

A neuro-based method for detecting context-dependent erroneous robot action

Stefan Ehrlich¹ and Gordon Cheng¹

Abstract—Validating appropriateness and naturalness of human-robot interaction (HRI) is commonly performed by taking subjective measures from human interaction partners, e.g. questionnaire ratings. Although these measures can be of high value for robot designers, they are very sensitive and can be inaccurate and/or biased. In this paper we propose and validate a neuro-based method for objectively validating robot behavior in HRI. We propose to detect from the electroencephalogram (EEG) of a human interaction partner, the perception of *inappropriate* / *unexpected* / *erroneous* robot behavior. To validate this method, we conducted an EEG experiment with a simplified HRI protocol in which a humanoid robot displayed context-dependent erroneous behavior from time to time. The EEG data taken from 13 participants revealed biologically plausible error-related potentials (ErrP) whose spatio-temporal distributions match well with related neuroscientific research. We further demonstrate that perceived erroneous robot action can reliably be modeled and detected from human EEG signals with classification accuracies on avg. $69.7 \pm 9.1\%$. These findings confirm principal feasibility of the proposed method and suggest that EEG-based ErrP detection can be used for quantitative evaluation and thus improvement of robot behavior.

I. INTRODUCTION

Assessing, validating and improving quality of human-robot interaction (HRI) is so far mostly performed by collecting subjective measures from human interaction partners in experimental HRI setups, e.g. questionnaire ratings capturing the participant’s subjective experiences during or after the experiment. These measures can be of high value to roboticists and designers when aiming at improving naturalness and efficiency in HRI. However, they are often inaccurate or biased due to various reasons as they are highly prone to variations of the participant’s individual interpretation of the questions, the participant’s daily constitution, priming due to the experimenter involuntarily influencing the participant, and many more [1].

In this paper we propose and validate a neuro-based method for objectively measuring robot behavior in HRI. Using *electroencephalography* (EEG) signals acquired from a human interaction partner we aim at detecting the perception of inappropriate/unexpected/erroneous robot behavior (see Fig. 1). Although depending on the context, perception of inappropriate robot behavior can vary based on the participant’s individual interpretation, our proposed method is optimally objective in the sense that it captures neural responses immediate to the participant’s conscious realization of that robot action (approx. 300-500 ms thereafter). We suggest this

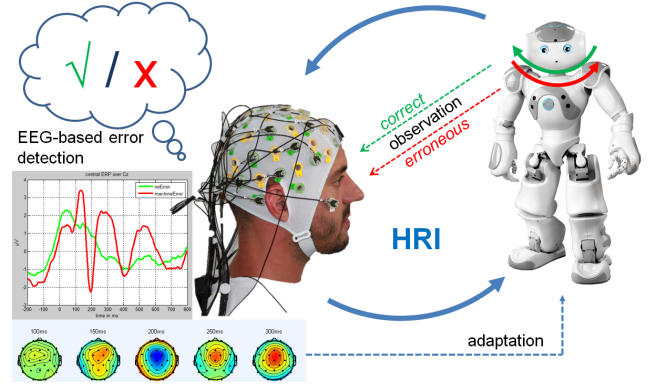


Fig. 1. Conceptual illustration of our approach: In interaction with a robot, a human interaction partner observes context-dependent inappropriate, unexpected or erroneous robot behavior. In response, the human electroencephalogram (EEG) shows deviating brain responses compared to appropriate, expected, and correct robot behavior. We suggest, decoding this information from the human EEG to be a valuable objective measure for quantifying and improving robot behavior in human-robot interaction (HRI).

measure can be of great help to roboticists in validating and improving/adapting robot behavior gradually or even on-the-fly during HRI.

We make the following contributions: We propose a neuro-based method to objectively evaluate robot behavior in HRI. For that, we (1) demonstrate that human perception of context-dependent erroneous robot behavior in HRI can be observed via EEG as specific neural responses, (2) we further demonstrate that these responses can reliably be detected from EEG signals, (3) and we provide a discussion on how we envisage to transfer these findings into a generalized method for objectively validating robot behavior. We believe the utilization of EEG to validate and improve HRI is a key component to progress in social robotics. This endeavour is as well reflected in one of our previous papers [2] in which we focused on human-robot eye-contact and engagement.

