NeuroImage 197 (2019) 544–556

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Performance monitoring beyond choice tasks: The time course of force execution monitoring investigated by event-related potentials and multivariate pattern analysis

Yohana Siswandari [a,*], Stefan Bode [a,b], Jutta Stahl [a]

[a] Department of Psychology, University of Cologne, Germany
[b] Melbourne School of Psychological Sciences, The University of Melbourne, Australia

A R T I C L E I N F O

Keywords:
EEG
Force monitoring
Multivariate pattern classification analysis (MVPA)
Motor control
Force production
Even-related potential

A B S T R A C T

Accurate force production is an essential motor function which, in most cases, requires continuous performance monitoring. Unlike choice-response tasks with two response alternatives, the accuracy in a force production paradigm is defined as an area between an upper and lower limit on the force continuum. In the present study, we investigated the neural mechanisms underlying force production. We used a force production task in which the participants ($n = 48$) were asked to exert a brief force pulse within a specific force range. This allowed: (1) investigation of action monitoring activity during force execution using response-locked and feedback-locked event-related potential (ERP) components known to be involved in error monitoring; (2) multivariate pattern analysis (MVPA) for ERPs. We found that the different force production ranges (characterised as *too low*, *correct*, and *too high* with respect to the target force range) showed no clear error-specific variations in the ERP components of interest. MVPA, on the other hand, allowed for successful classification, not only between the correct and the incorrect outcome conditions, but also between the two incorrect outcome conditions. This suggests that the classifier identified neural patterns reflecting the force magnitude rather than the *correctness* of a response. Moreover, additional support-vector regression (SVR) analyses showed that single-trial response parameters (i.e. peak force and time-to-peak) could be decoded from the brain activity pattern starting from 140 ms (for peak force) and 270 ms (for time-to-peak) before the response onset. These results indicate that the motor program defined the magnitude and timing of the force pulse before response execution, while the correctness of that response (in relation to the “default force” required) was not yet foreshadowed in neural signals. Finally, this study presents the first evidence of a *post-error force adjustment* mechanism, for which participants produced a higher force in trials after under-producing the required force, and a lower force in trials after over-producing the required force.

1. Introduction

To adaptively respond to different situations in everyday life, humans need not only a system that generates motor actions in response to various circumstances, but also a reliable monitoring system to ensure that an action will be performed as planned. Performance monitoring is a complex cognitive function, which is essential to detect errors immediately to modify inappropriate responses and to improve future behaviour by continuous adaptation; for example, when controlling an aircraft with a particular amount of force on the rudder pedal. Several components of the event-related potential (ERP) have been identified as indicators of different aspects of error processing. The frontocentral *error-(related) negativity* (Ne/ERN) reflects early error processing mechanisms and peaks 0–180 ms after error response onset (see Falkenstein et al., 1990; Scheffers et al., 1996). A component with similar characteristics, the *correct response negativity* (CRN), occurs after correct responses and has been found to index response uncertainty (e.g., Vidal, et al., 2000). The *error positivity* (Pe)—an indicator of an error detection mechanism related to error awareness (Nieuwenhuis et al., 2001; Steinhauser and Yeung, 2010)—peaks approximately 300 ms after an erroneous response with a more central scalp distribution. Around 250 ms after an error feedback onset, the frontocentral *feedback related negativity* (FRN) can be observed. According to the *reinforcement learning theory,* the FRN indicates prediction errors, such as if an outcome was worse than expected (Holroyd and Coles, 2002).

Although our everyday actions usually consist of continuous

* Corresponding author. Differentielle Psychologie und Psychologische Diagnostik, Universität zu Köln, Pohligstr 1, 50969 Köln, Germany.
E-mail addresses: ysiswan1@uni-koeln.de (Y. Siswandari), jutta.stahl@uni-koeln.de (J. Stahl).

https://doi.org/10.1016/j.neuroimage.2019.05.006
Received 24 July 2018; Received in revised form 25 April 2019; Accepted 2 May 2019
Available online 3 May 2019
1053-8119/

movements, action monitoring studies mostly investigated responses with *discrete* response parameters. This means that the responses to the choice alternatives were clearly separated by effectors (e.g., *left* and *right* hand responses in two-choice flanker tasks, for review see Gehring et al., 2012). However, it might be more difficult to monitor a continuous response parameter than binary left-or-right responses (e.g., Armbrecht et al., 2013; Stahl, 2010). De Bruijn et al. (2003) were the first to investigate action monitoring in a performance task with a *continuous* response parameter—the *response force*—in a task where the participants had to produce a specific amount of force by a finger press. According to the *Parallel Force Unit Model* (Ulrich and Wing, 1991), two mechanisms determine the amount of force produced by the muscles: (1) the number of recruited force-producing motor units (so called *force units*; i.e., the more force units are activated, the higher the resultant response force; note, one motor unit consists of the muscle fibres activated by one motor neuron); and (2) the duration of force unit activation (i.e., the longer the force units are activated, the higher the resultant response force). In De Bruijn's study, the participants pressed a key such that the maximum of the produced force (peak force, PF) reached an amount within an individually defined range in one of two force conditions: low target force (e.g., 250–549 cN) and high target force (e.g., 550–850 cN). Thus, all responses with a force above and below these ranges are incorrect, which means there is not just one incorrect outcome but an unlimited number (for the given example, PF < 250 cN or PF > 850 cN), which might explain why it is more difficult to monitor the quality of a continuous response parameter compared to binary left-or-right responses (e.g., Armbrecht et al., 2013; Stahl, 2010).

Interestingly, De Bruijn et al. (2003) showed that for the *correct* responses, higher CRN amplitudes were observed during the *high target force* condition, compared to the *low target force* condition. This indicates that the action monitoring system was also sensitive to the magnitude of force required. Furthermore, the Ne/ERN amplitude was increased only for force selection errors, but not for force execution errors. A *force selection error* means that, for example, in a low force condition a high force was selected (e.g., higher than 549 cN in the above-mentioned *low target force* condition), whereas a *force execution error* occurs when participants aimed for the correct *low force* range but produced an even lower force. Based on these findings, Armbrecht et al. (2013) reasoned that the force selection errors might not reflect the actual errors *per se*, but could be directly linked to controlling different aspects of the specific response dynamics (i.e., preparing the response force, or initiating the timing in response execution). To investigate this possibility, they assessed the effects of two response parameters on the Ne/ERN and CRN independently: (a) *peak response force* (PF) and (b) *Time-To-Peak* (TTP, i.e. the time period between response onset and maximum point of the force pulse). Participants were instructed to produce four types of isometric force pulses (i.e., high PF with a short TTP, high PF with a long TTP, low PF with a short TTP, and low PF with a long TTP). Independent of response accuracy, the amplitudes for both Ne/ERN and CRN were higher in the high target force condition compared to the low target force condition (similar to de Bruijn et al.‘s findings). Interestingly, Armbrecht et al. also found significantly reduced amplitudes for both Ne/ERN and CRN for long TTP compared to short TTP, independent of response accuracy. However, they found no error-related effects (which would correspond to *force execution errors* in Bruijn et al.‘s terminology) for Ne/ERN and CRN. Taken together, these studies provided evidence that Ne/ERN and CRN are affected by continuous response parameters (peak force and TTP) as well as by error-specific variations, but the functional meaning of the Ne/ERN and CRN in these tasks was still not clear (de Bruijn et al., 2003; Armbrecht et al., 2013).

Notably, neither De Bruijn et al. (2003) nor Armbrecht et al. (2013) found clear neural evidence of error processing during *force execution*. However, a previous study on pianists, who are highly trained in force execution, showed that expertise is linked to more precise response dynamics of finger presses (Parlitz et al., 1998). The results suggest that the action monitoring system is indeed capable of tracking response force during execution. Taking these results into consideration, the absence of effects during *force execution* stage in De Bruijn et al. (2003) and Armbrecht et al. (2013) studies could be due to the lack of substantial training for the participants. In the present experiment, we therefore simplified the force production task and reduced the complexity that might have been present in previous tasks to the abilities of "normal" participants, who are 'only' trained in several training blocks prior to the experiments. By reducing this complexity to only one target force range, we hoped that even "non-experts" would be able to track their force production more easily during execution. Another aspect that might have shaped previous studies' results was the absence of external feedback – in this case, visual feedback from one's own performance might not be enough to establish clear guidelines for successful performance monitoring. Introducing feedback in our study was therefore aimed to further facilitate the application of monitoring during force execution as participants could form clearer representations of the target force range in memory.

1.1. Objective of the present study

In the current study, we focused on investigating action monitoring during *force execution.* The general idea was to learn more about response dynamics in force production, and to shed light on when relevant information related to the monitoring processes (i.e., error specific information, response force parameters) becomes available and was reflected in brain activity. Therefore, we use two approaches in our investigation: (1) classical ERP components, as well as (2) a chronometric, multivariate approach (e.g., Bode and Stahl, 2014). The advantage of using this latter method, in addition to the classical univariate approach, is that it allows the onset of the availability of specific information regarding response parameters (here: correctness, force magnitude, TTP) in distributed patterns of brain activity to be identified. It can also be used to predict single-trial response parameters from brain activity patterns using multivariate regression. This multivariate approach has been used successfully in detecting movement intentions (e.g., Jochumsen et al., 2013; Jochumsen et al., 2016), classifying different movements when varying force and speed (Jochumsen et al., 2013), as well as detecting errors before an overt response in a digit-flanker task (Bode and Stahl, 2014).

