

# Detecting System Errors in Virtual Reality Using EEG Through Error-Related Potentials

Hakim Si-Mohammed<sup>\*†</sup>  
Inria, Univ. Rennes, CNRS, IRISA  
Rennes, France

Catarina Lopes-Dias<sup>\*</sup>  
Institute of Neural Engineering  
Graz University of Technology  
Graz, Austria

Maria Duarte  
Faculty of Science  
University of Lisbon  
Lisbon, Portugal

Ferran Argelaguet  
Inria, Univ. Rennes, CNRS, IRISA  
Rennes, France

Camille Jeunet  
CNRS, CLLE Lab, Univ. Toulouse  
Toulouse, France

Géry Casiez  
Univ. Lille, CNRS, Centrale Lille, Inria  
Lille, France

Gernot R Müller-Putz<sup>‡</sup>  
Institute of Neural Engineering  
Graz University of Technology  
Graz, Austria

Anatole Lécuyer<sup>§</sup>  
Inria, Univ. Rennes, CNRS, IRISA  
Rennes, France

Reinhold Scherer<sup>¶</sup>  
School of Computer Science  
and Electronic Engineering  
University of Essex  
Colchester, United Kingdom

## ABSTRACT

When persons interact with the environment and experience or witness an error (e.g. an unexpected event), a specific brain pattern, known as error-related potential (ErrP) can be observed in the electroencephalographic signals (EEG). Virtual Reality (VR) technology enables users to interact with computer-generated simulated environments and to provide multi-modal sensory feedback. Using VR systems can, however, be error-prone. In this paper, we investigate the presence of ErrPs when Virtual Reality users face 3 types of visualization errors: (Te) tracking errors when manipulating virtual objects, (Fe) feedback errors, and (Be) background anomalies. We conducted an experiment in which 15 participants were exposed to the 3 types of errors while performing a center-out pick and place task in virtual reality. The results showed that tracking errors generate error-related potentials, the other types of errors did not generate such discernible patterns. In addition, we show that it is possible to detect the ErrPs generated by tracking losses in single trial, with an accuracy of 85%. This constitutes a first step towards the automatic detection of error-related potentials in VR applications, paving the way to the design of adaptive and self-corrective VR/AR applications by exploiting information directly from the user's brain.

**Index Terms:** Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

## 1 INTRODUCTION

Virtual Reality enables to design computer generated environments inside of which, users can be immersed and interact using new types of input devices [17]. One key aspect of Virtual Reality is the sense of presence [33] i.e the sense of "being there" in the virtual environment, which can be lessened by different sorts of errors amongst which, tracking and rendering errors. These errors can sometimes be difficult to automatically detect. In this paper,

we propose to take a first step towards the use of Brain-Computer Interfaces (BCIs) in order to detect the occurrence of these errors from users' brains.

BCIs convert users' cognitive states into information. Various studies [1, 3, 29, 30] have revealed the possibility to estimate one's mental activity (e.g., movement imagination, mental 3D rotations) or the mental state (e.g., mental workload, engagement) from the measurement of their brain activity. One of the most popular imaging techniques is the electroencephalography (EEG), which measures the electrical activity of the brain from the scalp. The possibility to access this information provides new opportunities for improving Virtual Reality interaction by pro-actively adapting to the users depending on their mental state [22].

In particular, one neurophysiological signature, the error-related potential (ErrP) [4], holds great promise for improving interactive systems, as it appears shortly after an erroneous event onset. The generation of an ErrP happens when the user commits or witnesses an error and can be observed in the user's EEG signals as a specific EEG pattern. Yet, the concept of erroneous event is relative to the user's perspective, and the amplitude of the pattern has been shown to be correlated with the gravity or impact of the error as well as with the engagement of the user in the task [8]. Nevertheless, studies [9] have shown that it was stable across time for each subject, and that it was possible to detect ErrP in single trials.

Error-related potentials can be particularly interesting in human-computer interaction (HCI) contexts [4]. Being able to detect when a user perceives an error without requiring an explicit communication can be useful in many scenarios [38]. For example, it could be used to automatically assess the robustness of a system (e.g. linking the state of the system with erroneous events) or it could be used directly in the interaction process (e.g. automatically correct application or user errors). However, most of the existing works involve simple, unrealistic scenarios and do not study ErrPs in an ecological environment (realistic environment, involving commonly found tasks) [4, 9, 20, 25, 34]. Very few papers have investigated the presence of and the possibility to detect ErrPs in Virtual or Augmented Reality. Besides, to the best of the authors' knowledge, none have reported single trial classification of ErrPs in a VR scenario, meaning the automatic detection of ErrPs on single epochs rather than grand average analysis, nor have performed a systematic study on the different types of errors originating from the system.

Therefore, in this paper, we study the presence and the detection of ErrPs originating from Virtual Reality systems. In particular: what happens (1) when the tracking of an object being manipulated

<sup>\*</sup>These authors contributed equally

<sup>†</sup>e-mail: hakim.si-mohammed@inria.fr

<sup>‡</sup>e-mail: gernot.mueller@tugraz.at

<sup>§</sup>e-mail: anatole.lecuyer@inria.fr

<sup>¶</sup>e-mail: r.scherer@essex.ac.uk

by the user is lost; (2) when the system's response does not match the user's expectations; and (3) when an unrealistic event happens in the background. The possibility to detect the occurrence of these events from the users' brain activity paves the way to new means of improving interaction in Virtual/Augmented Reality. For the users, it could allow the design of mechanisms to automatically correct the system's behavior when they perceive an error, and consequently increase the systems' robustness. It could also help the systems' designers by providing them with a tool to better diagnose and review the interaction scenarios. One could think about automatically flagging the moments when users perceive errors, and marking them as points of interests.

In order to answer the questions mentioned herein-above, we designed a user study the aim of which was to highlight the neurophysiological pattern arising from situations representing 3 types of errors in VR/AR: (1) Tracking errors, which prevent the user from accomplishing a task—they were represented by a loss of tracking; (2) Feedback errors, corresponding to the situation where the obtained feedback is erroneous; and (3) Background anomalies, corresponding to the situation where an unrealistic event, unrelated to the task, happens in the background of the VR scene.

The remainder of the paper is structured as follows: Section 2 provides an overview of previous work on error-related potentials, and their use for improving Human-Computer Interaction, with a particular focus on Virtual Reality applications. Section 3 describes the user study that was conducted to investigate the presence of error-related potentials under three conditions of system errors in VR. Section 3 presents the materials and methods. The results are presented and summarized in Section 4 and discussed in Section 5.