II. RELATED WORK

When humans make or perceive errors, a distinct response can be detected in the EEG due to the person realizing this error [3]. This response is related to *error-/performance-monitoring* processes in the brain, which are crucial for successful goal-directed behavior, decision making, planning and execution of tasks, but also flexible adaptive behavior, error handling and learning [4]. It is suggested that performance monitoring happens throughout a distributed network involving many regions of the brain, but related electrophysiological (EEG) responses have most consistently

¹Gordon Cheng (professor and department head) and Stefan Ehrlich are with the Chair for Cognitive Systems (ICS), Technische Universität München (TUM), Germany. stefan.ehrlich@tum.de

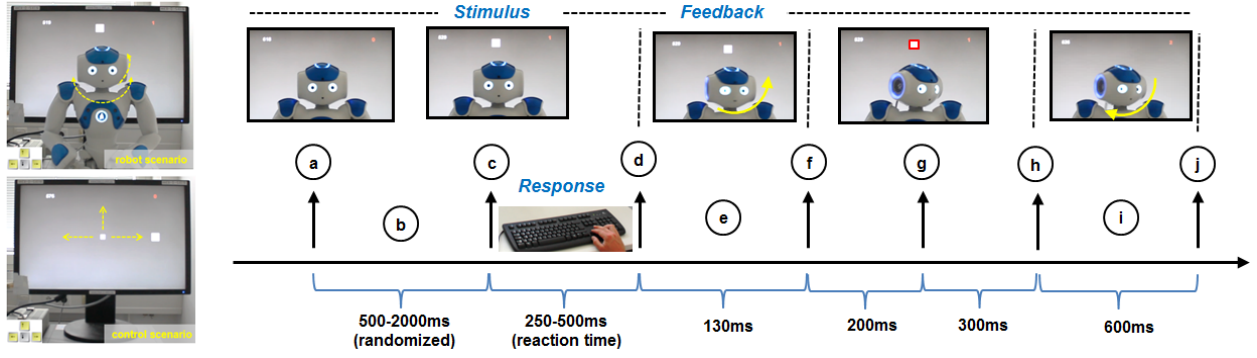


Fig. 2. Trial design: (a) Trial start, (b) break of random duration between 500-2000ms, (c) target stimulus presentation (in this case *up*), (d) participant response in form of arrow key press, start robot head movement (e) ongoing robot head movement (f) end robot head movement (g) target border feedback presentation (correct: *green*; incorrect: *red*), (h) disappearance target, start robot head turning back (i) ongoing robot head turning back (j) end of robot head turning back, updating average reaction time and error count.

been reported to originate from the posterior medial frontal cortex (pmFC) [4]. These responses are commonly referred as *error-related potentials (ErrP)*, a specific type of *event-related potentials (ERP)*. So far, fundamental research was mainly focused on self-inflicted errors, but recently researchers have observed similar responses when a person recognizes errors inflicted by others, such as a computer, a machine, or another human [5]. This finding has created attractive opportunities for engineering applications, specifically in the field of Brain-Computer Interfaces (BCI), which allow a person to communicate solely via brain activity. It has been shown that ErrPs can be automatically detected when a BCI delivers erroneous feedback [6] and that this information can be used for correction or adaptation of BCIs [7]. Typical ErrP single-trial classification accuracies are around 70-80% and relatively stable across recording sessions; ErrP classifiers have shown to still perform well up to 600 days after initial calibration [8] and can be applied across different task sets [9]. These works showed that decoding error-related potentials from human EEG can reliably be used to improve efficiency, performance and fluency in human-machine interaction. Most works along this line of research adopted (and sometimes modified) rather abstract experimental setups used in neuroscience, such as the *Choice-Reaction Time (CRT)* or the *Flanker* task. Some works have extended towards more naturalistic video-game based human-computer interaction, e.g. [10]. Only a few works have yet extended towards real physical machines such as industrial robotic arms, e.g. [11]. We present an extension of this line of research to the domain of human humanoid robot interaction for the measurement of context-dependent erroneous robot action.

III. EXPERIMENT

A. Short description, Objective, and Hypothesis

The purpose of the experiment was the investigation of perceived machine error in human-robot interaction. We developed a simplified HRI protocol in which a robot displayed context-dependent erroneous behavior from time to time. In order to link up well with the current state of research, we

developed a modified CRT-protocol: The participant's task was to react as quickly as possible to a target stimulus presented on the computer screen behind the robot with a corresponding key stroke. As a response, the robot head would move in the respective direction. From time to time, machine errors were introduced which manifested as a wrong feedback (robot head would turn in the wrong direction). We hypothesized to find ErrP responses in the human EEG comparable to those reported in previous literature, such as in [6], [8], [10], [11]. Based on this we further propose the principal feasibility to build EEG decoders capable of detecting these responses.

