘anova’. Trial change was entered as a categorical predictor reflecting whether the current trial change was match to mismatch, mismatch to match or match to match. Since the number of consecutive match trials was pseudo-randomized between 1 and 5 we decided to exclude trials where a mismatch trial occurred again after the first subsequent match trial. These trials corresponded to both the mismatch to match and match to mismatch trial change category. This resulted in the removal of 30 Trials per participant for the behavioral analysis.

2.7.2. Brain activity: EEG Preprocessing, Independent Component Analysis (ICA)

EEG data preprocessing and ICA were performed in Matlab 2019b (MATLAB, The MathWorks Inc., Natick, MA, USA), using the EEGLAB toolbox [35] and custom ‘BeMoBIL Pipeline’ scripts and functions. To detect bad channels for rejection, the ‘FindNoisyChannel’ function was used, which is selecting bad channels by amplitude, the signal to noise ratio and correlation with other channels [36]. Rejected channels were then interpolated while ignoring the EOG channel, and finally re-referenced to average reference (data A). The data was then filtered with a 1 Hz high-pass filter (data B) and a first adaptive mixture independent component analysis, AMICA [37], was used to identify eye related independent components (ICs) which were projected out of the sensor data. For this, the rank was reduced by one for the use of an average reference and further by the number of interpolated channels in the respective data set. To identify eye components, IClab [38] was used, whereas components exceeding a value of 0.7 for the ‘eye’ class were defined as eye components. Then, to detect segments of noisy data, an automated time domain cleaning (see [39]) was performed on narrowly filtered data from 1 to 40 Hz. The data was therefore first split into 1 second long segments for which the mean absolute amplitude and standard deviation of all channels as well as the Mahalanobis distance of all channel mean amplitudes were calculated. All three methods results were then joined together in order to rank all segments. The 12% highest ranking noisy segments were selected for rejection and an additional buffer of ± 0.49 sec was added around each segment resulting in about 15% rejected data for

Neural Sources Detecting Unrealistic VR Interactions

9

each subject. This data was rejected from data B and a second AMICA was calculated on this time domain cleaned data. A dipole fitting procedure was performed for each spatial filter using the 10-20 standard electrode locations and a boundary element head model (BEM) based on the MNI brain (Montreal Neurological Institute, MNI, Montreal, QC, Canada). The spatial filter information was then copied back to the preprocessed, interpolated and average referenced data set (see description of data A above).

To obtain indices of clean tap epochs, we leveraged EEGLAB's 'pop_autorej' function to remove epochs exhibiting large amplitude fluctuations. We used the functions default settings and entered epochs from -3 to 2 seconds surrounding the tap events. On average, 80.7 epochs were rejected (SD = 32.6) amounting to ~13% of the data.

Ultimately, all ICs with a probability smaller than .7 as indicated by the ICLabel 'brain' class were projected out of the data. This resulted in the final dataset including only very likely brain sources and their projections to the channels. Across the study set, 271 independent components were retained forming a representative sample of about 14.3 (SD = 5.0) components per participant. All subsequent EEG analyses were based on these data.

2.7.3. EEG Classifier, Classifier Scalp Projections and Localization of Components relevant to Classification In the current work, we present a processing pipeline with slight updates as compared to our earlier work [7]. To reproduce our previous findings, we report a permutation t-test of the ERP at electrode FCz. Activity at electrode FCz in the time window from 150 to 200 ms post mismatch event featured prominently in our earlier analysis and is frequently considered for MMN paradigms investigating ERPs at the scalp level, for modeling evidence see [19, 40]. For completeness, we report all electrodes that exhibited an amplitude difference at 200 ms post tap event. To this end, we computed a t-test of the amplitudes at 200ms post tap event. To correct for multiple comparisons, the false discovery rate (*fdr*) was computed with $\alpha = 05$ [41]. Channels whose p-value exceeded the *fdr* were plotted, see figure 3.