## 2 RELATED WORK

The first studies reporting a neurophysiological response to errors, were conducted in the early 90's [7, 12]. They demonstrated that shortly after subjects committed errors in a speed response choice task, a characteristic EEG event-related potential (ERP) was elicited. This ERP was designated as *error-related negativity* (ERN). This ERN mainly consisted in a negative potential deflection over the fronto-central scalp areas appearing 50 to 100ms after a subject's erroneous response, and was followed by a centro-parietal positive deflection [8].

Later studies further showed that the amplitude of the ERN was modulated by the importance of the error with regards to the given task [11], as well as with the subjective awareness of the error [8, 24, 35]. In addition to the ERN, a similar EEG pattern was shown to appear after an erroneous feedback. Called the *feedback-related negativity* (FRN) this potential appears between 200 and 300 ms after the feedback onset [14]. Today, these patterns are grouped under the term *error-related potential* (ErrP) which has become widespread in the HCI community [4].

In the literature, at least 4 different types of ErrPs have been introduced, depending on the circumstances in which they occur, namely: (1) Execution errors, which correspond to errors made by the user during the execution of a task ; (2) Outcome error, which corresponds to the situation where the outcome of an action differs from what was expected ; (3) Observation errors, which correspond to the observation of another person or agent committing an error. (4) Interaction errors, which correspond to the execution of an unintended action by a HCI. Here, we study the errors originating from the interactive system, which fall in the "interaction errors" category.

### 2.1 Error-Related Potentials for Human-Computer Interaction

Originally, error-related potentials have been described in the context of BCIs by Schalk and colleagues [32]. Error-related potentials can be used as a corrective signal, to correct a BCI's output when this

does not provide the intended command [10, 31] or in an adaptive manner, reducing the possibility of future errors [18, 19].

Additionally, the existence and decoding of ErrPs in asynchronous scenarios has been established [16, 20, 21, 25, 34]. Interested readers may refer to [4] for a more comprehensive review on error-related potentials.

These studies, suggest the possibility to detect and exploit ErrPs in interaction scenarios. However, most of them were conducted in non-ecological scenarios conditions. Recently, ErrPs have been studied in real-world situations, like driving a car [39], indicating the feasibility of decoding ErrPs in ecological scenarios.

### 2.2 Error-Related Potentials in Virtual Reality

A very limited number of papers addressed the presence and the detection of error-related potentials in realistic Virtual Reality environments.

Padrao et al. [27] were among the first to investigate the presence of error-related potentials in VR. Their objective was to study the neurophysiological correlates of violating agency, and the differences between self-generated and external errors. Their experiment consisted in a fast selection task, involving arm movement towards the left or the right. The self-generated errors corresponded to errors made by the subjects by choosing the wrong direction, while the externally generated errors corresponded to a wrong feedback from the avatar, with the virtual arm going to a different direction from the subject's arm. Their results on 24 subjects show a clear difference between the neurophysiological responses corresponding to the 2 conditions. While the self-generated errors displayed a fronto-central negativity around 100 ms after the error onset, the errors generated by the avatar elicited a parietal negativity around 400 ms after the error onset.

Yazmir et al. [37] investigated the presence of error-related potentials in the brain activity of subjects submitted to errors in a haptic-visual task. The participants were asked to move a virtual object horizontally from an origin to a destination using a Phantom haptic arm. The path was obstructed by a large cylinder so that the participants would lose sight of the target behind the cylinder. The errors were generated by tweaking the vertical position of the target when reappearing from behind the cylinder. In the correct trials, the object would reappear at the same vertical level, while it would be randomly translated vertically in the error trials. Their results on 5 subjects suggests the presence of an ERN shortly after the error onset.

Later, in [36] Yazmir et al. studied and compared the EEG correlates for success and failure in a "tennis-like" throwing game. The participants had to play the game against the computer. They considered 2 types of events: (1) Success or "hit", corresponding to the situation where the players scored a goal, meaning that they hit the ball and that the computer missed it. (2) Failure or "miss", corresponding to the situation where the players did not score, i.e. the computer repelled the ball. Their results show that negative peaks appeared in both of the events with different amplitudes and latencies.

In 2018, Pezzetta et al. [28] investigated the presence of error-related potentials in a situation where the errors were more frequent than correct trials. Their goal was to disentangle and assess that ERN were indeed associated with a perceived error rather than with the rarity of the errors. They combined EEG measurement with a CAVE (Cave Virtual Environment) in an observation task, where participants had to watch a 1-st person perspective avatar realizing a grasping task. In the correct trials, the virtual arm would succeed in grasping a virtual glass, whereas in the erroneous trials it would miss the glass. The originality of this work, lies in the fact that the proportion of errors was higher than the correct trials (70% and 30% respectively). Nevertheless, their results on 25 subjects show the eliciting of a fronto-central negativity around 300 ms after the error



Figure 1: Experimental setup used in our study. Left : participant wearing both an EEG cap (ANTNeuro EEGO sports) and an HTC Vive VR headset. Right : virtual environment. Participants were seated in front of a virtual table on which three targets were displayed. The virtual object (brown) was placed in the middle of the three targets and had to be picked up and placed on one target.

onset, demonstrating the presence of observation errors in VR as well as the fact that despite their higher probability, erroneous events still elicited error-related potentials. More recently, [13] explored the use of ErrPs in order to detect conflicts in visuo-haptic integration. Their results suggest a more pronounced early negativity during the trials where the visual and haptic feedback did not match.

The previous work clearly suggests the presence of error-related potentials in the brain activity of subjects facing errors in VR. However, none of them systematically studied the different kinds of errors that can emerge from a VR system.

### 3 MATERIALS AND METHODS

Our goal is to investigate the presence of error-related potentials in different erroneous situations happening in VR. In particular, we identified 3 types of situations that potentially elicit error-related potentials: (Te) The loss of tracking of a manipulated object ; (Fe) An unexpected or erroneous feedback ; (Be) An unrealistic background anomaly. To assess whether or not these conditions elicit error-related potentials, we conducted a user study to record EEG activity under each one of them. From the literature and previous work, our hypotheses were that (Te) would elicit Event-Related Negativity (ERN) associated with execution errors, that (Fe) would trigger a Feedback Related Negativity whereas (Be) would trigger an ERN for the users who notice the anomaly and that it is possible to classify and detect the elicited ErrPs in single trial.