B. Environment and Data Recording

The experiment took place in a quiet room which was partitioned into two sections by means of a curtain. On the right side of the room, a participant was seated approximately 1.5 meters in front of the humanoid robot NAO positioned in a crouched posture in front of a computer screen. NAO is a 58 cm tall humanoid robot with 21-25 degrees of freedom [12] which was controlled by a program running on an external PC connected to the robot via local area network (LAN). For the robot control and thus the implementation of the experiment protocol we used the Python-based *Naoqi* library; for presentation of stimuli on the computer screen we used the *Psychopy* library. The robot was equipped with a light emitting diode (LED) + a photodiode at the back of the its head in order to record the ground truth timing of the head movements synchronously to the recording of the EEG signals. This setup was not directly visible to the participants and thus not distracting. EEG data was acquired with a Brain Products actiChamp amplifier equipped with 32 active EEG electrodes arranged according to the international 10-20 system. All leads were referenced to the average of TP9 and TP10 (linked mastoids referencing) and the sampling rate was set to 1000 Hz. The impedance levels of all leads were kept below 10 k Ω . Three out of the remaining 30 channels (FT9, FT10, Oz) were used for capturing electrooculogram (EOG) signals in three locations of the participant's face (forehead, left and right outer canthi) according to a method

suggested by Schlögl *et al.* in 2007 [13]. In addition, the amplifier was connected to the PC executing the experiment protocol via parallel port over which event triggers were transferred to be stored synchronously with the EEG signals. The computer keyboard for capturing the participant responses was located in near distance to the participant to allow for comfortable access.

C. Experimental Protocol

As briefly outlined in the beginning of this chapter, the participants were presented randomly one out of three possible target stimuli on a computer screen located behind the robot. In the initial head position (yaw=0°, pitch=0°) the robot was gazing directly at the participant. The target stimuli were realized as white rectangulars of size 3x3 cm appearing either left, right, or above the robot head. The participant's task was to respond as quickly as possible with a corresponding arrow key stroke (left, right, up). Upon participant response, the robot head turned towards the target (left: yaw+40°, right: yaw-40°, up: pitch-20°) and reached the end position after 130 ms, kept the position for about 500 ms and then moved back to the initial head position (see Fig. 2). Machine errors (wrong robot response) were introduced randomly with a probability $pErr$, however only if the participant's response was correct. In addition, all participants executed a second (control) scenario with identical protocol, in which the robot was replaced by a simplified cursor on the computer screen (see Fig. 2, bottom left)¹.

The experimental protocol was divided into two recording sessions (one for each scenario) which took place one after the other, whereby half of the participants started with the control-scenario and the rest with the robot-scenario. Each scenario was further divided into 10 blocks of 50 trials each; the duration of one block was approximately 2.5 minutes. Thus, the total duration of the experiment was approximately 60 minutes. After each block the participant could take a rest and decide when to continue with the next block in a self-paced fashion. In order to avoid habituation to the machine error probability, half of the blocks were executed with a machine error probability of $pErr = 20\%$ and the other half with $pErr = 50\%$. The order was pseudo-randomized such that no more than 2 subsequent blocks would belong to the same error probability category and the first block was always set to $pErr = 20\%$. As such 500 trials were collected per participant and scenario out of which were approximately 35% machine error trials and a negligible number of human committed error trials. As an incentive to the participants, the left upper corner of the screen informed about the average reaction time per block in milliseconds, the right upper corner about the number of errors per block with no distinction of errors committed by the participant or the machine.

¹The data collected in the control scenario will not be further discussed here as this would go beyond the scope of this paper.

D. Participants

Thirteen healthy participants (age: 30.3 ± 7.3 ; 5 females, 8 males) participated in the experiment. The majority of educational background was in non-engineering, non-technical majors (9 out of 13). Prior experience and familiarity with humanoid robots scored rather low with 2.6 ± 1.8 on a scale of 1 "non-familiar" to 7 "familiar". The participants were equally instructed about the experiment protocol and agreed on having their data acquired by signing a consent form. Each participant was paid a honorarium of 8 EUR/h.