To focus on *force execution,* we modified the force production paradigms from previous studies (see De Bruijn et al., 2003; Armbrecht et al., 2013) by using just *one* target force range (with two possible *force execution errors*: responses with a force above the *target force* range, here referred to as the *"too high condition"* and below the *target force* range, referred to as the *"too low condition*") to fully eliminate *force selection errors* and to further eliminate the need of substantial training of different target forces. Participants were further trained to produce only one specific brief force pulse with a short TTP in each trial to minimize possible confounds which might be introduced by further TTP variations. Hence, participants could not confuse the required target range or target pulse length. Force parameters could be defined in terms of PF and TTP in each trial, while errors could solely be defined as missing the crucial target range.

This modified paradigm allowed us to link the classical ERP components to aspects of response dynamics. First, we tested the following contrasting hypotheses for Ne/ERN and CRN: if the Ne/ERN solely reflects error monitoring but not force magnitude monitoring, we would expect no Ne/ERN difference between the two error conditions (*too low*, *too high*). For both error types, however, the Ne/ERNs should be higher than the CRN. In case the observed components are sensitive to the magnitude of force but not to the quality/correctness (as suggested by Armbrecht et al., 2013), the amplitude should be scaled with condition and therefore smallest when a force lower than the target force range is produced, medium when the correct force range is reached, and highest when the target force range is exceeded. In case the components reflect a mixture of both error processing and force monitoring, clear differences between the conditions might be difficult to identify.

Second, we investigated the Pe component, which has been neglected in recent force monitoring studies, but also constitutes an interesting indicator of behavioural adaptation (Overbeek et al., 2005), error awareness (Nieuwenhuis et al., 2001) and error evidence accumulation (Steinhauser and Yeung, 2012). The Pe usually peaks at 300 ms after response onset, which is after the peak force usually reached, making the Pe a potentially highly interesting indicator for force-related error detection. Usually, Pe amplitudes correlated with behavioural indicators of post-error adaptation such as the *post-error slowing* (e.g., Nieuwenhuis et al., 2001). In our force production task, post-error slowing should not be relevant, because response speed did not matter, but we expected a *post-error force adjustment*. Therefore, we expected a higher force in trials following an erroneous response in the *too low condition* and a smaller force in trials following a *too high condition* as it is the more appropriate behavioural adaptation. Given the similarity in post-error adjustments between response slowing and force adaptation, we reasoned that investigating the Pe might provide us with useful information related to our newly developed behavioural adaptation indicator: the post-error force adjustment.

The third ERP component assessed in this study was the FRN, which was expected to follow the feedback in each trial. as this component reflects externally-induced error processing. However, according to the *first-indicator hypothesis* (Holroyd and Coles, 2002), in case an error is already detected at the time of responding, such feedback should not provide additional information, and error-related effects should only be reflected in the Ne/ERN but not in the FRN. However, for errors which went unnoticed at the time of responding, the feedback would be the first error indicator, which means that the FRN but not the Ne/ERN should reflect error-related information (Stahl, 2010).

Our second aim, which goes beyond the research questions derived from Armbrecht et al. (2013), was to better understand how *force execution* process unfolds. For this, we used a chronometric multivariate approach (MVPA; see Bode et al., 2012; Bode and Stahl, 2014), which takes into account spatial and temporal aspects of force production. We chose to use this approach since this method was able to predict upcoming errors (i.e. the use of the incorrect hand) in the spatially distributed pattern of ERPs up to 90 ms before an overt response in a two-choice flanker task (Bode and Stahl, 2014). Our goal was to investigate whether response-related information (correctness, error types, response force and TTP) could be predicted from the spatially distributed ERP signals in small, consecutive time-windows while the response was prepared and subsequently executed. This technique allows to detect *when* (before and/or after the response) any information that directly predicts the response dynamics of single trial response parameters (PF, TTP) becomes available, as well as *whether* the target force range will be missed (errors vs. correct responses), or, in case of an error, force will be under- or overproduced. These analyses might, beyond clarifying the role of classical ERP components in dynamic action monitoring, help explain how response force is produced, i.e. how early specific response information is set up and locked in, which would be essential for early error processing in our task.

2. Method

2.1. Participants

Seventy-eight participants (46 females), all students of the University of Cologne, were recruited and tested in this study. Out of our original datasets, 28 participants were excluded due to displaying an insufficient number of errors (less than 10 trials per condition), and a further two participants were excluded due to technical artefacts. Data from the remaining 48 participants, 33 females and 15 males (age range: 17–44 years; mean ± SD: 25.47 ± 5.44 years), were used for all analyses reported here. All participants were right handed and had normal or corrected-to-normal vision. The experiment was approved by the ethics committee of the German Psychological Association. Informed consent was obtained from all participants, and they were rewarded with course credit for participation.

2.2. Procedure

2.2.1. Apparatus

The behavioural data were recorded using a custom-made force-sensitive key. The response key was mounted on a board that provided full forearm support. To respond, participants briefly pressed the key with their right index finger. The response force was measured by a strain gauge attached on the fixed end of a leaf spring (110 × 19 × 2 mm), which was held by an adjustable clamp at one end of the key, leaving the other end free for the participant to press. When participants pressed the free end of the key, an analogue electrical signal, which corresponds to the exerted force, was produced. This signal was recorded with a sampling rate of 500 Hz. A chin rest (with an adjustable height) was used to maintain a constant and stable posture while keeping a distance of 56 cm from the monitor.

2.2.2. Maximum voluntary force

Before the experiment, we assessed each participant's *maximum voluntary force* (MVF) to determine the individual force ranges (defined in % MVF; see below) as MVF varies across participants. The participants were instructed to press the force key with their right index fingers as hard as possible without moving the forearm, and repeated this procedure ten times in a row. A start signal ("+++") initiated each of the ten key presses. The individual MVF was calculated by averaging the PF of the last seven key presses (e.g., Armbrecht et al., 2013). We then defined five force ranges relative to the individuals' MVF (*target range*: 46–54% MVF; *too high*: > 60% MVF; *slightly too high*: 54–60% MVF, *slightly too low*: 40–46% MVF; *too low* < 40% MVF).

2.2.3. Experimental task

Participants were tested individually in a 50-min experimental session. The experiment comprised six blocks of the force task, each consisting of 44 trials. Each trial began with a presentation of a white square on a black background (see Fig. 1A) for 400 ms in the middle of the screen, serving as a start signal. Participants were then required to produce a brisk, isometric force pulse with their right index finger. We trained the participants to produce a brief force pulse during the first experimental block to minimize TTP variability. This was achieved by using time feedback (the white square turned red if participants did not reach the peak force after 180 ms) in the first experimental block. In all blocks after a response was made, force feedback was presented for 700 ms to indicate whether the correct amount of force was reached or not (see Fig. 1B): The green area in the middle of the force ruler represented the 'correct' area. The upper red area represented the *too high* condition, while the bottom red area represented the *too low* condition. The two yellow areas indicated '*slightly too high*' and '*slightly too low*' conditions (these were not included as separate conditions in the final analyses, see below). A white cursor (see Fig. 1B) indicated the force level that was produced in the current trial. The feedback presentation was followed by a black screen (randomly jittered inter-trial duration: 300–600 ms).

2.3. Data acquisition

2.3.1. Behavioural data

Response time (RT) was defined as the first time point when the participant's response force exceeded 50 cN (i.e. response onset), measured from the stimulus onset. *PF* was defined as the maximum of the force in cN of a single trial force pulse. TTP was defined as the time point (measured from response onset) at which the peak force was reached. Frequency of correct responses, frequency of *too high* responses, frequency of *too low* responses, mean RT, mean PF, and mean TTP were determined separately for each condition and each participant.

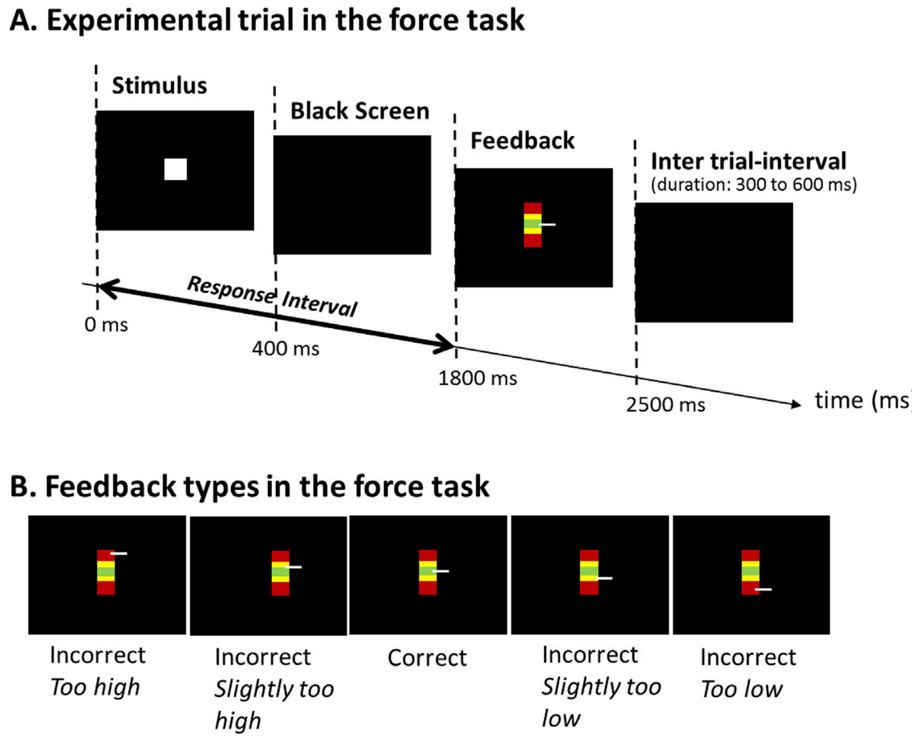

Fig. 1. (A) Time course of an experimental trial and (B) the feedback types in the force task. The response was indicated by a key press; feedback was presented according to the force produced in the current trial.