#### 3.1 Apparatus and participants

Fifteen healthy and not color-blind participants (7 women, 8 men) took part in the experiment (aged mean= 24.8 years, std=2.9). This study was designed in accordance to the declaration of Helsinki, data was anonymized and all subjects provided prior written informed consent. They were also informed of the informed that they could stop the experiment at any moment if they wished, with no consequences. The VR environment was displayed using used a HTC Vive device, which is a commercially available virtual reality device and can be bought and used by general population with no instruction or supervision.

EEG measurements were performed with an ANT Neuro equipment which has a CE certificate as a medical equipment. The equipment consists of a cap with electrodes, which is put over the participant's head. The electrodes on the cap were filled with a non-abrasive gel.

Sixty-four (64) EEG channels were recorded, grounded to the forehead, and referenced to CPz. The signal was amplified using an ANTNeuro Amplifier and recorded at 512 Hz.

The virtual scene and the experiment protocol were designed using the Unity software and C# scripting. The virtual scene was displayed using an HTC Vive Head-Mounted Display (HMD). The mounting of the HMD on top of the EEG cap was made in a way to avoid pressure on the fronto-central electrodes, by only fixating the HMD using the lateral elastic bands and leaving the upper elastic band looser.

The EEG data and experiment events were recorded using the Lab Recorder, which stores data sent using the Lab Streaming Layer (LSL) protocol [15].

#### 3.2 Experimental protocol

After the participants signed the consent form, they were equipped with the EEG cap and the VR headset. The virtual environment consisted of an office with a table placed in the middle, associated with a real table to provide participants with a passive haptic feedback (see Figure 1). Participants sat at the table, and had to perform a center-out pick and place task, using a Vive controller for grabbing, moving and dropping the objects using the controller trigger.

The experiment consisted in 4 conditions corresponding to the types of trials:

- **Correct:** These trials correspond to the normal realization of the task. In these trials, the participant grabbed the object and dropped it in the correct shape. They were provided with the correct feedback (Fc) (Figure 2: Top left).
- **Tracking error (Te):** These trials correspond to the situation in which the system lost tracking of the object before the participant placed it into the target shape. At a random point within 25% and 75% of the distance between the initial position and the destination shape, the object froze and was detached from the participant's hand. In this event, the participant was asked not to go back to the object and just wait for the next trial (Figure 2: Top right).
- **Feedback error (Fe):** These trials correspond to the participant getting an unexpected, erroneous feedback after completing the task. Even if the object was placed in the correct shape, the shape would turn *red*, which corresponds to a wrong feedback (Figure 2: Bottom left). No further instructions were given to the participant.
- **Background anomaly (Be):** These trials correspond to the situation where an unrealistic anomaly, not related to the task, appears in the background. Randomly, when the object was

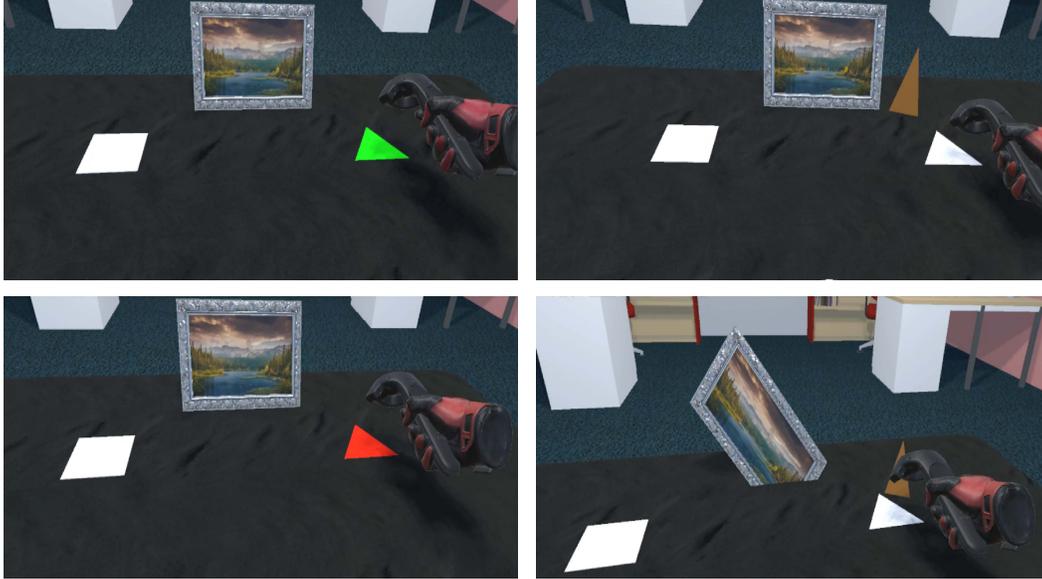


Figure 2: Illustration of 4 experimental conditions. Top left: *Correct Feedback (Fc)* condition. The participant receives a correct feedback after the completion of the task. Bottom left: *Feedback error (Fe)* condition. The participant receives a wrong feedback after the completion of the task. Top right: *Tracking error (Te)* condition. The object freezes and is detached from the participant's hand. Bottom right: *Background anomaly (Be)*. The picture frame in front of the participant flips and gets into an unrealistic position, penetrating the table.

within 25% and 75% of the total distance between the origin and the target shape, a frame placed in front of the participant on the table would flip and stay in an unrealistic position (crossing the table) until the end of the trial (see Figure 2: Bottom right).

Each trial started with the appearance of one of the three objects (a parallelepiped, a cylinder or a tetrahedron) placed at the center of the table. The participant's task was to grab the object and place it on the corresponding object's base shape (a square, a circle or a triangle respectively). After the participant placed the object on the corresponding shape, they were provided with a colored feedback for 2 s. The selected shape turned *green* if it was correctly associated with the object, and turned *red* otherwise. The timeline of the task is summarized in Figure 3.

The initial position of the object was equidistant to the 3 destination shapes (approx 25cm from the central position), and the whole setup could fit the user's field of view, so that the participant did not have to move the head to perform the task and thus to reduce muscle artifacts, which degrade EEG signals.

The experiment consisted of 10 runs of 39 trials (390 trials in total). In total there were approximately 69% of correct trials and 31% of erroneous trials. Each run consisted of 27 correct trials and 12 erroneous trials: 5 trials of condition *Te* (13%), 5 trials of condition *Fe* (13%) and 2 trials of condition *Be* (5%). For each run, the sequence of correct and error trials as well as the object in each trial were pseudo-randomly distributed, i.e., a random sequence was generated once and the same sequence was applied for all participants. The total duration of the experiment was around 1h30, including the time to set up the EEG cap, and, between runs, the participants could take breaks for as long as they wanted.