For classification of single-trial ERPs, we followed the approach introduced by [20]. A regularized linear discriminant analysis classifier was trained per participant with all mismatch trials constituting class 1 and a random sample of an equal number of match trials labeled class 2. Using the open-source toolbox BCILAB ver. 1.4, the classifier was trained on windowed means as features. First, EEG data were re-sampled to 100 Hz and band-pass filtered from 0.1 to 15 Hz. Average amplitudes of all channels in eight sequential 50 ms time windows between 0 and 400 ms after the cube was tapped were extracted as the windowed means feature vectors. A mean baseline taken in the -50 to 0 ms window was subtracted in order to compensate for event classes, match and mismatch, occurring at different stages of the ongoing movement. For robust performance estimation, a 5 x 5 nested cross-validation was used to calculate the shrinkage regularization parameter and assess the classifiers performance.

Classification accuracy was statistically evaluated using a two-sample T-test with

Neural Sources Detecting Unrealistic VR Interactions 10

the mean classifier accuracy per participant across folds and simulated chance level given trial numbers in each class [42].

In order to learn what regions of the brain the classifier specifically relied on, we first transformed the LDA filters at each time window to LDA patterns reflecting a mixture of scalp activations with regards to the discriminative source activity [43]. Subsequently, each independent component's relevance for classification was computed as the dot product of the LDA patterns per time window and the ICA unmixing matrix filter weights [20]. The equivalent current dipole models of independent components were then weighted by their relevance and ultimately visualized via EEGLAB 'dipoleDensity' plots [44]. The Harvard-Oxford atlas was consulted to extract cortical labels of regions of interest [45].

3. Results

Participants reached towards the target object after it appeared on the table. In the match trials without visuo-tactile VR glitches, participants took on average 1.04s (SD = .19) to complete the reach-to-tap.

We created visuo-tactile VR glitches by increasing the (bounding) object volume of the target. Hence, the collision detection registered prematurely. In these mismatch trials, participants took on average .73s (SD = .12) to complete the reach-to-tap, see figure 2b. Hence, increasing the (bounding) object volume for collision detection led to a spatio-temporal mismatch of approximately 300 ms as compared to the congruent, match, condition. The velocity profile in both conditions exhibited a narrow peak during outward reaching with a peak magnitude of .6 m/s and a broader and lower peak when the hand was retracted back to its origin, see figure 2b bottom.

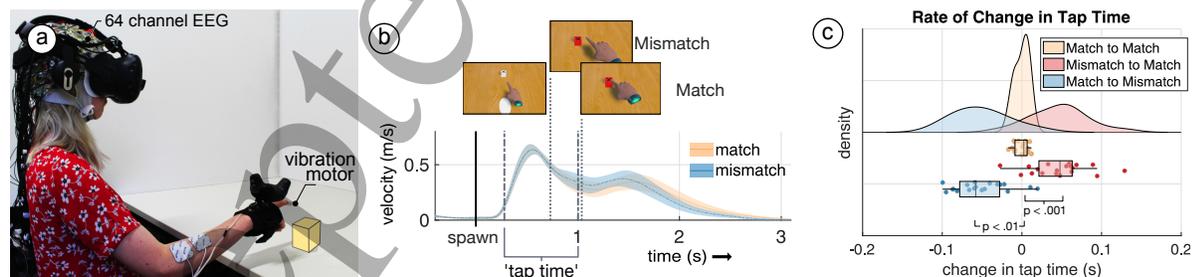


Figure 2. Task Structure and hand velocity profile. **a** Participants were instructed to reach to an appearing cube on a desk in front of them and tap it. They were equipped with a VR headset, a 64 channel EEG cap, electrode spacers and rigid body tracker on the hand. The vibration motor was placed under the fingertip of the index finger. **b** Top: Inside VR view of experimental scene. Bottom: Grand-average velocity with 95% confidence interval of both, match and mismatch conditions with event markers for object 'spawn', tap time start and end as well as moments of object tap in match and mismatch conditions. **c** Distribution of *rate of change* in tap time for the three trial change categories 'match to match', 'mismatch to match' and 'match to mismatch'.

Neural Sources Detecting Unrealistic VR Interactions

11

3.1. Prolonged Tap Time Following VR Glitches

‘Tap time’, the hand movement period from movement start to reaching the object, lasted on average .74s (SD = .15) in match- and .69s (SD = .15) in mismatch trials.