IV. DATA ANALYSIS AND MODELING

A. Data Preprocessing

The data of the first participant (s01) had to be excluded from any further analysis due to technical issues during the experiment. All EEG data preprocessing was carried out in MATLAB, in part using functions provided by the EEGLAB toolbox [14]. The following steps were carried out for each of the remaining datasets (12 participants) separately: (1) Band-pass filtering of EEG and EOG channels using a zero phase Hamming windowed sinc FIR filter with cutoff frequencies of 1 Hz and 40 Hz. (2) Identification and interpolation of artifacted EEG channels using kurtosis and a threshold of 5%. (3) Automatic EOG artifact correction according to a method suggested by Schlögl *et al.* in 2007 [13]. This method does not only account for eye-blink artifacts but also horizontal eye-movements. (4) Re-referencing EEG channels to common average (CAR). (5) Determination of ground truth timing of onset of robot head movement using the signals captured by the photodiode. (6) Downsampling EEG data to 500 Hz in order to reduce processing time in all further analysis steps.

For further analyses, the data were epoched by extracting segments around the time² of feedback presentation (robot head turning). In particular, segments of -500 to 1500 ms with respect to the onset of feedback presentation were extracted. These epochs were further separated into 3 categories: (1) no error trials, (2) machine error trials, (3) human error trials. Per participant, we extracted on average approximately 325 no error trials, 159 machine error trials, and 16 human error trials. Since human errors were not in the focus of our investigation and their average number turned out to be very low, we discarded them from any further analysis.

B. Neurophysiological Analysis of Error-Related Potentials (ErrP)

We averaged the ERP time-courses of each category separately without applying any baseline correction and visualized them in Figure 3. As we expected the main effect to be centrally located (based on related literature), only the ERPs of channel Cz were depicted. Panel a shows the grand average of all participants for each category (noError, machineError). In addition we visualized the grand

²Please note that exact timing of onset of robot head movement was retrieved from the captured photodiode signals during data pre-processing

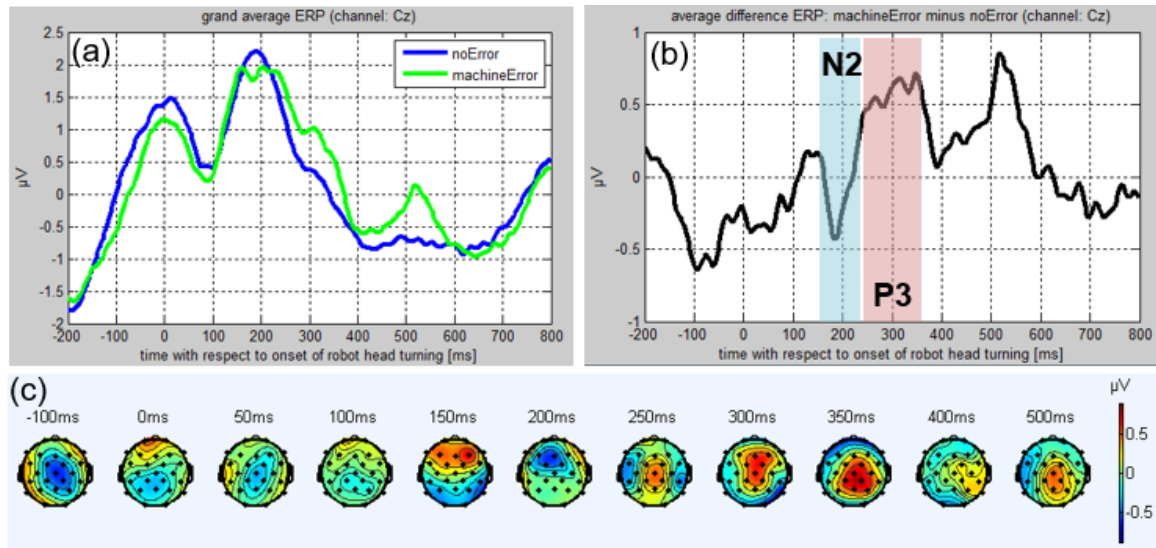


Fig. 3. Upper panel (a): grand average ERPs per category of channel Cz. Upper panel (b): grand average difference ERP (machineError minus noError). Lower panel (c): topographic spatio-temporal visualization of the grand average difference ERP.