### 3.3 Electrophysiological analysis

Before being preprocessed, the signal was resampled to 256 Hz and re-referenced to the average of left and right mastoids (using electrodes P7 and P8). In order to perform artifact rejection, the signal was filtered between 1 and 30 Hz and only the channels F3, F1, Fz,

F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CP2 and CP4 were considered. Epochs were rejected by threshold ( $\pm 125$  microvolts), probability and kurtosis. Regarding the methods based on probability and kurtosis, a threshold of 5 standard deviations was used as excluding criteria [6]. For 14 participants (all except participant 10), up to 12.5% of trials per condition were removed. For participant 10, up to 25% of trials per condition were removed.

For the electrophysiological analysis, the data was band-pass filtered between 1 and 10 Hz with a zero-phase Butterworth filter of order 4.

In order to compare erroneous and correct responses, we defined three types of onset events in correct trials (*Tc*, *Fc* and *Bc*) occurring at comparable time-points to the error onsets in error trials (*Te*, *Fe* and *Be*, respectively).

For trials of type *Fc*, we considered as onset the moment in which the target shape turned green (correct feedback). Trials of type *Tc* and *Bc* had no intrinsic onset, so we considered virtual onsets corresponding to each participant's average delay of the corresponding error onsets (*Te* and *Be*, respectively), in relation to the moment in which the object was grabbed.

To display the electrophysiological analysis of all conditions, we considered a 1 s window of starting -0.2 s before the corresponding onset and ending 0.8 s after it. The results are displayed as topoplots, calculated using all channels, and as EEG traces at channel *FCz* due to it being a representative channel.

### 3.4 Single Trial Classification

For the classification of the tracking condition (*Te* vs *Tc*), we considered 250 ms epochs starting 0.150 ms after either the error onset (*Te*) in the error trials or the virtual onset (*Tc*) in correct trials. For the classification of the feedback condition (*Fe* vs *Fc*), we considered 250 ms epochs starting 0.100 ms after either the error onset (*Fe*) in the error trials or the virtual onset (*Fc*) in correct trials. In both situations, the amplitudes of the channels *FCz* and *Cz* within these windows were used as features to train a shrinkage-linear discriminant analysis (LDA) classifier [2]. To evaluate the classification

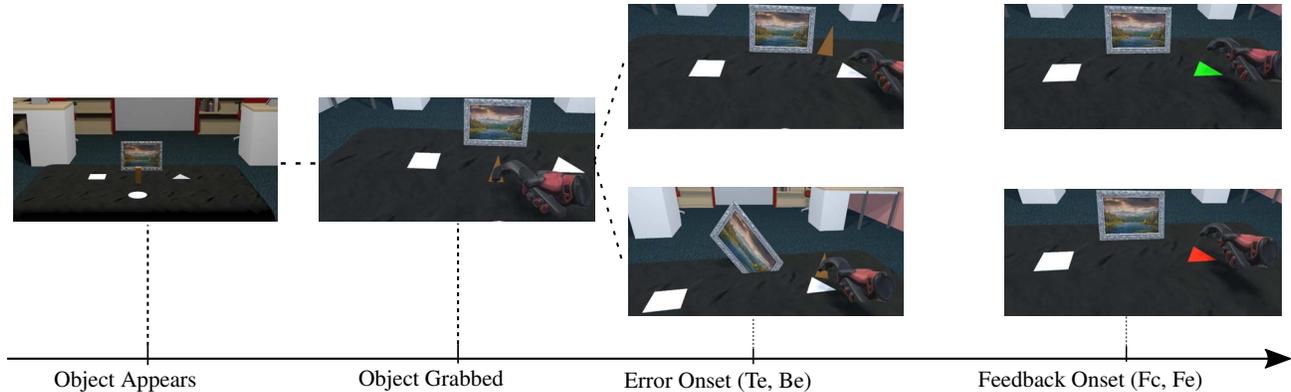


Figure 3: Timeline of a trial with the different possible events. At the beginning of the trial, the object appears at the center of the table and the participant had to grab the object. During the *tracking error* ( $T_e$ ) trials, the object freezes (top). During *background anomaly* ( $B_e$ ) trials, the frame on the tables flips (bottom). When the participant achieves the task, a correct feedback (top) is given in the *correct trials*, and an error feedback (bottom) is given on *Feedback error* ( $F_e$ ) trials.

results we performed 10 times a 5-fold cross-validation. To address the imbalance between the classes, we considered, in each fold of the cross-validation, only a randomly selected subset of the correct trials, matching the number of error trials. This way, the classification was performed using a balanced data set.

For the classification of the tracking condition ( $T_e$  vs  $T_c$ ),  $T_e$  was considered as the positive class and  $T_c$  as the negative class. Similarly, for the classification of the feedback condition ( $F_e$  vs  $F_c$ ),  $F_e$  was defined as the positive class and  $F_c$  as the negative class. The classification results are reported in terms of accuracy, true positive rate (TPR) and true negative rate (TNR). The theoretical chance-level for all measures is 50%. The significance level (SL) was individually calculated for every participant, as the upper-limit of the 95% confidence interval of the theoretical chance-level, calculated based on the binomial cumulative distribution [5, 23].

## 4 RESULTS

Here we present the results regarding the electrophysiological analysis and the single-trial classification.

### 4.1 Electrophysiological Analysis

#### Tracking Error vs Correct Tracking Analysis ( $T_e$ and $T_c$ )

The results of the electrophysiological analysis of the tracking conditions ( $T_e$  and  $T_c$ ) are summarized in Figure 4. Figure 4 (Left) depicts the topoplots of the grand average correct ( $T_c$ ) and error ( $T_e$ ) signals from  $t = 0$  s to  $t = 0.55$  s in intervals of 50 ms (top and bottom rows, respectively).

Figure 4 (Right), the green line represents the grand average signal corresponding to correct trials ( $T_c$ ) at channel FCz. The red line represents the grand average signal corresponding to error trials ( $T_e$ ) at channel FCz. The correct epochs are not aligned to a real event but rather to a virtual onset, and therefore no event-related potential is expected. The grand average correct signal obtained is characterized by a relatively flat curve. The grand average error signal obtained is characterized by a positive peak at time  $t = 0.206$  s with amplitude  $0.4 \mu\text{V}$ , followed by a negative peak at time  $t = 0.257$  s with amplitude  $-4.0 \mu\text{V}$ . This is followed by a positive peak at time  $t = 0.335$  s with amplitude  $8.6 \mu\text{V}$ , by a negative peak at time  $t = 0.421$  s with amplitude  $0.4 \mu\text{V}$  and, finally, by a positive peak at time  $t = 0.491$  s with an amplitude of  $2.5 \mu\text{V}$ .