We calculated the *rate of change* in ‘tap time’ as a metric of post-error slowing and observed that trial change categories impacted ‘tap time’ ($F_{(2)} = 53.7, p < .001, R^2 = .66$). Following match trials, ‘tap time’ in the subsequent trial did not change, i.e. 0 ms (SD = 20 ms). However, following mismatch trials, ‘tap time’ was increased in the subsequent trial on average by 47 ms (SD = 37 ms, $t_{18} = 5, p < .001$). For completeness, ‘tap time’ decreased from match to mismatch trials on average by 50 ms (SD = 33 ms, $t_{18} = -5.4, p < .001$).

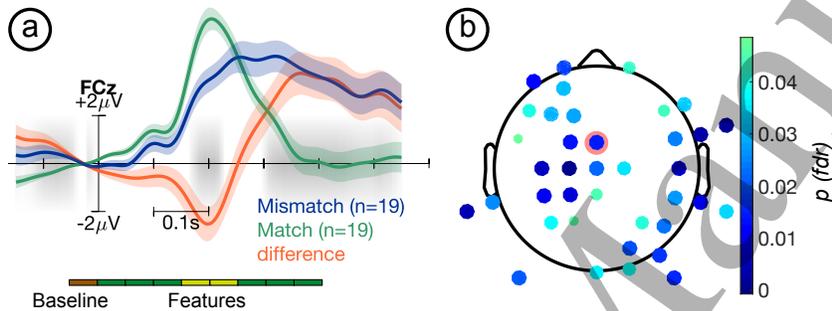


Figure 3. **a:** Grand-average ERP ($n = 19$) of projected source mixtures at electrode FCz with significant class differences marked in grey. Bottom: Time windows used to compute features for classification (all greens). Windows in light green indicate time windows of interest for classifier source localization. **b:** Electrodes with a significant amplitude difference (*fdr* corrected) between match and mismatch trials at 200 ms post tap event. Electrode locations are color scaled by their respective p-value, colder colors correspond to a lower p-value. Scalp location of electrode FCz (see subplot **a**) is highlighted with a red background.

3.2. ~77 % Classification Accuracy Detecting VR Glitches using ERPs

We found significant differences between match and mismatch trials in the grand-average event-related potential (ERP) at several scalp locations, see figure 3 showing the ERP at electrode ‘FCz’ for an example. Hence, we reproduced our previous findings in [7] with an altered processing pipeline. At electrode ‘FCz’, amplitude differences at 200 ms indicated a significant difference between mismatch, i.e. the VR glitch condition and the matching trials ($t_{18} = -5.34, p < .001$). Differences were observed most strongly in the 150-280 ms time window, at 250 ms and in later windows starting at 350 ms, see figure 3.

To assess the potential for single-trial online applications, a discriminative classification system was cross-validated. The system, using windowed mean ERP features, succeeded in detecting VR glitches. Mismatch and match trials were correctly labeled to the corresponding class with an average accuracy of ~77 percent ($SD = 9.12$). The classification accuracy exceeded chance level at ~ 56 percent, $t_{(18)} = 42.1, p < .001$.

3.2.1. *Classification Driven by Midline Cingulate and Occipital EEG Sources* To draw conclusions about the cortical origin of the discriminatory signal we investigated which EEG sources contributed maximally to the classification. With regards to the system's applicability as robust neural interface technology, this source reconstruction served two purposes: (1) Asserting that the classifier did not rely *primarily* on artifact EEG sources, and (2) to gain additional information about the contributing brain regions to allow interpretations about cognitive processing.

In the fourth time window of the eight windowed mean features (150 - 200 ms, see the first light green shaded window at the bottom of figure 3 and in figure 4a top) classifications were driven primarily by activity originating in right lateral parieto-occipital cortical sources (BA19; MNI: $x = 30, y = -70, z = 30$), see figure 4a bottom. In the following time-window (200 - 250 ms, see the second light green shaded window at the bottom of figure 3 and in figure 4b top) the classification signal draw from distributed source activity in occipital areas as well as from sources located in anterior midline cingulate gyrus (near BA23; MNI: $x = 0, y = -10, z = 30$), see figure 4b bottom. Some ocular sources were not classified as such by our automated processing pipeline and carried information relevant to classification in this time window of interest.

