Correcting Robot Mistakes in Real Time Using EEG Signals

Andres F. Salazar-Gomez*, Joseph DelPreto†, Stephanie Gil†, Frank H. Guenther*, and Daniela Rus†

***Abstract*— Communication with a robot using brain activity from a human collaborator could provide a direct and fast feedback loop that is easy and natural for the human, thereby enabling a wide variety of intuitive interaction tasks. This paper explores the application of EEG-measured error-related potentials (ErrPs) to closed-loop robotic control. ErrP signals are particularly useful for robotics tasks because they are naturally occurring within the brain in response to an unexpected error. We decode ErrP signals from a human operator in real time to control a Rethink Robotics Baxter robot during a binary object selection task. We also show that utilizing a secondary interactive error-related potential signal generated during this closed-loop robot task can greatly improve classification performance, suggesting new ways in which robots can acquire human feedback. The design and implementation of the complete system is described, and results are presented for real-time closed-loop and open-loop experiments as well as offline analysis of both primary and secondary ErrP signals. These experiments are performed using general population subjects that have not been trained or screened. This work thereby demonstrates the potential for EEG-based feedback methods to facilitate seamless robotic control, and moves closer towards the goal of real-time intuitive interaction.**

I. Introduction

Using brain signals to control robots could offer exciting possibilities for intuitive human-robot interaction. Although capturing and identifying such signals represents a considerable challenge given current technology, recent research has shown that the *error-related potential* (ErrP) signal is generated by the brain in response to observing or making a mistake. If this signal could be leveraged to facilitate human-robot control even in a restricted class of situations, it would enable many new applications of natural human-robot collaboration. For example, humans could remotely supervise robots on factory floors and communicate "stop" instantaneously when the robot makes a mistake without needing to type a command or push a button.

Reliably detecting this error-related potential could enable communication via a signal that occurs naturally in the brain during interaction with, or observation of, a collaborating robot. This could potentially alleviate the extensive user training, extra cognitive load, or constant visual stimuli often required by common brain-computer interface (BCI) systems. Due to the inherent difficulty of quickly extracting such signals from a subject's brain activity using electroencephalography (EEG), studies involving error-related potentials are often performed in controlled settings and for simulated or open-loop tasks. However, robotic applications demand closed-loop scenarios in real-world environments; this paper therefore explores the applicability of EEG-measured ErrP signals to real time closed-loop robotic tasks.

*Boston University Speech Lab, Boston, MA 02215
{salacho, guenther}@bu.edu
†MIT Distributed Robotics Lab, Cambridge, MA 02139
{delpreto, sgil, rus}@csail.mit.edu

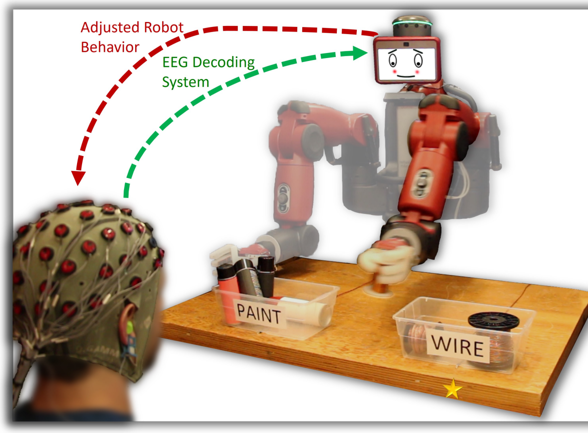


Fig. 1: The robot is informed that its initial motion was incorrect based upon real-time decoding of the observer's EEG signals, and it corrects its selection accordingly to properly sort an object.

Towards this end, a feedback system is developed for human-robot collaboration that hinges upon online identification of error-related potentials. In particular, a Rethink Robotics Baxter robot performs object selection while being observed by a human. The human operator's EEG signals are collected and decoded in real time; if an ErrP signal is detected, the robot immediately corrects its trajectory. Figure 1 depicts the system's operation during a sorting task, which is a simple extension of the object selection task.

An important result obtained by these experiments is the observation and analysis of interactive error-related potentials, namely *secondary errors*, generated in online closed-loop experiments due to the human's active participation. They naturally occur in response to a real-time misclassification of the EEG signal, i.e. when the robot does not properly obey the human's feedback. These secondary errors are often significantly easier to classify than the initial error-related potential (primary error). This signal can therefore improve system performance and greatly aid the development of EEG-based closed-loop controllers for robotic tasks; this paper provides preliminary offline analysis towards this goal.