#### Feedback Error vs Correct Feedback Analysis ( $F_e$ and $F_c$ )

The results of the electrophysiological analysis of the feedback conditions ( $F_e$  and  $F_c$ ) are shown in Figure 5. Figure 5 (Left) depicts

the topoplots of the grand average correct ( $F_c$ ) and error ( $F_e$ ) signals from  $t = 0$  s to  $t = 0.55$  s in intervals of 50 ms (top and bottom rows, respectively).

Figure 5 (Right), the green line represents the grand average signal corresponding to correct trials ( $F_c$ ) at channel FCz. The red line represents the grand average signal corresponding to error trials ( $F_e$ ) at channel FCz.

Both correct and erroneous feedback trials are time-locked to an event (start of feedback) and were expected to elicit an event-related potential.

The grand average correct signal obtained is characterized by a positive peak at time  $t = 0.257$  s with amplitude  $3.05 \mu\text{V}$ . The grand average error signal obtained is characterized by a negative peak at time  $t = 0.148$  s with amplitude  $-2.52 \mu\text{V}$ , followed by a positive peak at time  $t = 0.253$  s with amplitude  $5.22 \mu\text{V}$ .

#### Background Anomaly Trials Analysis ( $B_e$ and $B_c$ )

Figure 6 displays results of the electrophysiological analysis of the background anomaly conditions ( $B_e$  and  $B_c$ ). Figure 6 (Left) depicts the topoplots of the grand average correct ( $F_c$ ) and error ( $F_e$ ) signals from  $t = 0$  s to  $t = 0.55$  s in intervals of 50 ms (top and bottom rows, respectively).

Figure 6 (Right), the green line represents the grand average signal corresponding to correct trials ( $B_c$ ) at channel FCz. The red line represents the grand average signal corresponding to error trials ( $B_e$ ) at channel FCz. Correct epochs are not aligned to a real event but rather to a virtual onset and therefore were not expected to be associated with an event-related potential. Error trials were time-locked to the moment in which the frame flipped and were expected to cause an event-related potential, in case participants noticed the occurrence. The grand average correct signal obtained is characterized by a relatively flat curve. The grand average error signal obtained is also characterized by a relatively flat curve, although a small positive peak at time  $t = 0.292$  s with amplitude  $1.43 \mu\text{V}$  is visible.

### 4.2 Single Trial Classification

In order to illustrate the feasibility of detecting error-related potentials in single trial in a VR experiment, we decided to test the single trial classification on the tracking condition ( $T_e$  vs  $T_c$ ) and on the feedback condition ( $F_e$  vs  $F_c$ ). The EEG signals regarding the tracking condition were very distinct and  $T_e$  condition presented a clear ErrP pattern. These were good indicators that these could lead to a good classification performance. The signals in the feedback

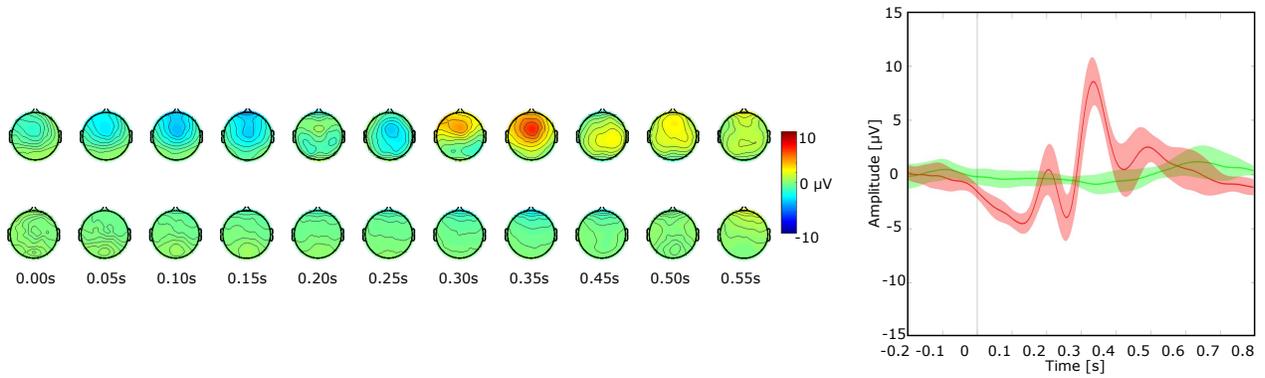


Figure 4: (Left) Topoplots of the grand average error (Te) and correct (Tc) conditions (top and bottom rows respectively), displayed from  $t = 0$  s to  $t = 0.55$  s in intervals of 50 ms. (Right) Grand average correct (Tc) and error (Te) signals at channel FCz (green and red lines, respectively). The shaded areas represent the 95% confidence interval for the average curves. The black vertical line represents the onset events.

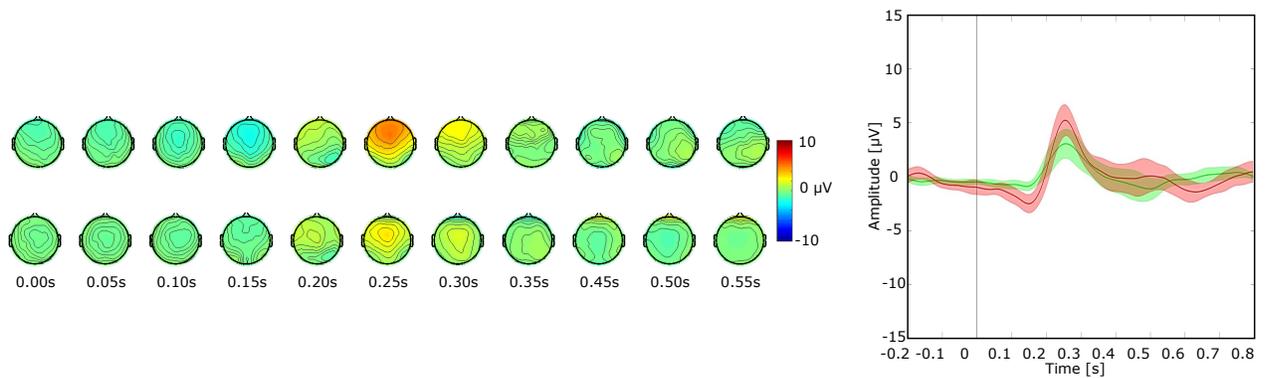


Figure 5: (Left) Topoplots of the grand average error (Fe) and correct (Fc) conditions (top and bottom rows, respectively), displayed from  $t = 0$  s to  $t = 0.55$  s in intervals of 50 ms. (Right) Grand average correct (Fc) and error (Fe) signals at channel FCz (green and red lines, respectively). The shaded areas represent the 95% confidence interval for the average curves. The black vertical line represents the onset events.