average difference ERP (machineError minus noError) in panel b. The difference ERP shows the characteristic N2 component, a negativity around 200ms post stimulus, that has been reported consistently in relation to cases of high conflict (e.g. incongruent stimuli) [15]. When looking at the spatio-temporal topographical activation patterns (panel d) it becomes evident that the N2 component is more frontally located. Also, the characteristic P3 component (centrally located around 300 ms post stimulus) is visible and enhanced in case of an error. P300 modulations have consistently been associated with stimuli expectancy (enhanced P3 in case of unexpected events) [?]. In addition we observe another error-related enhanced positive component around 500 ms post stimulus which is distributed over central-parietal sites. This effect might be related to the P600 component which has been associated with syntactic/semantic anomalies [?]. The observed ERPs are biologically plausible in terms of spatio-temporal topographical distribution. In fact, the observed effects qualitatively resemble previously reported error-related potentials from different but related experimental contexts, e.g. in [6], [10].

C. Data Modeling and Evaluation

The purpose of modeling the data was to demonstrate the feasibility to discriminate based on EEG data, the two classes of mental states under investigation (class1: response to noError event, class2: response to machineError event). This was carried out in three steps: (1) extraction of relevant features in the time- as well as in the frequency-domain, (2) classifier training, and (3) evaluation. We intentionally decided to employ a relatively simple modeling approach to demonstrate principal feasibility, foster transparency, and illustrate the lower boundary of achievable classification performance.

1) *Feature Extraction*: We used standard feature extraction techniques described in related literature. Most literature on ERP modeling and classification employ temporal features e.g. [6], some few have explored spectral features as well, e.g. [10].

Time-domain features: As for temporal features we computed the arithmetic mean in pre-defined time-slots of 0-100, 100-200, 150-250, 200-300, 250-350, 300-400, 400-500, 500-600, 600-800 ms with respect to the onset of feedback presentation. As such we obtained 243 temporal features per epoch (27 channels x 9 time slots).

Frequency-domain features: As for spectral features we first bandpass filtered each epoch using two subsequent Chebychev Type II zero-phase filters (separate high-pass with cutoff frequency 4Hz and low-pass with cutoff frequency 13Hz) with maximum passband attenuation of 3 dB and minimum stopband attenuation of 50 dB. Previous literature reported the ErrP to be mainly represented in theta-band (4-8Hz) [16]; our own experiences, however, show that in some participants the manifestation of ErrP can reach higher in frequencies and we decided therefore to include the alpha band. We computed the spectral power in the same time-slots as those used for temporal features. The spectral power was obtained by computing the natural logarithm of the variance of the signal of each time-slot ($\ln(\text{var}(x))$). As such we obtained 243 spectral features per epoch.

2) *Classification*: As for modeling this binary classification problem we used linear discriminant classifiers (LDA) and evaluated the two different feature types separately. In addition we evaluated the combination of both feature types. As we had to deal with a relatively high-dimensional feature space, dimensionality reduction was crucial to overcome the curse of dimensionality. Instead of employing feature-

TABLE I
10-TIMES 10-FOLD CROSS-VALIDATION RESULTS

ID	Temporal ($\lambda = 0.48$)			Spectral ($\lambda = 0.91$)			Combined ($\lambda = 0.77$)		
	noErr [%]	machErr [%]	accuracy [%]	noErr [%]	machErr [%]	accuracy [%]	noErr [%]	machErr [%]	accuracy [%]
s02	74.5	60.1	69.9	64.0	51.3	59.9	74.0	58.4	69.0
s03	75.2	66.1	72.0	65.3	59.7	63.3	74.1	65.7	71.1
s04	84.9	71.6	79.9	80.7	66.5	75.3	89.8	71.8	83.0
s05	70.9	61.9	67.8	63.2	56.0	60.7	70.5	62.1	67.6
s06	85.9	71.8	80.6	67.5	70.4	68.6	86.4	70.6	80.4
s07	82.3	69.9	78.3	69.6	59.9	66.5	82.9	68.4	78.2
s08	64.8	58.6	62.5	54.2	53.2	53.8	64.4	56.6	61.5
s09	75.4	62.0	70.7	61.2	56.8	59.6	75.8	62.5	71.2
s10	60.8	47.6	56.4	54.0	52.8	53.6	61.4	42.9	55.2
s11	72.6	61.2	68.6	63.8	60.3	62.6	71.9	58.9	67.4
s12	58.0	39.4	51.7	56.5	44.9	52.6	63.8	39.3	55.6
s13	78.8	70.6	75.8	71.1	63.9	68.4	81.4	69.0	76.8
AVG \pm SD	73.7 \pm 9.9	61.7 \pm 9.9	69.5 \pm 9.0	64.3 \pm 7.6	58.0 \pm 7.0	62.1 \pm 6.9	74.7 \pm 9.1	60.5 \pm 10.3	69.7 \pm 9.1

selection/-pruning algorithms we decided to use a regularized version of the LDA classifier [17] which controls the computation of the covariance estimates using a shrinkage function (see equation (2)).