In order to investigate behaviour adaptation after force error commission, we calculated *post-error force adjustment* ($\Delta PF_{post(error)}$). Analogous to the standard procedure for post-error slowing (Nieuwenhuis et al., 2001), we calculated the difference between the PF of an error trial *n* (separately for, *too high* and *too low condition*) and the PF of the trial following an error trial (*n*+1). To account for the general difference between two subsequent trials, we also calculated the difference between a response following a correct response ($\Delta PF_{post(correct)}$) in the same way as we did with the error response:

$$\Delta PF_{post(f)}\ (n) = PF_f\ (n+1) - PF_f\ (n)$$

Separately for the different *force ranges* (*f*; too high, too low, and correct) and trial *n* (number of trial). The difference scores were then averaged across trials for each participant separately for the three force ranges.

2.3.2. Electrophysiological data

EEG data were recorded from 61 scalp electrode sites (FP1, FP2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C3′, C1, Cz, C2, C4, C4’, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2) according to the standard international 10–20 system (Jasper, 1958). The active Ag/AgCl electrodes (actiCAP, Brain Products) were referenced against the left mastoid. The vertical electrooculogram (EOG) was recorded from an electrode infraorbital to the left eye and the horizontal EOG was recorded at the outer canthi of each eye. The EEG was continuously recorded at a sampling rate of 500 Hz using BrainAmp DC (Brain Products).

The electrophysiological data were time-locked to (a) the response onset; (b) the *Time-To-Peak*; and (c) the feedback onset. In the case of *response-locked* and *feedback-locked* analyses, data were epoched ranging from 100 ms before until 600 ms after response (or feedback) onset. For the *Time-To-Peak-locked* analyses, data were epoched ranging from 300 ms before until 300 ms after peak force was reached. Baseline correction was performed with the period of 100 ms before response/feedback onset for the *response-locked* and *feedback-locked analyses,* and with the period of 100 ms before stimulus onset for *TTP-locked analyses*. An ocular correction algorithm was applied in order to reduce the impact of eye movements (Gratton et al., 1983), followed by a second baseline correction. Afterwards, an artefact-rejection procedure was carried out to eliminate contaminated trials exceeding maximum/minimum amplitudes of ±100 μV. The remaining trials were averaged. A current source density (CSD) analysis was then performed on the averaged ERP waveforms. This analysis accounted for the curvature of the head using a spline algorithm (Perrin et al., 1989), and was performed to reduce the effect of neighbouring currents. The CSD signals (order of splines = 4; lambda = 10^{-5}; maximal degree of Legendre polynomials = 10) were computed for each electrode site by taking the second derivative of the distribution of the voltage over the scalp. Lastly, the ERP components (Ne/ERN, CRN, Pe/Pc for correct trials and FRN; both peak amplitudes and mean amplitudes as the standard indicator for the area under the curve in the defined time ranges – hereafter referred to as ‘area’) were determined separately from the individual mean CSD-ERP waveforms at the electrode sites FCz and Cz. Ne/ERN (CRN) was defined as the most negative peak in a time window ranging from 0 to 180 ms after response onset at electrode site FCz. Pe (Pc) was defined as the most positive peak in a time window ranging from 150 to 300 ms after response onset at electrode site Cz. Finally, the FRN was defined as the most negative peak in a time window ranging from 150 to 250 ms after feedback onset at electrode sites Cz and FCz. The selected time windows and electrode sites for each component were in line with previous studies (e.g., De Bruijn et al., 2003; Gehring et al., 1993; Falkenstein et al., 2000) and confirmed by visual inspection of the waveforms and topography maps (see Supplementary Fig. 7).

2.4. Univariate statistical analyses

Separate repeated-measures analysis of variance (ANOVA) were conducted for the within-subject factor *force range* (*correct, too high, too low*) for all behavioural measures (RT, PF, TTP, ΔRT_{post}, ΔPF_{post}) using

SPSS 23. Due to very low number of error trials (less than 6 trials), we could not use the *slightly too high* condition and *slightly too low* condition for the separate analyses (but the trials were included for the regression approach of MVPA, see below). Further, separate ANOVAs with the within-subject factor *force range* were performed for the respective ERP components' peak amplitudes (Ne/ERN, CRN, Pc, Pe) and the components' area measures. A two-way ANOVA with the within-subject factors *force range* (*correct, too high, too low*) and *electrode site* (FCz and Cz) was performed for the FRN peak amplitude. The electrode factor was used here as the source of the component was less clear compared to the other components. Significant ANOVA results were followed up using Bonferroni adjusted post-hoc tests. Level of significance were adjusted using Geisser and Greenhouse (1958) correction in case the sphericity assumption was violated. Effect sizes are reported in terms of partial eta^2 (η_p^2).

2.5. Multivariate pattern classification analysis

Multivariate pattern classification analysis (MVPA) was used to find the earliest time point after stimulus presentation that allowed for decoding the response outcome from distributed spatio-temporal patterns of ERPs. We applied two types of MVPA by using the Decision Decoding ToolBOX (Bode et al., 2019). For the first set of analyses, we used support *vector machine classification* (SVC) in a moving analysis window approach to decode the following conditions in the force task: (1) *correct* vs. *too high* condition; (2) *correct* vs. *too low* condition; (3) *too high* vs. *too low* condition. Additionally, we aggregated *too high* and *too low* conditions to one condition which we refer to as the *incorrect* condition, and conducted an additional SVC analysis for *incorrect* vs. *correct*. As the number of trials in the *too high* and *too low* condition was not equal (with the ratio of 2:3), we balanced the number of trials for these two conditions for both the training and test data sets of the *incorrect* condition. In all cases, the moving analysis windows were applied to the brain activity pattern before and after the response onset. For the second set of analyses – aimed at identifying the first time point at which the response parameters (PF, TTP) were decodable – we used support *vector regression* (SVR), which allows the decoding of continuous outcome variables.

2.5.1. Classification analyses

For the first set of classification analyses using SVC, each participant's artefact-cleaned data was sorted into three groups with respect to the force produced: (1) *correct*; (2) *too high*; (3) *too low*. For the additional *correct* vs. *incorrect* analysis, group 2 (*too high*) and group 3 (*too low*) were combined into one group (*incorrect* condition). The EEG epochs (for all groups) were time-locked to the response onset and (additionally) the time point when the peak force was reached (TTP in figure). Then, we conducted six classification analyses, analysing all three pairwise classification analyses for both response-locked and TTP-locked data, respectively. For the main classification analyses (response-locked), we included data starting from 150 ms before response onset since Bode and Stahl (2014) were able to predict whether responses would be erroneous or correct from about 100 ms before response onset. As the task was to reach a specific maximum force, and the TTP point marked the end point of the force production phase, it was of interest to decode information in neural signals leading up to this time point, mirroring the classical ERP analyses. For these classification analyses, data starting from 300 ms before the peak force and after the peak force was reached were included.

For each individual pattern classification analysis (*correct* vs. *too high*; *correct* vs. *too low*; *too low* vs. *too high; correct* vs. *incorrect*) the following steps were performed. A non-overlapping *spatio-temporal* analysis time-window of 10 ms that contained 5 data points for each of all 61 channels was used, covering the entire epoch (cf., Bode and Stahl, 2014). For each trial, all data points included in this window were transformed into vectors, to represent the *spatio-temporal* patterns associated with each condition. These were then randomly assigned to ten separate sets. The linear classifier (using the default regularisation parameter C = 1) was trained on vectors from the two conditions of interest by interfacing with the LIBSVM toolbox (Chang and Lin, 2011). Importantly, only 90% of the data (i.e., 9 of the 10 sets) were randomly drawn and were used to train the classifier. Based on these exemplars, the classifier estimated a decision boundary that optimally separated exemplars from the two classes (i.e. categories). The vectors from the remaining independent 10% of the data were subsequently used for testing the classifier (Bode et al., 2019). Note that in order to avoid biased results because of an imbalanced trial numbers (i.e., samples), we balanced the trial numbers for conditions before dividing the data into training and test sets. We used only the smaller number of trials of the two conditions, and the trials of the condition with larger trial numbers were randomly drawn to match this number. This means, there was an equal number of trials in both the training set and the test set, and the number of exemplars was always the same between conditions.

The percentage of correct classifications (decoding accuracy) is based on the estimated decision boundary served as the outcome measure. Above chance classification indicates that the data from the respective analysis time window contained information about the two conditions, while chance-level classification (50% with two classes) suggests no evidence for such information. In order to minimize the risk of false positive results when determining the decoding accuracy, the classification process was first repeated using a 10-fold cross-validation procedure in which each set containing 10% of the data served as *test data* once while the classifier was re-trained on the remaining 90% of the data, until all of the sets were independently used for testing once. In addition, to avoid potential drawing biases, all ten cross-validation steps were then repeated ten times in an identical fashion, but with newly randomly drawn 10 sets for training and testing, resulting in a total of 100 analyses. The average of all of these analysis steps constituted the estimate of the individual classification accuracy for the respective analysis time window.

Finally, statistical testing was performed at group-level. For this, we conducted a series of Bonferroni-corrected t-tests for each (10 ms) analysis time window, testing the performance of the classifier against empirical chance results from the near-identical shuffled-labels analysis (Bode et al., 2019), in which the same data and the same labels were used in the same number of cross-classification steps (and iterations thereof), but the assignment of labels to data was randomly assigned for each step. This means, both the real and the empirical chance distributions were composed of the average accuracies (or Fisher-Z transformed correlation coefficients for SVR) across 100 analyses per participant (10×10 cross-validation steps). The decision to use the empirical chance distributions was made because it provided a stricter test than testing against the theoretical chance level of 50% (Bode et al., 2012; Bode and Stahl, 2014).