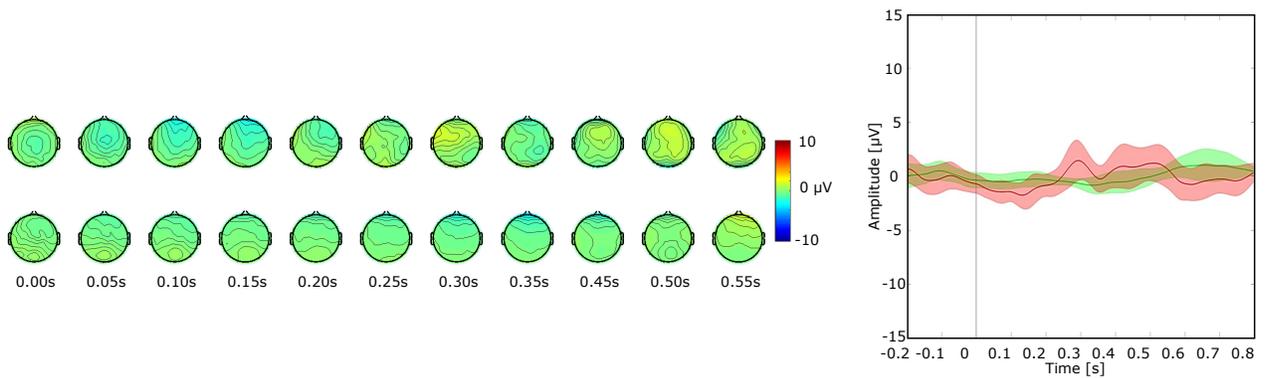


Figure 6: (Left) Topoplots of the grand average error (Be) and correct (Bc) conditions (top and bottom rows, respectively), displayed from  $t = 0$  s to  $t = 0.55$  s in intervals of 50 ms. (Right) Grand average correct (Bc) and error (Be) signals at channel FCz (green and red lines, respectively). The shaded areas represent the 95% confidence interval for the average curves. The black vertical line represents the onset events.

Participants	Accuracy (%)	TNR (%)	TPR (%)	SL (%)
P1*	87.1 ± 7.5	90.0 ± 10.0	84.3 ± 10.4	60.2
P2*	91.8 ± 6.5	95.7 ± 6.2	88.0 ± 11.8	60.0
P3*	89.0 ± 6.3	88.9 ± 8.7	89.1 ± 10.2	60.3
P4*	91.7 ± 6.0	95.1 ± 6.8	88.2 ± 10.9	60.3
P5*	80.3 ± 9.6	81.3 ± 13.2	79.3 ± 12.9	60.3
P6*	81.6 ± 8.2	86.0 ± 13.2	77.3 ± 13.3	60.3
P7*	91.2 ± 6.0	94.6 ± 7.0	87.7 ± 11.8	60.1
P8*	73.8 ± 9.1	80.2 ± 11.5	67.3 ± 17.4	60.3
P9*	76.9 ± 11.5	80.2 ± 13.8	73.4 ± 15.7	60.2
P10*	81.9 ± 8.8	84.5 ± 11.5	79.3 ± 15.3	61.0
P11*	77.8 ± 9.7	79.6 ± 12.8	76.0 ± 14.8	60.3
P12*	79.9 ± 9.9	81.6 ± 13.0	78.4 ± 15.3	60.2
P13*	95.4 ± 4.4	98.7 ± 3.6	92.2 ± 7.5	60.1
P14*	85.1 ± 6.3	95.1 ± 6.9	75.2 ± 12.0	60.0
P15*	90.2 ± 5.6	94.4 ± 6.3	86.1 ± 9.8	60.1
Average	84.9 ± 6.5	88.4 ± 6.8	81.5 ± 7.1	

Table 1: Single-trial classification results of the tracking condition (Tc vs Te) in terms of accuracy, true negative rate (TNR) and true positive rate (TPR) for every participant (mean ± std) and their average. The Tc class was defined as the negative class and the Te class was defined as the positive class. The significance-level (SL) indicates the the upper-limit of the 95% confidence interval of the theoretical chance-level. Participants whose average measures were above significance level were marked with '\*'. \*

condition were very similar to each other, posing challenges to a reliable classification.

#### 4.2.1 Tracking condition

Table 1 displays the single trial classification results of Tc vs Te trials in terms of accuracy, TNR and TPR. The average accuracy obtained was 84.9%. The average TNR obtained was 88.4%. The average TPR obtained was 81.5%. The obtained results are significantly above chance-level ( $\alpha = 0.05$ ).

#### 4.2.2 Feedback condition

Table 2 displays the single trial classification results of Fc vs Fe trials in terms of accuracy, TNR and TPR. The average accuracy obtained was 59.3%. The average TNR obtained was 59.0%. The average TPR obtained was 59.6%. The average results are not significantly above chance-level ( $\alpha = 0.05$ ).

## 5 DISCUSSION

In this experiment we analyzed the electrophysiological signature during three possible system errors (tracking loss, feedback error and background anomaly) in a virtual reality task, and in particular the appearance of error-related potentials on the EEG data. Taken together, the results presented above suggest that it is possible to accurately classify "tracking error" vs "correct" conditions in single trial, and that it seems more difficult to detect "feedback error" and "background anomaly" conditions.

The first observation that it was possible to make regarding the "tracking error" errors is that participants could easily notice them. As when they happened, participants were no longer able to achieve the pick-and-place task. Thus, this kind of error jeopardized the task and generated ErrPs. Besides, the pattern of the ErrP obtained in our study is very similar and well consistent with state-of-the-art literature regarding error signals in 2D paradigms [25, 26, 34].