The classifier training is given by equations (1)-(3):

$$\vec{w} = (\mu_2 - \mu_1)(\sigma_{reg1} + \sigma_{reg2})^{-1} \quad (1)$$

with \vec{w} being the classifier weight vector, $\mu_{1,2}$ being the class-wise means and $\sigma_{reg1,2}$ the regularized class-wise covariance estimates, with:

$$\sigma_{reg1,2} = (1 - \lambda)\sigma_{1,2} + \lambda \nu I \quad (2)$$

and λ being the shrinkage parameter with $\lambda \in [0, 1] \subset \mathbb{R}$.

$$b = -\vec{w}(\mu_1 + \mu_2)/2 \quad (3)$$

with b being the classifier bias.

The classifier decision function is given by equation (4):

$$y = \text{sign}(\vec{w}\vec{x}^T + b), \quad (4)$$

with \vec{x} being the feature vector and y the class decision (-1: noError; 1: machineError).

3) *Evaluation*: We used 10-fold cross-validation to train the classifier on 9 out of 10 splits and tested it with the remaining split. Then folds were shuffled until each fold had once been used for testing. The results were averaged and reported as class-wise and mean classification accuracies. Evaluation was performed on each participant separately.

Please note that the number of observations per class were initially unbalanced (machineError observations were approx. 35% of the total number of observations). Therefore, before classifier training the number of training samples were balanced by random observation picking from class *noError*. Thus, cross-validation was repeated for 10 times to yield average results. Testing was always performed on the unbalanced split. The shrinkage parameter λ was manually tuned based on the data of the first participant (s02) for

each feature type and their combination separately using grid-search. The obtained λ -values were then applied to the remaining participants.

4) *Results*: Table I shows the cross-validation classification results. We obtained the highest mean classification accuracy of 69.7% with combined features, however not significantly higher than with temporal features alone. Temporal features generally outperformed spectral features by approx. 7%. The mean classification accuracies are significantly above chance level $p < 0.05$ for all participants and feature types (temporal, spectral, combined)³. In addition, we observed a consistent bias towards the class *noError*, meaning, the specificity (true negative rate: correct classification of *noError*) is higher than the sensitivity (true positive rate: correct classification of *machineError*). Since we took care of class balancing, we assume this phenomenon to be related to slight but systematic deviations between the distributions of both classes. Based on this assumption we propose classifiers with less sensitivity to class distribution (e.g. Support Vector Machines) may lead to more balanced results.

V. DISCUSSION

Our findings indicate (1) that the effects we elicited with our experiment and captured via EEG originate from performance monitoring related brain processes, since both spatial and temporal distributions resemble previously reported ErrPs in related but different experimental contexts. (2) We demonstrated that these responses can successfully be modeled and detected on a participant-dependent level. We understand these results as an entry point to the use of ErrP detection in validating robot behavior in HRI. In this section, we aim at highlighting a series of implications and challenges we envisage along the path of implementing this concept in HRI validation:

³Validated with one-sample t-test against 50% based on 100 individual classification results obtained by 10-times repeating 10-folds cross-validation.

Classification performance: - *Is the average classification accuracy of 70% high enough and are the classifiers robust enough with regard to the proposed method?* (1) Classification performance should be understood in a probabilistic way in the context of this work, meaning: out of a series of erroneous robot actions, how many would actually be detected as erroneous given that correct actions would robustly be detected as correct. By shifting the classifier bias to near 100% sensitivity we still obtain approx. 40% specificity, meaning, every third erroneous robot action would reliably be detected as such and be usable for adaptation of the robot's behavior. (2) ErrPs have been shown to be reliable across time [8], therefore we expect our classification models to be re-usable in subsequent recording/online-application sessions with comparable performance.