2.5.2. Support vector regression (SVR)

For this analysis we included all artefact-cleaned data without further grouping into any condition. For SVR, a slightly longer interval of the same EEG epochs (response-locked, −400 ms–300 ms related to response onset) and the same set of parameters for the *spatio-temporal* analysis (non-overlapping analysis time windows of 10 ms that included 5 data points from each of the 61 channels) were used as for the classification analysis. The decision to start the analyses at 400 ms before response onset for SVR was made because if evidence about erroneous response was present in the brain activity pattern already around 100 ms before response onset (Bode and Stahl, 2014), evidence regarding response related parameters (PF and TTP) might exist in the brain activity pattern even earlier. Training the model was again conducted on vectors from all data using LIBSVM (we used the default parameters epsilon-SVR = 3 and C = 1). To estimate the regression model, data from all trials (including the *slightly high force/slightly low force conditions*) were again randomly divided into 10 equally-sized sets. Out of these, 9 (90%) sets were randomly drawn and used to train the model, while the left-out one (10%) was used for testing. As before, a 10-fold-cross validation procedure was then applied for which each data set was used for testing once and the other 9 were used for training. The same strict 10-times repetition of the entire cross-validated analysis was conducted as described above. The only difference to the classification analysis was that the results of SVR are individual correlation coefficients between the *predicted* variable (PF or TTP) based on the regression model and their *true* values, averaged across all iterations for each given time window. Finally, we conducted Fisher-Z transformation for the correlation coefficients, and the final measure was one average coefficient per participant per analysis time window, reflecting information regarding the condition of interest (e.g., Bode et al., 2014). In other words, a significant coefficient means that information regarding PF (or TTP) was represented in brain activity in the respective analysis time window.

As before, the same analysis was then conducted with randomly shuffled labels (i.e., PF or TTP values) for each participant and each analysis time window to obtain an empirical distribution of regression results under the null hypothesis for each analysis step. Group level Bonferroni-corrected t-tests were again used to compare the real empirical results with the shuffled label SVR results for each time window.

3. Results

3.1. Behavioural data

First, we investigated whether responses for the different peak force ranges classified as *too high*, *correct* and *too low* also differed with respect to TTP. The respective ANOVA showed that Force Range had a significant effect, $F(2,47) = 22.79$, $p < 0.001$, $\eta_p^2 = 0.245$. The longest TTP was observed in the *too high* condition (252.74 ± 13.42 ms), followed by the *correct* condition (245.38 ± 15.56 ms) and the *too low* condition (212.95 ± 12.53 ms). Follow-up post-hoc tests confirmed significant differences between the *correct* and the *too low* condition ($p < 0.001$), as well as between the *too high* condition and the *too low* condition ($p < 0.001$), but not for the *too high* condition and the *correct* condition ($p = 0.999$). No significant Force Range effect was observed for RT, $F(2,47) = 0.325$, $p = 0.723$. We additionally investigated the distribution of response types and found that the correct responses were most frequent ($46.44 \pm 1.39\%$), followed by the *too low* condition ($18.82 \pm 1.30\%$) and the *too high* condition ($12.81 \pm 0.82\%$).

The ANOVA on post-response force adjustment (ΔPF_{post}) showed a significant effect of Force Range, $F(2,47) = 1513.54$, $p < 0.001$, $\eta_p^2 = 0.970$. Following trials in which insufficient force was produced, participants produced on average more force (621 ± 21.55 cN), whereas after producing too much force they produced significantly less force on average (-788 ± 24.71 cN). After a correct force trial, participants responded on average with a similar amount of force (-22.7 ± 12.03 cN; for all comparisons, $p < 0.001$). However, the equivalent ANOVA for RT showed that no post-response slowing (ΔRT_{post}) was observed after correct and incorrect responses, $F(2,47) = 1.59$, $p = 0.208$, $\eta_p{}^2 = 0.033$.

3.2. Electrophysiological data

3.2.1. Response-locked and TTP-locked averages

A series of ANOVAs was conducted, each testing the effect of the within-subjects factor Force Range (conditions: *too low, correct, too high*) on the peak amplitude in the time window of interest (specific for each ERP component) at the respective channels, as described in the method section. Identical analyses for the same components were conducted with data time-locked to the response and to the TTP, respectively. The Force Range did neither affect the response-locked Ne/ERN peak amplitude (CRN for the *correct* condition), $F(2,47) = 3.08$, $p = 0.050$, $\eta_p^2 = 0.062$, nor Ne/ERN area, $F(2,47) = 0.86$, $p = 0.424$ (see also Fig. 2A). The additional TTP-locked analyses for Ne/ERN (see Fig. 2B) showed a similar pattern of results. Neither Ne/ERN peak amplitude, $F(2,47) = 0.59$, $p = 0.554$, $\eta_p^2 = 0.025$, nor Ne/ERN area, $F(2,47) = 1.68$, $p = 0.197$, $\eta_p^2 = 0.068$, showed significant effects.

Significant effects of Force Range were detected for both response-locked Pe peak amplitude, $F(2,47) = 3.77$, $p = 0.030$, $\eta_p^2 = 0.141$, and Pe area, $F(2,47) = 6.98$, $p = 0.002$, $\eta_p^2 = 0.233$, at Cz (Fig. 2C). Post-hoc tests showed significant differences for Pe area between the *correct* condition ($11.29 \pm 1.40\,\mu V/cm^2$) and the *too high* condition ($15.42 \pm 2.26\,\mu V/cm^2$, $p = 0.017$), as well as between the *too low* condition ($9.95 \pm 1.26\,\mu V/cm^2$) and the *too high* condition ($15.42 \pm 2.26\,\mu V/cm^2$, $p \leq 0.001$). We also observed a near-identical pattern of results for the TTP-locked analyses for Pe peak amplitude and areas (see Fig. 2D). Significant effects of Force Range were observed for both Pe peak amplitude, $F(2,47) = 7.521$, $p < 0.001$, $\eta_p^2 = 0.246$, and Pe area, $F(2,47) = 15.569$, $p < 0.001$, $\eta_p^2 = 0.471$, at Cz (Fig. 2D). Post-hoc tests showed significant differences for Pe area between the *correct* condition ($10.36 \pm 1.27\,\mu V/cm^2$) and the *too high* condition ($16.72 \pm 1.85\,\mu V/cm^2$, $p < 0.001$), as well as between the *too low* condition ($9.87 \pm 1.24\,\mu V/cm^2$) and the *too high* condition ($16.72 \pm 1.85\,\mu V/cm^2$, $p \leq 0.001$). Highly similar post-hoc results pattern for Pe peak amplitude were observed for both response-locked and TTP-locked results (see Supplementary Table 1).

3.2.2. Feedback-locked averages

The ANOVAs for FRN peak amplitude (see Fig. 2E and F) revealed a significant effect of Force Range at Cz, $F(2,47) = 6.43$, $p = 0.002$, $\eta_p^2 = 0.120$. Post-hoc comparisons showed that FRN peak amplitude of the *too low* condition ($-5.08 \pm 1.95\,\mu V/cm^2$) was significantly more negative than of the *correct* condition ($0.45 \pm 1.88\,\mu V/cm^2$; $p = 0.002$). No significant differences were observed between the *too high* condition ($-3.05 \pm 1.87\,\mu V/cm^2$) and the other two conditions (*ps* > 0.10). A near-identical pattern of results was observed when the FCz electrode was used to determine the FRN (see Supplementary Table 1). Here, we did not investigate FRN area because for a relative negative component, the area is usually not very informative.

Finally, we tested if any of the peak amplitude measures for these ERP components were correlated with post-error force adjustment (ΔPF_{post}). However, no significant correlations were observed (Ne/ERN: $r < 0.255$; Pe: $r < 0.197$; and FRN: $r < 0.206$; for all *ps* > 0.09).

3.3. Multivariate pattern classification

To investigate *whether* error-specific information is predictable from the distributed ERP signal, and to identify *when* (i.e., time point(s)) this error specific information is decodable from brain activity, we ran the first six sets of multivariate classification analyses. The first three classifications were conducted using response-locked ERP data (see Fig. 3A–C), and the other three were conducted using TTP-locked ERP data (see Fig. 3D–F). For each analysis, each pair of the defined force ranges were used as distinct classes for the classifier, each time using a moving-window approach with 10 ms analysis time windows: *correct* vs. *too high* (Fig. 3A, C), *correct* vs. *too low* Fig. 3B, D), and *too high* vs. *too low* (Fig. 3C, F). We also conducted a *correct* vs. *incorrect* classification analysis (Fig. 4) to investigate whether a more 'general' error-related processes could be reflected in the ERP signal. The univariate equivalent for this *correct* vs. *incorrect* analysis as well as the corresponding statistical results can be seen in the Supplementary Fig. 8 and Supplementary Table 2.