The results of the analysis of the "feedback error" condition does not seem to show a clear ErrP pattern. Both correct and error trials were associated with an evoked potential. The erroneous condition displayed a small negative peak, that was not present in the correct condition, and displayed a positive peak with slightly higher

Participants	Accuracy (%)	TNR (%)	TPR (%)	SL (%)
P1	56.7 ± 12.2	54.0 ± 16.3	59.3 ± 19.0	60.3
P2*	66.1 ± 9.3	65.2 ± 16.0	67.2 ± 12.6	60.2
P3*	68.1 ± 9.3	68.0 ± 14.1	68.2 ± 13.7	60.3
P4	54.4 ± 8.3	57.2 ± 16.3	51.7 ± 15.7	60.2
P5	60.5 ± 10.0	60.1 ± 16.0	61.2 ± 16.0	60.2
P6	62.8 ± 9.7	59.6 ± 17.4	66.2 ± 17.4	60.3
P7	54.5 ± 10.6	58.0 ± 13.2	51.0 ± 16.2	60.4
P8*	62.7 ± 9.9	63.1 ± 15.1	62.2 ± 17.1	60.3
P9	58.3 ± 10.6	58.3 ± 15.4	58.4 ± 18.7	60.6
P10*	64.2 ± 10.6	63.8 ± 15.6	64.8 ± 16.1	61.0
P11	48.1 ± 11.2	47.5 ± 17.8	48.8 ± 14.7	60.2
P12*	66.4 ± 9.5	60.3 ± 14.7	72.5 ± 15.5	60.2
P13	57.2 ± 10.5	57.4 ± 16.6	57.1 ± 17.7	60.2
P14	56.7 ± 10.8	61.3 ± 16.4	52.2 ± 18.8	60.2
P15	51.8 ± 9.3	51.1 ± 13.1	52.2 ± 18.9	60.2
Average	59.3 ± 5.8	59.0 ± 5.3	59.6 ± 7.3	

Table 2: Single-trial classification results of the feedback condition (Fc vs Fe) in terms of accuracy, true negative rate (TNR) and true positive rate (TPR) for every participant (mean ± std) and their average. The Fc class was defined as the negative class and the Fe class was defined as the positive class. The significance-level (SL) indicates the the upper-limit of the 95% confidence interval of the theoretical chance-level. Participants whose average measures were above significance level were marked with '\*'. \*

amplitude than the correct condition. One possible explanation for this result is the fact that some participants stopped paying attention to the color of the given feedback. As the difficulty of the task was almost inexistent, participants did not rely on the feedback to evaluate their own performance. Possibly, some of them moved back to the center of the scene immediately after dropping the object, hence not paying attention to the feedback.

Regarding the single trial classification accuracy of the "feedback error" condition, even though the average results was not significantly above chance level, the individual average accuracy of five (5) participants was significantly above chance level. One hypothesis may be that these participants, possibly paid closer attention to the provided feedback, hence eliciting ErrP patterns. Even though this observation may be encouraging, considering the data at hand, it is difficult to draw any conclusion regarding the detection of feedback error. Further investigations should however be conducted in future work.

Lastly, our results showed that "Background anomalies" did not seem to generate any ErrP. One potential explanation could be that such kind of error had no impact at all on the execution or outcome of the task. The background anomaly could even remain unnoticed by some participants. Additionally, the low number of background anomaly trials can also hinder the results obtained (20 per participant, before artifact rejection). Thus, additional experiments need be conducted to further characterize the reactions to this specific kind of system error.

In addition, a challenge in combining EEG measurements with VR is the fact that body and eyes movements contaminate the signal quality. Indeed, in our experiment, some participants' data was still contaminated with artifacts, even after artifacts rejection. For most participants, the tracking condition did not seem affected by artifacts, as participants were instructed to remain motionless when detecting a tracking error. No particular instructions were given regarding the feedback condition nor the background anomaly condition, making them more sensitive to artifacts. Additionally, the feedback condition was associated with the dropping of the object and the start of a new trial, making it prone to movement and blink artifacts.

Overall, our results confirm previous results about the generation

of ErrPs in Virtual Reality. In our case, tracking errors generated ErrPs, while background anomalies did not generate any ErrPs. However, considering the ecological condition of the experiment and the fact that the participants were free of their movements, chances are significant that artifacts disrupted the ErrPs signals in some conditions, even after the artifact rejection procedure. Nevertheless, our results suggest that tracking errors could be detected in VR/AR using EEG analysis. This could possibly be due to tracking errors having higher impact on a user than feedback errors. In the context of VR/AR applications' assessment or debugging this would enable to automatically detect and grade the errors that a user might experience during a given session.

Moreover, the single-trial classification results for "feedback condition" were, on average, not significantly above chance-level. Even though 30% of the subjects were above. This could be a consequence of the ecological conditions of the study, leading some participants to stop paying attention to the color of the feedback. The single-trial classification accuracy for the "tracking error" condition was 85% and thus significantly above chance-level, for all participants. This means a very good discrimination between classes "tracking error" and "no tracking error". Such high-level of detection performance is a starting point for asynchronous and real-time decoding of error-related potentials in VR/AR, as it was shown possible in 2D paradigms [20, 25, 34].

## 6 CONCLUSION

In this paper we have explored the possibility to detect error-related potentials in virtual reality when the system generates an unexpected error. In particular, we have investigated three types of errors: tracking errors, feedback errors and a background anomalies. The different errors represented a different aspect on the interaction process (task, outcome, context) and had a different impact on the realization of the task. The results indicated that only the errors which had an impact on the task (tracking losses) were able to generate ErrPs. Additionally, we showed that the single trial detection of ErrPs related to tracking errors in VR was possible with a high accuracy. Taken together, these results show the viability of using BCIs to detect high-cost error-related potentials in a VR application which can lead to different use cases. In the future, this information could be used for monitoring purposes, in order to detect application flaws, or in can be used for corrective purposes, in order to automatically correct errors.

## 7 ACKNOWLEDGMENTS

This work was supported by *Region Bretagne, Université Bretagne et Loire* and Horizon 2020 ERC Consolidator Grant 681231 Feel Your Reach.