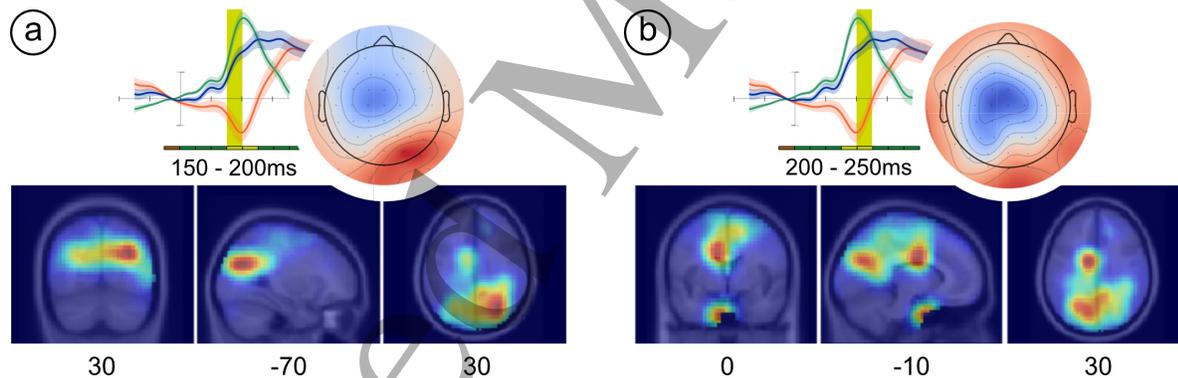


Figure 4. An LDA classifier was trained on eight windowed means of 50 ms size from 0 to 400 ms following the cube tap, see figure 3 bottom. Two classes of synchronous and asynchronous trials were labeled for training and cross-validation. **a, b** Scalp maps of difference-between-classes activity for the 4th (150-200) and 5th (200-250ms) time windows and the equivalent source localization (MNI coordinates of the location of maximum activity).

4. Discussion

With this study, we contributed a new approach to automatically detect conflicts in visuo-tactile sensory integration in VR based on a classifier using ERPs.

Our work aimed at elucidating whether ERP-based classification can help address the challenge of *continuous* labeling of a user's immersion. This work contributes towards the overarching goal to develop a continuous method to validate the effectiveness of haptic devices that foster presence experience.

We achieved a $\sim 77\%$ classification accuracy detecting visuo-tactile glitches in a reach-to-tap task in VR. The midline cingulate cortex as well as a distributed network of parieto-occipital EEG sources enabled the classification success.

We believe our experimental setup, and/or data, can be used to calibrate a classifier that labels unrealistic VR interactions in near real-time. Consider the example of grabbing a hammer in VR with a game controller. When the available sensory channels are misaligned, for example the hammer ‘snaps’ to the virtual hand before the controller’s vibration simulates physical contact, the interaction is labeled ‘unrealistic’ by the ERP-based classifier. When interacting with objects in VR, ERP-based classification is a promising endeavour due to the abundance of events.

4.1. Post-error Adaptation Following Visuo-Tactile Mismatches during Interaction with Virtual Worlds

We hypothesized a trial-to-trial adaptation in movement behavior as an implicit behavioral reflection of the effectiveness of our ‘VR glitch’ experimental manipulation. [34] have shown that correlations of brain signals with global averages of post-error slowing metrics may be moderated by confounding *cognitive* processes, e.g. fluctuating concentration levels. Therefore, we looked at the rate of change in ‘tap time’. Between two subsequent match trials there was no change in ‘tap time’. However, ‘tap time’ increased in the trial following a mismatch trial. Similarly, we observed a decrease in ‘tap time’ in the mismatch trials following match trials. While the effect is comparable in absolute value to the increase in the ‘match to mismatch’ trial change condition, we note that it is challenging to separate the processes underlying these changes in behavior. While one might happen more immediately due to the glitch manipulation, the other one may happen as a delayed consequence of it.

In general, these findings confirmed our manipulation, since the visuo-tactile VR glitch impacted behavior. Consequently, this ruled out the possibility of an automatic behavior in which participants were unaware of the manipulation. Since the task featured 600 trials, this was not unlikely.