Our main (response-locked) classification results (Fig. 3A–C) showed significant classification accuracies for all three analyses. For all three analyses, we did not find evidence for substantial prediction before response onset (0 ms), but we identified several significant time windows between response onset and the average TTP (ranging between 213 and 253 ms, see behavioural data), indicating that information about the quality of the response outcome was available in brain activity before the actual maximum force was reached. Our additional TTP-locked classification analyses (Fig. 3E and F) confirmed that information regarding the quality of the response outcome was available as early as 80–110 ms before peak force was reached. Interestingly, both

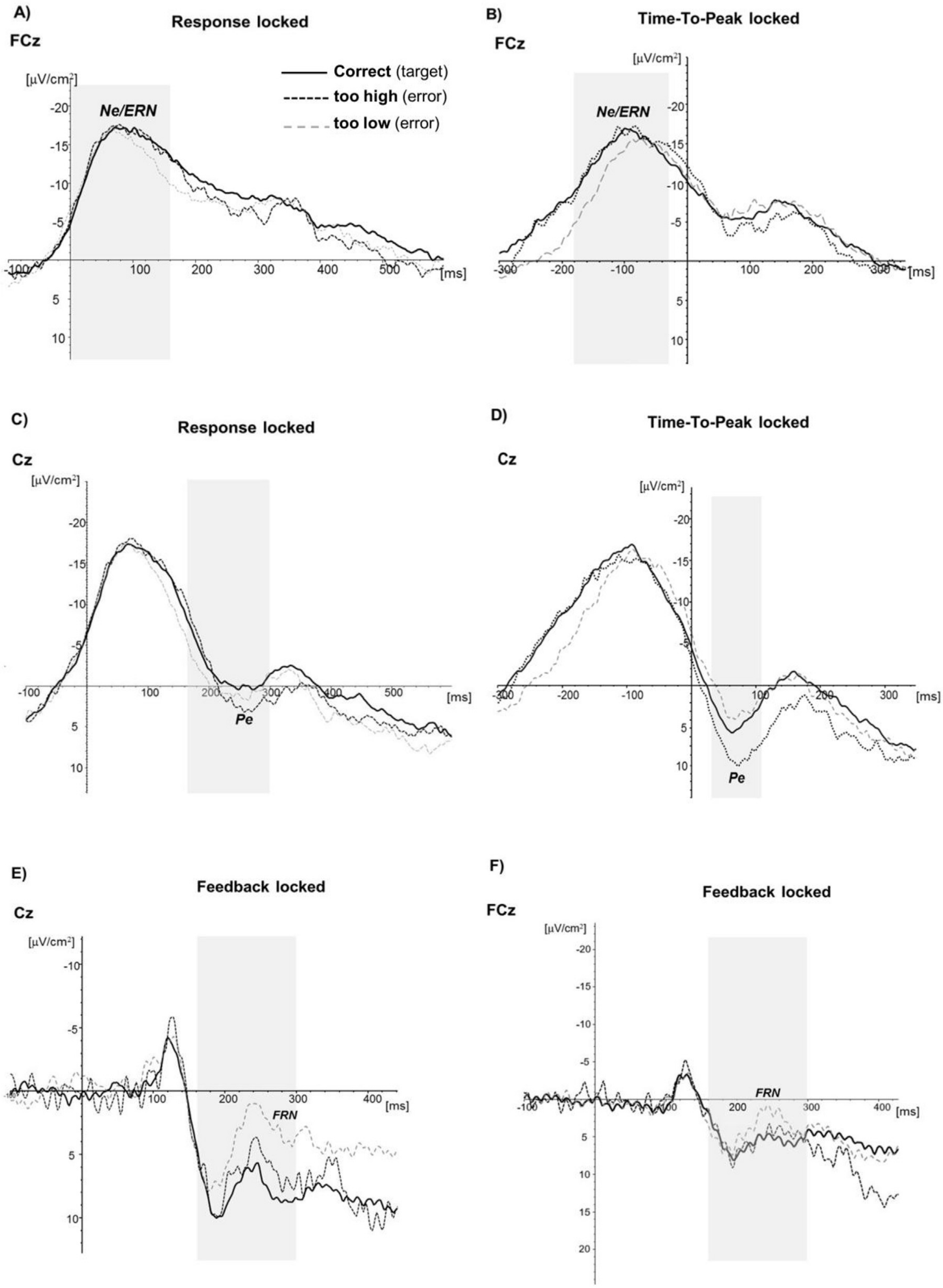

Fig. 2. Averaged event-related potentials of the force task time-locked to the response onset (A, C), time-locked to the Time-To-Peak (B, D), and time-locked to the feedback-onset (E, F) for the two electrode sites of interest FCz (A, B, F) and Cz (C, D, E) for error–related negativity (Ne/ERN), error positivity (Pe) and feedback-related negativity (FRN).

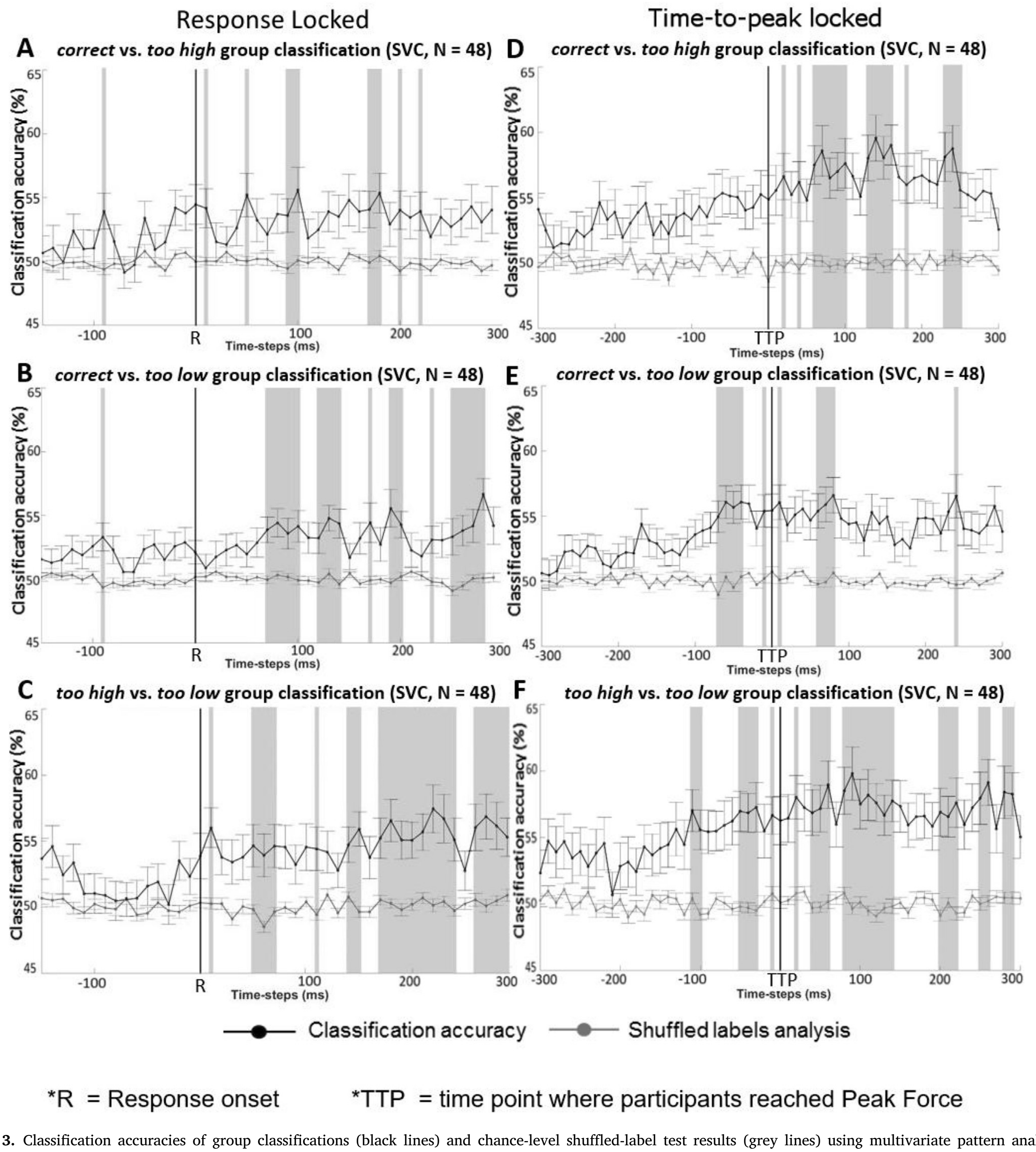

Fig. 3. Classification accuracies of group classifications (black lines) and chance-level shuffled-label test results (grey lines) using multivariate pattern analysis (Support Vector Classification, SVC), time-locked to the Response Onset (A–C) and time-locked to the TTP (D–F), for (A,D) *correct* (target force) vs. *too high* (error) classification, (B,E) *correct* (target force) vs. *too low* (error) classification, (C,F*) too high* (error) vs. *too low* (error) classification; corrected for multiple comparisons ($N = 48$; $p < 0.05$; error bars indicate standard errors of the mean; grey bars indicate significant accuracy in analysis time window).

the response-locked and TTP-locked analyses (Fig. 3C, F) showed that the classifier also identified differences between the two error conditions (*too high* vs. *too low*). Furthermore, the absence of any significant time windows for the *correct* vs. *incorrect* analysis (Fig. 4) suggested that the classifier was not able to pick up information regarding a more general error-related processes from the brain activity. Added together, these results showed that instead of accuracy-related evidence of the response, we presumably decoded information regarding force magnitude.