Event-based detection: - *What if robot actions were continuous, non-event-based?* At the current stage, our proposed method relies on events which are clearly isolated and separated in time. Also, our method relies on precise information about when in time these events happened. Depending on the context, however, HRI can be less event-based, but more continuous or a superposition of events on continuous interaction. As some works have successfully shown non time-locked ErrP detection (continuous), e.g. [18], [10], we do not expect a serious drop in classification performance when transferring our method to more continuous interaction protocols without precise information about timing of events. However, it should be noted that the proposed method does rely on events, namely the human recognition of something suddenly odd, inappropriate, or erroneous in the robot's behavior, even if these events are embedded in continuous trajectories.

Task specificity: - *Are the observed effects and the classification models specific to the task employed in this experiment?* Among other works, e.g. [9], our work confirms again the inspecificity of ErrP responses to the experimental context. The qualitative resemblance of the observed ErrP with results reported by others speaks in favor for generic underlying neural processes which seem not to be specifically influenced by the type of task/interaction or the stimulus appearance/meaning. In this regard we do not expect significant deviations of the observed ErrP when changing the experimental conditions, e.g. from simplified head turning to other gestures or even language based interaction, etc.

VI. CONCLUSION

In this paper we proposed a neuro-based method for quantitative evaluation of robot behavior with respect to inappropriate / erroneous actions. We validated this method in an experiment with a simplified HRI protocol and showed that ErrPs can be observed in the EEG signals of a human interaction partner when observing a humanoid robot performing context-dependend erroneous actions. We further

demonstrated that these responses can successfully be modeled and detected from human EEG with avg. accuracies of $69.7 \pm 9.1\%$. With these findings we confirm principal feasibility of our proposed method and suggest that EEG-based ErrP detection can be used for quantitative evaluation and improvement of robot behavior and HRI.

REFERENCES

- [1] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 33–40.
- [2] S. Ehrlich, A. Wykowska, K. Ramirez-Amaro, and G. Cheng, "When to engage in interaction-and how? EEG-based enhancement of robot's ability to sense social signals in HRI," in *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014, pp. 1104–1109.
- [3] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: a tutorial," *Biological psychology*, vol. 51, no. 2, pp. 87–107, 2000.
- [4] M. Ullsperger, C. Danielmeier, and G. Jocham, "Neurophysiology of performance monitoring and adaptive behavior," *Physiological reviews*, vol. 94, no. 1, pp. 35–79, 2014.
- [5] H. T. van Schie, R. B. Mars, M. G. Coles, and H. Bekkering, "Modulation of activity in medial frontal and motor cortices during error observation," *Nature neuroscience*, vol. 7, no. 5, pp. 549–554, 2004.
- [6] P. W. Ferrez and J. d. R. Millán, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE transactions on biomedical engineering*, vol. 55, no. 3, pp. 923–929, 2008.
- [7] 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*, vol. 24, no. 10, pp. 1120–1127, 2011.
- [8] R. Chavarriaga and J. d. R. Millán, "Learning from eeg error-related potentials in noninvasive brain-computer interfaces," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 18, no. 4, pp. 381–388, 2010.
- [9] T. Plewian, E. Wascher, M. Falkenstein, and S. Hoffmann, "Classifying response correctness across different task sets: A machine learning approach," *PloS one*, vol. 11, no. 3, p. e0152864, 2016.
- [10] 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*, vol. 9, p. 155, 2015.
- [11] I. Iturrate, R. Chavarriaga, L. Montesano, J. Mínguez, and J. d. R. Millán, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control," *Scientific reports*, vol. 5, 2015.
- [12] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, "The NAO humanoid: a combination of performance and affordability," *CoRR abs/0807.3223*, 2008.
- [13] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical neurophysiology*, vol. 118, no. 1, pp. 98–104, 2007.
- [14] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.
- [15] T. O. Gruendler, M. Ullsperger, and R. J. Huster, "Event-related potential correlates of performance-monitoring in a lateralized time-estimation task," *PloS one*, vol. 6, no. 10, p. e25591, 2011.
- [16] N. Yeung, R. Bogacz, C. B. Holroyd, S. Nieuwenhuis, and J. D. Cohen, "Theta phase resetting and the error-related negativity," *Psychophysiology*, vol. 44, no. 1, pp. 39–49, 2007.
- [17] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.
- [18] J. Omedes, I. Iturrate, and L. Montesano, "Asynchronous detection of error potentials," in *Proceedings of the 6th Brain-Computer Interface Conference 2014*, no. EPFL-CONF-205139, 2014.