We believe our findings further evidence the literature on post-error adaptation, with participants taking a slightly more cautious approach following VR glitches [46]. This speed decline has frequently been observed to facilitate increasing accuracy on subsequent trials [47]. Here, cognitive control processes inhibit motor execution, presumably by closely monitoring, and raising, cortical activity thresholds [48].

In summary, we believe our data captures ‘prediction errors’, violations of goal-directed behavior under consideration of the predicted action outcome. Further, we hypothesize cognitive control mechanisms to monitor and adjust subsequent motor output with one objective being an increase in behavioral accuracy. Only observing minor differences in absolute values between ‘mismatch to match’ and ‘match to mismatch’ may be due to the missing framing of our task: (1) not giving participants an incentive to optimize the speed-accuracy trade-off and (2) not providing any feedback

Neural Sources Detecting Unrealistic VR Interactions 14

on their task performance. Not collecting an accuracy metric is a shortcoming that should be improved in subsequent works, for example as in [49].

4.2. Towards BCI based on Embodied Predictive Coding during Interaction with Virtual Worlds

In order to describe a robust ERP feature representing prediction errors, we localized their EEG source origin and found two loci: (1) the midline cingulate cortex and (2) a distributed network in parieto-occipital areas.

The role of midfrontal EEG source activity has frequently been linked to cognitive control processes [47, 50, 51]. Crucially, these studies employed a stationary setup, limiting participants' interaction with the task environment. However, environmental affordances surpass the visual domain. Particularly in humans, a proclivity to use both hands to act on the environment has emerged and is greatly trusted upon, for example when finding your way in the dark [52, 53, 54]. In fact, many tasks can be completed without concurrently consulting the visual domain, such as typewriting or even reaching for a cup; these typically rely heavily on the tactile and proprioceptive sense and as such are denoted as eyes-free interactions.

In our voluntary, albeit instructed, tapping task, we observed a frontal prediction error negativity, 'PEN', (at electrode FCz) to exhibit a familiar time course as compared to stationary setups reporting MMN, see difference wave in figure 3. In line with the MMN literature, we observed a stronger early negative deflection in mismatch trials as compared to the match trials, see figure 3. [20] report a single 'PEN' source origin in anterior cingulate cortex, however in our study, besides midline cingulate sources, sources in parieto-occipital areas contributed to the successful classification. Previously, [15] observed an effect of prediction errors about the visual consequences of the current motor action in a reaching task in parietal electrodes contralateral to the reaching hand. The authors conclude a role of the dorsal processing stream, processing the 'where' in visuomotor 'PEs'. Interestingly, first evidence now indicates that cortical cross-talk between the motor area corresponding to the active, for example reaching, hand and parietal regions may in fact reflect the strength of the body illusion in VR, the illusion that an avatar's hand is in fact mine [55]. Considering the classifier weighting in our task in the time periods between 150 to 200 ms as well as 200 - 250 ms, we take note of a parieto-occipital reliance for match/mismatch separation in the earlier time window which was followed by a more pronounced weighting on frontal midline sources in a following time window.

Hence, we observed our classifier to first rely on parieto-occipital sources and subsequently on frontal midline sources in the typical time range of the MMN. This may indicate the role of embodied affordances in our immersive reaching task. Attention modulating cognitive control during 'PEs' may be represented in the MMN. In stationary setups or 'motor-passive' paradigms, such as in [20], no countermeasure to correct the 'PE' exists. This may explain a heavy role of midline cingulate activity in classification

with no other sources contributing. However, even for a simple task like reaching for an object, several sensory signals (e.g., visual, tactile and proprioceptive feedback) are continuously gathered and analyzed to efficiently interact with/in a dynamically changing environment. To compensate for the sensory noise within the nervous system and for the uncertainties in the dynamic environment different movement-related sensory cues have to be integrated. To gain an overall representation of the body position, movement and acceleration, the most reliable sensory information must be enhanced while the most noisy ones must be diminished [56], i.e. multisensory integration. With increasing immersion, processing gets more accurate and therefore might trigger a hierarchical cascade of ‘PE’ processing [57]. Here, multisensory integration in parieto-occipital regions precedes action outcome evaluation and cognitive control supported by midline cingulate cortex structures. However, the fact that our classifier relied on parieto-occipital source activity is direct evidence for ‘PE’ processing. One possible explanation is that sensory ‘PEs’ may already be resolved at early stages in the processing cascade instead of an inefficient signal forwarding to frontal brain areas.