Since our classification results suggested that the ERP activity patterns might reflect information regarding force magnitude instead of correctness *per se*, we then investigated *when* specific force parameters of the actually executed force could be predicted. To address this point, we used SVR to predict the single-trial PF (single-trial TTP) from the ERP patterns. The trials were not separated by conditions any longer but were included in the same analysis. Our first SVR analysis (Fig. 5A) showed that the PF could be predicted from 140 ms before the response onset up to the point in time of the average TTP. As TTP and PF reflected different

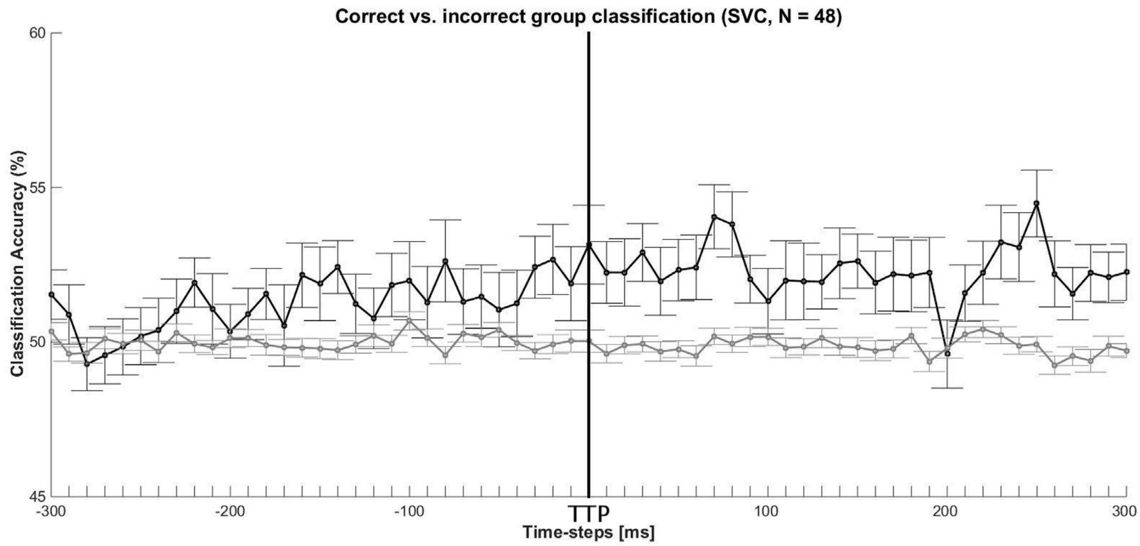

Fig. 4. Classification accuracies of group classification and permutation test result using multivariate pattern analysis (Support Vector Classification, SVC) for *correct* vs. *incorrect (too high & too low)* classification; corrected for multiple comparisons (N = 48; $p < 0.05$; error bars indicate standard errors of the mean; no analysis time window was significant).
* TTP =
time point where Participants reached Peak Force.

aspects of a response (across trials, the mean correlation was small; $r = 0.19$), we further predicted TTP in an independent SVR analysis (Fig. 5B). These findings showed that TTP, which reflects the temporal dynamics of a force pulse, could be predicted starting from around 270 ms before the response up to 300 ms after the response (remarkably, during the entire time interval after response onset), pointing towards the fact that not only the force magnitude but also the timing of the force pulse was processed (prepared) clearly before the response onset and approximately 100 ms earlier compared to PF.

4. Discussion

To date, little is known regarding the processes involved in action monitoring in tasks with continuous response parameters, such as response force. The present study aimed to fill this gap and investigated action monitoring during *force execution* using a force production task with one target force range. We assessed three error-related ERP components, the Ne/ERN/CRN, Pe(Pc), and FRN, which are well-established indicators of different error processing mechanisms. Furthermore, we used a novel multivariate approach (MVPA) to investigate the time course of pre- and post-response force execution monitoring. Although we were not able to identify clear error-specific variations from the ERP components, our MVPA results provide first evidences that response related parameters (TTP, PF) are decodable from the brain activity clearly before the time of response.

4.1. Force execution monitoring

From previous work, it was not clear whether the medial-frontal negative ERP components (i.e., Ne/ERN and CRN) were truly sensitive to errors in force execution because there were some confounding factors in these studies (Armbrecht et al., 2012, 2013; De Bruijn et al., 2003). We attempted to answer this question in the present study by using a *pure* force execution task, therefore eliminating *force selection* processes (and therefore potential confounding effects that comes with having several target ranges to select from). However, we could neither find clear error-specific variation between Ne/ERN and CRN, nor could we clearly replicate the components' sensitivity to the force ranges with the present data, while others have shown that higher force ranges were associated with higher peak amplitudes (Armbrecht et al., 2013; De Bruijn et al., 2003). One reason that might explain these ambiguous findings for Ne/ERN and CRN is variability in TTP, which we did not strictly control, in order to focus the participants on correct peak force production.

Our study was the first to investigate the Pe in a force production task, and we observed clear Pe variations between force ranges. Interestingly, the Pe was observable not only in the two incorrect conditions but also in the correct condition. According to previous studies, the Pe is discussed to reflect conscious evaluation of an error (see Nieuwenhuis et al., 2001) and error evidence accumulation (Steinhauser and Yeung, 2010). For example, in studies investigating left-right hand errors, a clear Pe was observed in error trials but not in correct trials (e.g., Falkenstein et al., 2000; Nieuwenhuis et al., 2001). The existence of a Pc (which is a Pe-like component in the *correct* condition) in our task might be an effect of the participants' uncertainty in distinguishing a correct from an erroneous response, in particular in trials close to the lower limit of the correct force range. The higher Pe in the *too high* condition compared to the *too low* condition could be a result of more error evidence accumulation in a shorter time period (Steinhauser and Yeung, 2012)—hence, more error awareness—after a stronger key press than a weaker key press (in the erroneous conditions).

Interestingly, the FRN amplitude was significantly more negative in the *too low* condition compared to the *correct* condition, but not when comparing the *too low* condition to the *too high* condition, nor when comparing the *correct* condition to the *too high* condition. According to the *first indicator hypothesis* (Holroyd and Coles, 2002; Stahl, 2010), the FRN amplitude should be higher in conditions where the error was not detected at the time of responding. These results therefore appear to be complementary to the Pe results, which showed significant differences only between the *correct* and *too high* conditions, suggesting that these differences were detected immediately, while the other error, the *too low* condition, was not. Hence, feedback was needed to become aware of errors committed in the *too low* force production range, which would lead to enhanced FRN amplitudes for the *too low* condition compared to *correct* condition, but not the *too high* condition. However, the lack of significant difference in the FRN between the two error conditions slightly challenges this interpretation. An alternative theory, which might explain this specific pattern of results is the *expectancy violation hypothesis* (Hajcak et al., 2005). Previous studies have shown that the FRN was larger in

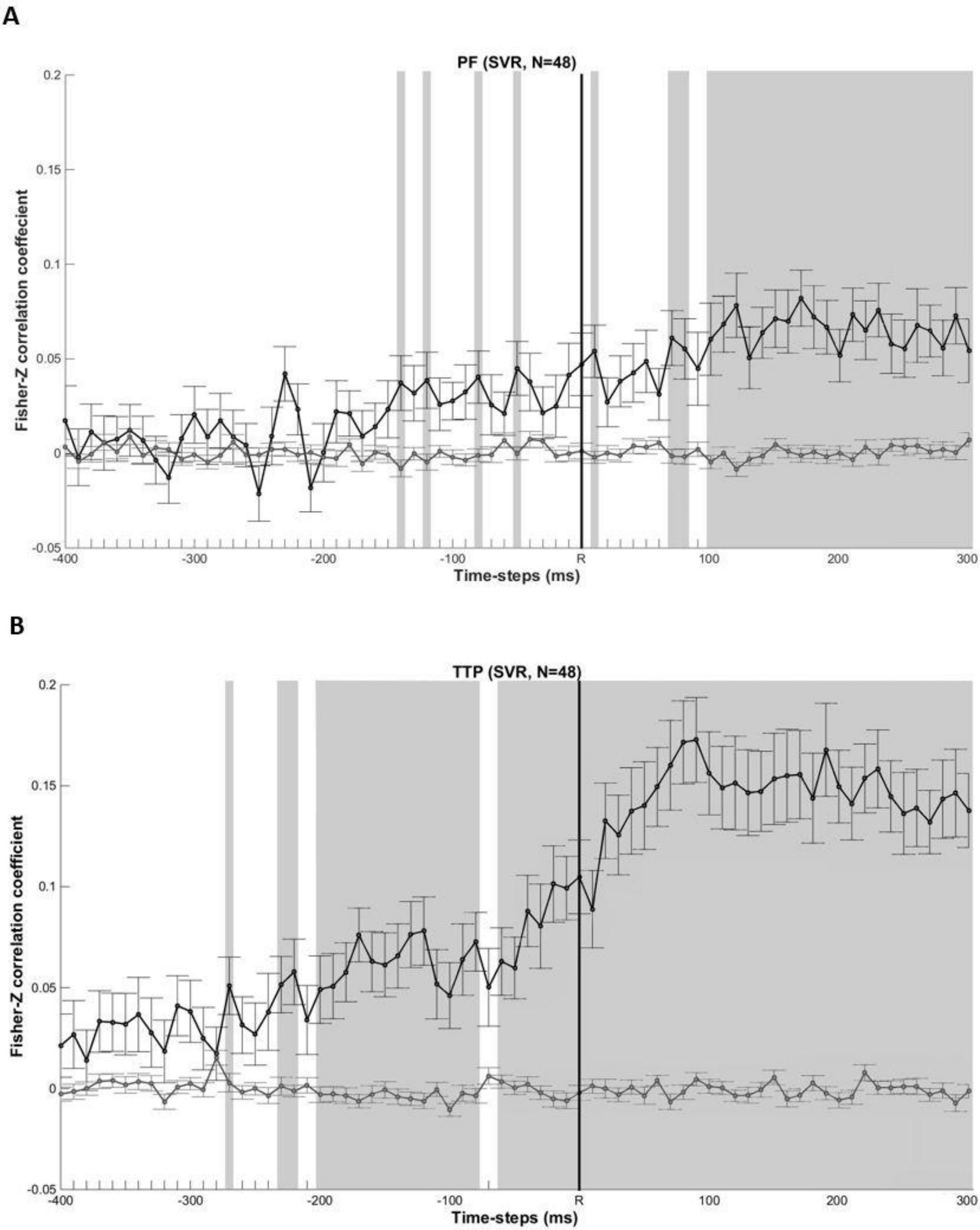

Fig. 5. Fisher Z-correlation coefficients and permutation test results using multivariate pattern analysis (Support vector regression, SVR, A, B) for (A) force prediction (PF), and (B) time to peak prediction (TTP); corrected for multiple comparisons (N = 48; $p < 0.05$; error bars indicate standard errors of the mean; grey time windows indicate significant correlation).
*R =
Response onset.

conditions with a lower probability of a negative outcome, because a negative outcome was less expected (see Holroyd et al., 2009; Hajcak et al., 2007; Holroyd and Krigolson, 2007). If the high Pe in the *too high* condition actually reflected successful error detection, the participants should have already expected getting a feedback that an error was committed. As a consequence, there was no violation of expectation in this condition, and consequently a smaller FRN would be observed compared to the *too low* condition, in which the error feedback was surprising, because the error went undetected, as suggested by the Pe results. In other words, receiving error feedback in the *too low* condition could have represented a violation of expectation, and in turn resulted in a comparably larger FRN.