## REFERENCES

- [1] M. Andujar and J. E. Gilbert. Let's learn!: enhancing user's engagement levels through passive brain-computer interfaces. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp. 703–708. ACM, 2013.
- [2] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller. Single-trial analysis and classification of erp components — a tutorial. *NeuroImage*, 56(2):814 – 825, 2011. Multivariate Decoding and Brain Reading. doi: 10.1016/j.neuroimage.2010.06.048
- [3] L. Bonnet, F. Lotte, and A. Lécuyer. Two brains, one game: design and evaluation of a multiuser bci video game based on motor imagery. *IEEE Transactions on Computational Intelligence and AI in games*, 5(2):185–198, 2013.
- [4] R. Chavarriaga, A. Sobolewski, and J. d. R. Millán. Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in neuroscience*, 8:208, 2014.
- [5] E. Combrisson and K. Jerbi. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250:126 – 136, 2015. Cutting-edge EEG Methods. doi: 10.1016/j.jneumeth.2015.01.010
- [6] A. Delorme, S. Makeig, and T. Sejnowski. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. 01 2001.
- [7] M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke. Effects of crossmodal divided attention on late erp components. ii. error processing in choice reaction tasks. *Electroencephalography and clinical neurophysiology*, 78(6):447–455, 1991.
- [8] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein. Erp components on reaction errors and their functional significance: a tutorial. *Biological psychology*, 51(2-3):87–107, 2000.
- [9] P. W. Ferrez and J. d. R. Millán. Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE transactions on biomedical engineering*, 55(3):923–929, 2008.
- [10] P. W. Ferrez and J. d. R. Millán. Simultaneous real-time detection of motor imagery and error-related potentials for improved bci accuracy. Technical report, Proceedings of the 4th Intl. Brain-Computer Interface Workshop and Training Course (Graz), 2008.
- [11] M. J. Frank, B. S. Worocho, and T. Curran. Error-related negativity predicts reinforcement learning and conflict biases. *Neuron*, 47(4):495–501, 2005.
- [12] W. J. Gehring, B. Goss, M. G. Coles, D. E. Meyer, and E. Donchin. A neural system for error detection and compensation. *Psychological science*, 4(6):385–390, 1993.
- [13] L. Gehrke, S. Akman, P. Lopes, A. Chen, A. K. Singh, H.-T. Chen, C.-T. Lin, and K. Gramann. Detecting visuo-haptic mismatches in virtual reality using the prediction error negativity of event-related brain potentials. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 427:1–427:11. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300657
- [14] G. Hajcak, C. B. Holroyd, J. S. Moser, and R. F. Simons. Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, 42(2):161–170, 2005.
- [15] C. Kothe. Lab streaming layer (lsl). <https://github.com/scn/labstreaminglayer>. Accessed on October, 26:2015, 2014.
- [16] A. Kreiling, C. Neuper, and G. Müller-Putz. Error potential detection during continuous movement of an artificial arm controlled by brain-computer interface. *Medical & biological engineering & computing*, 50:223–30, 03 2012. doi: 10.1007/s11517-011-0858-4
- [17] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.
- [18] A. Llera, V. Gómez, and H. J. Kappen. Adaptive classification on brain-computer interfaces using reinforcement signals. *Neural Computation*, 24(11):2900–2923, 2012.
- [19] A. Llera, M. A. van Gerven, V. Gómez, O. Jensen, and H. J. Kappen. On the use of interaction error potentials for adaptive brain computer interfaces. *Neural Networks*, 24(10):1120–1127, 2011.
- [20] C. Lopes-Dias, A. I. Sburlea, and G. R. Müller-Putz. Masked and unmasked error-related potentials during continuous control and feedback. *Journal of Neural Engineering*, 15(3):036031, apr 2018. doi: 10.1088/1741-2552/aab806
- [21] C. Lopes-Dias, A. I. Sburlea, and G. R. Müller-Putz. Online asynchronous decoding of error-related potentials during the continuous control of a robot. *Scientific Reports*, 9(1):17596, 2019. doi: 10.1038/s41598-019-54109-x
- [22] F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. Lécuyer, and R. Leeb. Combining BCI with Virtual Reality: Towards New Applications and Improved BCI. In Allison, B. Z., Dunne, Stephen, Leeb, Robert, Millán, J. D. R., Nijholt, and Anton, eds., *Towards Practical Brain-Computer Interfaces*. Springer, 2013.
- [23] G. R. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller. Better than random: a closer look on bci results. *International Journal of Bioelectromagnetism*, 2008.
- [24] A. Navarro-Cebrian, R. T. Knight, and A. S. Kayser. Error-monitoring and post-error compensations: dissociation between perceptual failures and motor errors with and without awareness. *Journal of Neuroscience*, 33(30):12375–12383, 2013.

- [25] J. Omedes, I. Iturrate, J. Minguez, and L. Montesano. Analysis and asynchronous detection of gradually unfolding errors during monitoring tasks. *Journal of Neural Engineering*, 12(5):056001, jul 2015. doi: 10.1088/1741-2560/12/5/056001
- [26] J. Omedes, I. Iturrate, and L. Montesano. Brain connectivity in continuous error tasks. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3997–4000, Aug 2014. doi: 10.1109/EMBC.2014.6944500
- [27] G. Padrao, M. Gonzalez-Franco, M. V. Sanchez-Vives, M. Slater, and A. Rodriguez-Fornells. Violating body movement semantics: neural signatures of self-generated and external-generated errors. *Neuroimage*, 124:147–156, 2016.
- [28] R. Pezzetta, V. Nicolardi, E. Tidoni, and S. M. Aglioti. Error, rather than its probability, elicits specific electrocortical signatures: a combined EEG-immersive virtual reality study of action observation. *Journal of neurophysiology*, 2018.
- [29] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134, 2001.
- [30] R. N. Roy, S. Bonnet, S. Charbonnier, and A. Campagne. Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive bci. In *2013 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6607–6610. IEEE, 2013.
- [31] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus. Correcting robot mistakes in real time using EEG signals. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6570–6577, May 2017. doi: 10.1109/ICRA.2017.7989777
- [32] G. Schalk, J. R. Wolpaw, D. J. McFarland, and G. Pfurtscheller. EEG-based communication: presence of an error potential. *Clinical Neurophysiology*, 111(12):2138 – 2144, 2000. doi: 10.1016/S1388-2457(00)00457-0
- [33] M. Slater and A. Steed. A virtual presence counter. *Presence: Teleoperators & Virtual Environments*, 9(5):413–434, 2000.
- [34] M. Spüler and C. Niethammer. Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity. *Frontiers in Human Neuroscience*, 9:155, 2015. doi: 10.3389/fnhum.2015.00155
- [35] J. R. Wessel. Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, 6:88, 2012.
- [36] B. Yazmir and M. Reiner. I act, therefore i err: EEG correlates of success and failure in a virtual throwing game. *International Journal of Psychophysiology*, 122:32–41, 2017.
- [37] B. Yazmir, M. Reiner, H. Pratt, and M. Zacksenhouse. Brain responses to errors during 3d motion in a haptic-visual vr. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pp. 120–130. Springer, 2016.
- [38] T. O. Zander, L. R. Krol, N. P. Birbaumer, and K. Gramann. Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proceedings of the National Academy of Sciences*, 2016. doi: 10.1073/pnas.1605155114
- [39] H. Zhang, R. Chavarriaga, Z. Khalilgardali, L. Gheorghe, I. Iturrate, and J. d R Millán. EEG-based decoding of error-related brain activity in a real-world driving task. *Journal of Neural Engineering*, 12(6):066028, nov 2015. doi: 10.1088/1741-2560/12/6/066028