4.3. Limitations & Open Challenges

To allow for a full reproduction of our results, we provide the BIDS formatted data as well as all processing code alongside this publication. We chose automated over manual processing. However, problems remain in automated labeling of EEG sources. As evident in figure 4 sources localized to the eyes did contribute to the classification. More stringent vetting of EEG sources could make the results to rely exclusively on brain sources. However, using the eye activity features for classification is useful as long as it is reliable. This is especially relevant when using low-density EEG systems, such as consumer market products and it would be interesting to dissociate the different sources’ contribution to the classification.

One way to validate ‘PEN’ as a correlate of sense of presence would be a correlation with established presence questionnaires [3, 2]. We believe that due to the highly repetitive experimental design and very subtle experimental manipulation, frequently asking questions would not yield valid results in the sense that frequently breaking the ongoing presence experience would bias the very construct we aimed at measuring. Recently [58] reported a correlation of late ERP components in central electrodes with the subjective presence experience. The authors used an auditory irrelevant probe paradigm and showed that ERP fluctuations to auditory distractor probes coincided with presence ratings. This approach relies on probing the user’s overall attentional state. Instead, our proposed approach directly probes the user’s internal model of their environment providing a precise task-relevant marker for neuroadaptive interfaces [59].

Due to the high-level of immersion in VR, classifying event-related activity such as that occurring during object interaction, may be improved by an unfolding of overlapping activity. For example, when picking up a virtual hammer in VR, several ‘events’ may co-occur. A visual event in the background may coincide with the haptic

Neural Sources Detecting Unrealistic VR Interactions 16

event of the controller when making contact with the hammer. The exact timing and rich descriptions of all such co-occurring events readily exist in VR simulations making near real-time overlap correction feasible. Novel approaches to ‘unfold’ such EEG data exist [60]. We believe real-time ERP-based user experience classification in VR will benefit significantly from such overlap correction, amplifying the signal while attenuating the noise.

4.4. Conclusions & Outlook: Towards a Robust Metric of Presence Experience in Virtual Worlds

Midline cingulate EEG sources contributed to prediction error ERPs, ‘PENs’, and may serve as a robust source to detect violations of user’s predictions about the interaction with virtual worlds [7, 8, 20]. This source origin can be specifically and repeatedly probed for real-time BCI purposes, informing the technical system about the user’s mental representation generating the predictions [59, 20]. If follow up studies replicate and extend on our classification success several benefits emerge: (1) the ERP-based measure to continuously evaluate haptic immersion gains significant robustness and reliability. (2) This will in turn motivate further research on the PE paradigm moving towards implicit measures of the user’s subjective experience. (3) We believe this paves the way for fast and reliable real-time adaptation as the EEG feature search space is significantly reduced. However, our results also show a network of distributed parieto-occipital EEG sources contributing to the classification success. This indicates the challenges remaining in scenarios with a higher level of physical immersion.

With this work, we hope to contribute to the design of a new method based on neural interface technology to assess the effectiveness of haptic devices that foster the emergence of presence experience.

Acknowledgments

This research was supported by a grant from the German Federal Ministry of Education and Research (01GQ1511) to KG

References

- [1] Slater M 2009 *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364** 3549–3557
- [2] Schubert T W 2003 *Zeitschrift für Medienpsychologie* **15** 69–71
- [3] Witmer B G and Singer M J 1998 *Presence* **7** 225–240
- [4] Slater M 1999 *Presence* **8** 560–565
- [5] Kiltner K, Groten R and Slater M 2012 The sense of embodiment in virtual reality
- [6] Sanchez-Vives M V, Spanlang B, Frisoli A, Bergamasco M and Slater M 2010 *PLoS One* **5** e10381
- [7] Gehrke L, Akman S, Lopes P, Chen A, Singh A K, Chen H T, Lin C T and Gramann K 2019 Detecting visuo-haptic mismatches in virtual reality using the prediction error negativity of event-related brain potentials *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19 CHI '19* (New York, New York, USA: ACM Press) pp 427:1–427:11