Taken together, we found no clear error-specific variations for the Ne/ERN, but some evidence indicating force sensitivity in the other ERP components: (1) error evidence accumulation in the *correct* and *too high* conditions was reflected in the Pe (and Pc); and (2) a more negative FRN in the *too low* condition, possibly as a result of the participants' need for external feedback to detect errors and/or violation of expectations.

4.2. Post-error force adjustment

We found a significant PF adjustment after error trials in both error conditions (i.e., *too low* and *too high*). Specifically, participants produced a higher force (approx. 620 cN higher) in trials after under-producing the required force, and a lower force (approx. 790 cN lower) in trials after over-producing the required force. No such systematic force adjustments were observed in trials following correct trials. To the best of our knowledge, this is the first evidence of a *post-error force adjustment* mechanism, which might be aimed at performance improvement depending on the previous error type. In tasks with hand errors (i.e. response selection tasks in which force is irrelevant), usually a positive correlation between post-error slowing and Pe was reported, implicating a role of (conscious) behavioural adaptation reflected by Pe (see Hajcak et al., 2003; Nieuwenhuis et al., 2001). However, in our study, we did not observe a correlation between the Pe (or any of the other ERP components) and the magnitude of force adaptation. Thus, the present results do not provide further evidence regarding the origins of this mechanism and whether it reflected conscious adjustments. It is interesting to note, however, that the absence of a Pe effect for the *low force* range suggests that under-production errors went undetected, but yet a reliable adjustment process was observed. This points to the importance of the external feedback for this process.

4.3. Decoding of force production specific information from ERP signals

Upcoming errors were previously found to be reflected in the ERP up to 90 ms before an overt response (see Bode and Stahl, 2014). Thus, we conducted a further set of analyses using a more sensitive method – MVPA – to see *if* (furthermore, *when*) information regarding any of the response-related parameters, which contribute to the quality of a single trial response (i.e., correctness, error types, single trial PF and TTP) was actually decodable from the ERP signal.

Using classification analyses, we were able to decode correct trials from the two incorrect conditions reliably after response onset. The response-locked and TTP-locked classification results for *correct* (target force) vs. *too low* (error) and *correct* (target force) vs. *too high* (error) could be explained by a variation of force magnitude (i.e., *too low* < *correct* (target force) < *too high*), by error monitoring activity (*correct* vs. *incorrect*), or by a combination of both. However, we could also decode differences between the two incorrect conditions. If the ERP patterns solely reflected whether a response was correct or not, the classifier should not have been able to identify the difference between the two incorrect conditions. This was supported by an additional classification analysis for which we combined the two erroneous conditions (*too low* and *too high*) into one 'incorrect' condition and decoded *correct* vs. *incorrect* force. The result from this analysis confirmed that the classifier could not predict general error-related process from the ERP pattern. Taking into account both findings, we concluded that the classifier mainly decoded force magnitude information instead of merely error specific information. Furthermore, if this was the case, the result from the additional *correct* vs. *incorrect* classification would be expected, because both "incorrect" patterns (for *too high* and *too low* conditions) would be markedly different from each other, and the correct pattern would indeed lie in-between them, diminishing the classifier's ability to reliably distinguish between conditions. It is also noteworthy that the significant classification results were consistently distributed from the time of responding until 300 ms after response onset, and therefore not likely to only mirror differences in the ERP components described above. While on average mostly not significantly different from chance, accuracies appeared to be enhanced in some time windows even before response onset, which could point to some trials in which the final PF was determined very early, which in turn was already reflected in brain activity patterns. The cognitive origins of these patterns are rather diffuse. Based on these analyses alone, however, it remains difficult to clearly link these results to specific cognitive processes. Another important thing to note is, our methods were utilized to estimate a conservative lower bound which we can trust not to be a false positive result. There are various ways in which classifiers optimisation could be achieved (e.g. feature elimination, kernel optimisation, steps reduction in the cross-validation process, modification of the analysis time window width, and many others). We chose to sacrifice absolute accuracy in favour of statistical rigour since absolute accuracy could falsely imply that it is a measure of how much of the process can be predicted from the brain activity. Instead of optimizing accuracy, for our research questions it is more important to make sure that it is truly above chance. Note that most researches in the field adapt a similar approach, with accuracy rarely exceeding 55–60% (e.g., Hogendoorn and Burkitt, 2018, Neuroimage; Fahrenfort et al., 2017, Sci Rep). Thus, we decided to go with techniques that are likely to show lower average accuracies in the end, but allowed us to avoid over interpreting our results and reduced the risk of false positives.

The subsequent SVR analyses showed that the single-trial PF could already be decoded from the EEG pattern at least 140 ms before the response onset. This means that the information about the actual response outcome was clearly available before response initiation in the periphery (i.e. when the minimal force of 50 cN required to be registered as a button press was exceeded). Similarly, a previous study using force sensitive keys has shown that the correctness of simple decisions (i.e., erroneous responses in a flanker task) could be decoded ~100 ms before the overt response was registered (Bode and Stahl, 2014). This previous study, however, did not attempt to predict the actual response force, and differences in force range were not a component of the task either. However, taken together, our results suggest that the brain appears to process information that is indicative of subsequent response errors long before response execution. This must of course not mean that this information is available for immediate response monitoring, as it could simply reflect neural patterns associated with incorrect response planning.

The second finding of the SVR analysis in the present study was that the single-trial TTP was decodable during the entire period after responding even more consistently than PF. This result further pointed out that the temporal dynamics of force production—as reflected by TTP—seem to be affected by processes even earlier than the maximum of the force itself, because it was decodable already 270 ms before response onset. Ulrich et al. (1995) demonstrated that TTP and PF are not fully independent parameters, but they are more than just two sides of the same coin. This is in line with our finding that both parameters were only marginally (and not significantly) correlated. The *Parallel Force Unit Model* (see Ulrich and Wing, 1991) provides a logical explanation for this partial independence: If participants produce force in a specific time, they have to adjust two response parameters, the duration of force unit activation (which is reflected in the TTP) and the number of recruited force units (i.e., force-producing motor units, as reflected in the PF).

Taken together, findings from both SVR analyses (for PF and TTP) provided important evidence that the magnitude and timing of the produced force pulse were planned already before response execution. As a force pulse is a ballistic process (Cordo, 1987; Desmedt, 1982), however, with some controllable aspects, for example, by a modification of the number of the involved force units (Ulrich and Wing, 1991), an early definition of the response parameters in the 'motor program' seems to be an efficient strategy for a fast force production and could serve as information for an error detection process. However, this early force magnitude monitoring process seemed to precede (if they were not completely independent) the process of determining the *correctness* of a certain response, in relation to 'the default force' required for a response to be deemed correct. Thus, although the brain appeared to be fast enough in term of planning the response parameters, the *correctness* aspect of the response itself was not yet foreshadowed in the neural signals.

4.4. Limitations and future research

Aiming to investigate monitoring and error processing of force execution, our modified paradigm has allowed us to eliminate potential confounds caused by *force selection* (De Bruijn et al., 2003) and therefore allowed us to investigate aspects of response dynamics (PF and TTP) in a pure force execution task. Our first sets of analyses was conducted to see whether ERP components reflected error monitoring, force monitoring, or a mixture of both. However, our ERP results showed no clear differentiation between error processing activity from force-time related activity, which is presumably related to the nature of this kind of task, as the information about the correctness is naturally contained in the force magnitude and modulated by its temporal dynamic. Our multivariate analyses results showed that force magnitude – instead of *correctness* was decodable from brain. However, we cannot exclude the possibility that error specific information was in fact encoded in the brain, and our method might not have been sensitive enough to pick up this information.

One potential avenue for future research is to ask participants directly to rate their force production as correct or incorrect after each response and before providing feedback. This would allow separating *detected* and *undetected errors,* whose differences might provide further insights about error detection processes in force production. This would allow whether information regarding *correctness* is available in trials with *detected* errors to be revealed. As all participants had little practice in performing our task, a further interesting question is whether it is possible to detect incorrect force production after more practice (using several practice sessions or by investigating experienced pianist; Parlitz et al., 1998) and if this will be reflected in a clearer differentiation between the early brain activity of correct and incorrect responses (CRN, Ne/ERN). Another possible modification of the current paradigm could be to experimentally vary TTP (i.e., short and long TTP; similar to Armbrecht et al., 2013) to systematically investigate how the timing of the peak force might serve as an aggravating or a supporting factor for successful force error detection. Another limitation is that although we balanced the trial numbers for each condition for each classification analysis, which always resulted in equal number of trials for both conditions in all training and test sets, there is always a risk that differences in variance between conditions could bias the classifier. However, there was no evidence for differences in variance in the higher-dimensional space in our patterns, and the SVR analyses did not include different conditions, which makes it unlikely that any systematic biases have occurred. Lastly, in the present study we found clear *post-error force adjustments*; however, we cannot determine whether the information resulted from internal or external feedback. Thus, an investigation of highly experienced participants without feedback presentation could provide further insights into error processing in force production, and for instance, answer the question whether *post--error adjustments* were based on internally acquired error information, or, in other words, whether this process operates without externally acquired error information (i.e., feedback).