- 1
2
3 *Neural Sources Detecting Unrealistic VR Interactions* 17
4
5 [8] Si-mohammed H, Lopes-dias C, Duarte M, Jeunet C and Scherer R 2020 653–661
6 [9] Singh A K, Chen H T, Cheng Y F, King J T, Ko L W, Gramann K and Lin C T 2018 *IEEE Access*
7 **6** 24617–24624
8 [10] Knill D C and Pouget A 2004 *Trends Neurosci.* **27** 712–719
9 [11] Clark A 2013 *Behav. Brain Sci.* **36** 181–204
10 [12] Friston K 2010 *Nat. Rev. Neurosci.* **11** 127–138
11 [13] Rao R P and Ballard D H 1999 *Nat. Neurosci.* **2** 79–87
12 [14] Friston K 2012 *Cogn. Process.* **13 Suppl 1** S171–7
13 [15] Savoie F A, Thénault F, Whittingstall K and Bernier P M 2018 *Sci. Rep.* **8** 12513
14 [16] Holroyd C B and Coles M G H 2002 *Psychol. Rev.* **109** 679–709
15 [17] Bendixen A, SanMiguel I and Schröger E 2012 *Int. J. Psychophysiol.* **83** 120–131
16 [18] Stefanics G, Kremláček J and Czigler I 2014 *Front. Hum. Neurosci.* **8** 666
17 [19] Lieder F, Daunizeau J, Garrido M I, Friston K J and Stephan K E 2013 *PLoS Comput. Biol.* **9**
18 e1002911
19 [20] Zander T O, Krol L R, Birbaumer N P and Gramann K 2016 *Proc. Natl. Acad. Sci. U. S. A.* **113**
20 14898–14903
21 [21] Makeig S, Gramann K, Jung T P, Sejnowski T J and Poizner H 2009 *Int. J. Psychophysiol.* **73**
22 95–100
23 [22] Gramann K, Gwin J T, Ferris D P, Oie K, Jung T P, Lin C T, Liao L D and Makeig S 2011 *Rev.*
24 *Neurosci.* **22** 593–608
25 [23] Gramann K, Ferris D P, Gwin J and Makeig S 2014 *Int. J. Psychophysiol.* **91** 22–29
26 [24] Jungnickel E, Gehrke L, Klug M and Gramann K 2019 Chapter 10 - MoBI—Mobile Brain/Body
27 imaging *Neuroergonomics* ed Ayaz H and Dehais F (Academic Press) pp 59–63
28 [25] Töllner T, Wang Y, Makeig S, Müller H J, Jung T P and Gramann K 2017 *J. Neurosci.* **37**
29 2504–2515
30 [26] Chatrian G E, Lettich E and Nelson P L 1985 *Am. J. EEG Technol.* **25** 83–92
31 [27] Polich J 2007 *Clin. Neurophysiol.* **118** 2128–2148
32 [28] Liao Y, Gramann K, Feng W, Deák G O and Li H 2011 *Psychophysiology* **48** 1412–1419
33 [29] Wiersema J R, van der Meere J J and Roeyers H 2007 *Neuropsychologia* **45** 1649–1657
34 [30] Donchin E and Coles M G H 1988 *Behav. Brain Sci.* **11** 357–374
35 [31] Hart S G and Staveland L E 1988 Development of NASA-TLX (task load index): Results of
36 empirical and theoretical research *Advances in Psychology* vol 52 ed Hancock P A and Meshkati
37 N (North-Holland) pp 139–183
38 [32] Gehrke L, Akman S, Chen A, Lopes P and Gramann K 2021 "prediction error"
39 [33] Ojeda A, Bigdely-Shamlo N and Makeig S 2014 *Front. Hum. Neurosci.* **8** 121
40 [34] Dutilh G, van Ravenzwaaij D, Nieuwenhuis S, van der Maas H L J, Forstmann B U and
41 Wagenmakers E J 2012 *J. Math. Psychol.* **56** 208–216
42 [35] Delorme A and Makeig S 2004 *J. Neurosci. Methods* **134** 9–21
43 [36] Bigdely-Shamlo N, Mullen T, Kothe C, Su K M and Robbins K A 2015 *Front. Neuroinform.* **9** 16
44 [37] Palmer J, Kreutz-Delgado K and Makeig S 2011 *San Diego, CA: Technical report, Swartz Center*
45 *for Computational Neuroscience* 1–15
46 [38] Pion-Tonachini L, Kreutz-Delgado K and Makeig S 2019 *Neuroimage* **198** 181–197
47 [39] Gramann K, Hohlefeld F U, Gehrke L and Klug M 2021 *Scientific reports* **11** 1–12
48 [40] Lieder F, Stephan K E, Daunizeau J, Garrido M I and Friston K J 2013 *PLoS Comput. Biol.* **9**
49 e1003288
50 [41] Benjamini Y and Hochberg Y 1995 *J. R. Stat. Soc.* **57** 289–300
51 [42] Müller-Putz G R, Scherer R, Brunner C, Leeb R and Pfurtscheller G 2007 Better than random: a
52 closer look on BCI results *2007 1st COST Neuromath Workgroup Meeting, Rome, Italy*
53 [43] Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J D, Blankertz B and Bießmann F 2014
54 *Neuroimage* **87** 96–110
55 [44] Krol L, Mousavi M, De Sa V and Zander T 2018 Towards classifier visualisation in 3D source space
56
57
58
59
60

1
2
3 *Neural Sources Detecting Unrealistic VR Interactions* 18

- 4
5 *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* pp 71–76
- 6 [45] Makris N, Goldstein J M, Kennedy D, Hodge S M, Caviness V S, Faraone S V, Tsuang M T and
7 Seidman L J 2006 *Schizophr. Res.* **83** 155–171
- 8 [46] Rabbitt P and Rodgers B 1977 *Q. J. Exp. Psychol.* **29** 727–743
- 9 [47] Ridderinkhof K R, Ullsperger M, Crone E A and Nieuwenhuis S 2004 *Science* **306** 443–447
- 10 [48] Botvinick M M, Braver T S, Barch D M, Carter C S and Cohen J D 2001 *Psychol. Rev.* **108**
11 624–652
- 12 [49] Purcell B A and Kiani R 2016 *Neuron* **89** 658–671
- 13 [50] Cavanagh J F and Frank M J 2014 *Trends Cogn. Sci.* **18** 414–421
- 14 [51] Cooper P S, Karayanidis F, McKewen M, McLellan-Hall S, Wong A S W, Skippen P and Cavanagh
15 J F 2019 *Neuroimage* **189** 130–140
- 16 [52] Gehrke L, Iversen J R, Makeig S and Gramann K 2018 The invisible maze task (IMT): Interactive
17 exploration of sparse virtual environments to investigate Action-Driven formation of spatial
18 representations *Spatial Cognition XI* vol 11034 LNAI ed Creem-Regehr S, Schöning J and Klippel
19 A (Cham: Springer International Publishing) pp 293–310
- 20 [53] Gehrke L and Gramann K 2021 *Eur. J. Neurosci.* 1–18
- 21 [54] Miyakoshi M, Gehrke L, Gramann K, Makeig S and Iversen J 2021 *Eur. J. Neurosci.*
- 22 [55] Casula E P, Tieri G, Rocchi L, Pezzetta R, Maiella M, Pavone E F, Aglioti S M and Koch G 2022
23 *J. Neurosci.* **42** 692–701
- 24 [56] Fetsch C R, Pouget A, DeAngelis G C and Angelaki D E 2011 *Nat. Neurosci.* **15** 146–154
- 25 [57] Singh A K, Gramann K, Chen H T and Lin C T 2021 *Neuroimage* **226** 117578
- 26 [58] Grassini S, Laumann K, Thorp S and Topranin V d M 2021 *Brain Behav.* 1–16
- 27 [59] Krol L R, Haselager P and Zander T O 2020 *J. Neural Eng.* **17** 012001
- 28 [60] Ehinger B V and Dimigen O 2019 *PeerJ* **7** e7838
- 29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60