5. Conclusions

The present study aimed to investigate performance monitoring in a pure *force execution* task by the means of classical ERP analyses (Ne/ERN, CRN, Pe, Pc, FRN) and MVPA, which allowed identifying the time when force- and time-related information of a force pulse was decodable from spatially distributed patterns of ERPs. Interestingly, we observed *post-error force adjustments* depending on the outcome of the previous trial (i.e., more or less force than the intended force) in the following trial, which seems to be equivalent to post-error slowing in speeded choice tasks (Notebaert et al., 2009). We found no clear error-specific variations but some evidence indicating force sensitivity in the classical ERP components. The more sensitive MVPA enabled us to determine that response-related parameters could be decoded from brain activity. The results of our first six sets of MVPA revealed that information regarding force magnitude – instead of *error-specific* processes – were decodable from the brain activity pattern. In support of this conclusion, *correct* vs. *incorrect* force ranges, without accounting for the *specific* force range, could not be decoded from brain activity patterns. Furthermore, the regression-based MVPA indicated that information about the single-trial response parameters (PF and TTP) was already decodable from brain activity patterns before response onset, which would be necessary (but not sufficient) for fast error detection. These findings suggested that although force-relevant parameters (PF and TTP) were already set in motor programs prior to the response onset, a clear representation of the response *correctness* was not fully developed (and hence present in the neural data) until around the time when a response was executed. As it is well-known from studies with piano players and other practitioners with fine motor skills (e.g., Ericsson and Lehmann, 1996; Maidhof et al., 2009) that these people are able to correct imperfect behaviour instantaneously, an interesting question for future research is how the internal representation of *correctness* is established for continuous response parameters and how deviations from the correct parameters can be detected fast and reliably.

Funding

This research was partially funded by Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP RI).

Acknowledgements

We are grateful to Maria Herzhauser, Lisa Marie Koch and Lena Wogenstein for assistance with data acquisition and Katharina Ohmann, Kilian Kummer, Manuela Hundrieser, Markus Mück, Sebastian Dummel, Ulrike Thesing, and Wilfried Follman for the helpful discussions and feedback on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.05.006.

References

Armbrecht, A.S., Gibbons, H., Stahl, J., 2012. Monitoring force errors: medial-frontal negativity in a unimanual force-production task. Psychophysiology 49 (1), 56–72.
Armbrecht, A.S., Gibbons, H., Stahl, J., 2013. Effects of response force parameters on medial-frontal negativity. PLoS One 8 (1), e54681.
Bode, S., Bennett, D., Stahl, J., Murawski, C., 2014. Distributed patterns of event-related potentials predict subsequent ratings of abstract stimulus attributes. PloS One 9 (10) e109070.
Bode, S., Feuerriegel, D., Bennett, D., Alday, P.M., 2019. The Decision Decoding ToolBOX (DDTBOX)–A multivariate pattern analysis toolbox for event-related potentials. Neuroinformatics 17 (1), 27–42.
Bode, S., Sewell, D.K., Lilburn, S., Forte, J.D., Smith, P.L., Stahl, J., 2012. Predicting perceptual decision biases from early brain activity. J. Neurosci. 32 (36), 12488–12498.

Bode, S., Stahl, J., 2014. Predicting errors from patterns of event-related potentials preceding an overt response. Biol. Psychol. 103, 357–369.
Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) 2 (3), 27.
Cordo, P.J., 1987. Mechanisms controlling accurate changes in elbow torque in humans. J. Neurosci. 7 (2), 432–442.
De Bruijn, E.R., Hulstijn, W., Meulenbroek, R.G., Van Galen, G.P., 2003. Action monitoring in motor control: ERPs following selection and execution errors in a force production task. Psychophysiology 40 (5), 786–795.
Desmedt, J.E., 1982. Size principle of motoneuron recruitment and the calibration of muscle force and speed in man. Adv. Neurol. 39, 227–251.
Ericsson, K.A., Lehmann, A.C., 1996. Expert and exceptional performance: evidence of maximal adaptation to task constraints. Annu. Rev. Psychol. 47 (1), 273–305.
Fahrenfort, J.J., Grubert, A., Olivers, C.N., Eimer, M., 2017. Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. Sci. Rep. 7 (1), 1886.
Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1990. Effects of errors in choice reaction tasks on the ERP under focused and divided attention. Psychophys. Brain Res. 1, 192–195.
Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J., 2000. ERP components on reaction errors and their functional significance: a tutorial. Biol. Psychol. 51 (2), 87–107.
Gehring, W.J., Goss, B., Coles, M.G., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4 (6), 385–390.
Gehring, W.J., Liu, Y., Orr, J.M., Carp, J., 2012. The error-related negativity (ERN/Ne). In: Oxford Handbook of Event-Related Potential Components, pp. 231–291.
Geisser, S., Greenhouse, S.W., 1958. An extension of box's results on the use of the *F* distribution in multivariate analysis. Ann. Math. Stat. 29 (3), 885–891.
Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artefact. Electroencephalogr. Clin. Neurophysiol. 55 (4), 468–484.
Hajcak, G., McDonald, N., Simons, R.F., 2003. To err is autonomic: error-related brain potentials, ANS activity, and post-error compensatory behaviour. Psychophysiology 40 (6), 895–903.
Hajcak, G., Holroyd, C.B., Moser, J.S., Simons, R.F., 2005. Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology 42, 161–170.
Hajcak, G., Moser, J.S., Holroyd, C.B., Simons, R.F., 2007. It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. Psychophysiology 44 (6), 905–912.
Hogendoorn, H., Burkitt, A.N., 2018. Predictive coding of visual object position ahead of moving objects revealed by time-resolved EEG decoding. Neuroimage 171, 55–61.
Holroyd, C.B., Coles, M.G., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109 (4), 679.
Holroyd, C.B., Krigolson, O.E., 2007. Reward prediction error signals associated with a modified time estimation task. Psychophysiology 44 (6), 913–917.
Holroyd, C.B., Krigolson, O.E., Baker, R., Lee, S., Gibson, J., 2009. When is an error not a prediction error? An electrophysiological investigation. Cognit. Affect Behav. Neurosci. 9 (1), 59–70.
Jasper, H.H., 1958. The 10/20 international electrode system. EEG Clin. Neurophysiol. 10, 371–375.
Jochumsen, M., Niazi, I.K., Mrachacz-Kersting, N., Farina, D., Dremstrup, K., 2013. Detection and classification of movement-related cortical potentials associated with task force and speed. J. Neural Eng. 10 (5), 056015.
Jochumsen, M., Niazi, I.K., Dremstrup, K., Kamavuako, E.N., 2016. Detecting and classifying three different hand movement types through electroencephalography recordings for neurorehabilitation. Med. Biol. Eng. Comput. 54 (10), 1491–1501.
Maidhof, C., Rieger, M., Prinz, W., Koelsch, S., 2009. Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. PLoS One 4 (4), e5032.
Nieuwenhuis, S., Ridderinkhof, K.R., Blom, J., Band, G.P., Kok, A., 2001. Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. Psychophysiology 38 (5), 752–760.
Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W., Verguts, T., 2009. Post-error slowing: an orienting account. Cognition 111 (2), 275–279.
Overbeek, T.J., Nieuwenhuis, S., Ridderinkhof, K.R., 2005. Dissociable components of error processing: on the functional significance of the Pe vis-à-vis the ERN/Ne. J. Psychophysiol. 19 (4), 319–329.
Parlitz, D., Peschel, T., Altenmüller, E., 1998. Assessment of dynamic finger forces in pianists: effects of training and expertise. J. Biomech. 31 (11), 1063–1067.
Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72 (2), 184–187.
Scheffers, M.K., Coles, M.G., Bernstein, P., Gehring, W.J., Donchin, E., 1996. Event-related brain potentials and error-related processing: an analysis of incorrect responses to go and no-go stimuli. Psychophysiology 33 (1), 42–53.
Stahl, J., 2010. Error detection and the use of internal and external error indicators: an investigation of the first-indicator hypothesis. Int. J. Psychophysiol. *77* (1), 43–52.
Steinhauser, M., Yeung, N., 2010. Decision processes in human performance monitoring. J. Neurosci. 30 (46), 15643–15653.
Steinhauser, M., Yeung, N., 2012. Error awareness as evidence accumulation: effects of speed-accuracy trade-off on error signalling. Front. Hum. Neurosci. 6, 240.
Ulrich, R., Wing, A.M., 1991. A recruitment theory of force-time relations in the production of brief force pulses: the parallel force unit model. Psychol. Rev. 98 (2), 268.
Ulrich, R., Wing, A.M., Rinkenauer, G., 1995. Amplitude and duration scaling of brief isometric force pulses. J. Exp. Psychol. Hum. Percept. Perform. 21 (6), 1457.
Vidal, F., Hasbroucq, T., Grapperon, J., Bonnet, M., 2000. Is the 'error negativity' specific to errors? Biol. Psychol. 51 (2), 109–128.