A. Human-Robot Interaction using ErrP Communication

There are many ways that a human can interact with a robot using ErrP signals for communication. The human can be actively included or excluded from the control loop, and the computation can be performed in real time or after the experiment is completed. Specifically:

Closed-Loop implies that the robot and human directly affect each other throughout the task; the behavior of one affects the other. For example, the EEG classifier detects an ErrP and communicates a trajectory correction to Baxter. This adjustment is immediately observed by the human, affecting their behavior and EEG signals, completing the closed loop.

Open-Loop implies that the robot performs the task without any feedback from the human. The robot is still observed by the human and EEG data is acquired, but it is not decoded in real time to command the robot. The human's role may seem similar to the closed-loop scenario, but the lack of collaboration significantly affects the subject's engagement.

Online Performance implies that the EEG classification system operates in real time, as is necessary for closed-loop feedback. To achieve this, the system must acquire EEG data for less than one second and then make a classification decision within milliseconds; longer latencies would deteriorate the effectiveness of closed-loop feedback.

Offline Performance is obtained by running the EEG classification system on pre-recorded EEG data after an open-loop or closed-loop experiment is completed. Offline analysis has no constraint on computation time, so it generally achieves better performance by optimizing the classification pipeline.

Due to the nature of real-world robotic applications, this work focuses on developing an online closed-loop system. EEG data is acquired for brief windows during the task and decoded in a few hundred milliseconds to provide immediate feedback to the robot. 12 subjects participated, only one of which had ever used an EEG system before, and none had been previously trained for our task. One of the subjects was in a meditative state, and the data of that subject was excluded from the present analysis for consistency. The classifier used in online sessions was trained on a single-experiment basis using only 50 initial trials. Offline analysis is also presented, with optimal classification parameters using all available data; with advances in computation and hardware, similar results could eventually be achieved online.

B. Paper Contributions

This paper develops and tests an online feedback system for an object selection task where human directives are communicated to the robot via an EEG classification system. In particular, this work presents the following:

- Demonstration of the existence and applicability of error-related potentials to robotic collaboration tasks
- Closed-loop ErrP classification, allowing a human to influence a robot in real time via natural thought patterns and the robot to influence the human's EEG activity
- System performance results for online and offline analysis of both open-loop and closed-loop scenarios
- Demonstration of secondary ErrPs in closed-loop scenarios, and their offline utilization to boost performance

This work therefore makes progress towards the goal of intuitive human-robot interaction by exploring methods of applying EEG data to real-time robotic control.

II. Literature Review

A. EEG-based methods for Robot Tasks

The emergence of non-invasive electroencephalography (EEG) and its natural applicability to human-robot collaboration tasks has spurred much research over the past decade. For example, EEG signals have been used with motor imagery tasks to control robots such as a quadcopter or wheelchair [1], [2]. While this certainly showcases the promise of using EEG for robot applications, typical approaches often require several training phases to pre-screen operators based on task proficiency and for the human operator to learn how to modulate thoughts appropriately. The most common tasks for EEG-based control use signals produced in response to stimuli (usually visual). These include the P300 signal, in conventional grids [3] or during rapid serial visual presentation (RSVP) [4], and the steady state visually-evoked potentials (SSVEP) [5], [6]. While these have shown good average classification performance, they require constant attention of the operator, add cognitive or visual burden, and usually require many repeated prompts in order to decode a single command. Such challenges make these approaches less amenable to closed-loop control tasks.

B. The Error-Related Potential Signal

A more desirable approach would be one that utilizes a naturally occurring brain signal, thereby generalizing to many different tasks and not requiring extensive training or active thought modulation by the human. One such signal that has been observed is the *error-related potential* (ErrP) [7]. This signal is consistently generated when a human consciously or unconsciously recognizes that an error has been committed, even if the error is made by someone else [8], [9], [10], [11]. Researchers believe ErrPs are an integral component of the natural trial-and-error learning process. Interestingly, researchers have found that these signals demonstrate a characteristic shape even across users with no prior training (see Figure 2). In addition, they are typically detectable within 500 ms of when the human observes the error [7], [9]. This suggests that error-related potentials are particularly well-suited for use in robotic tasks.

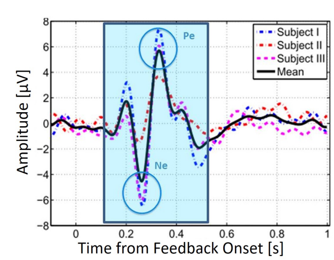


Fig. 2: Error-Related Potentials exhibit a characteristic shape across subjects and include a short negative peak, a short positive peak, and a longer negative tail. Image adapted from [8].

C. Applicability of Error-Related Potentials to Robotics

Recently, efforts have been made to apply error-related potentials to robotic tasks [12], [13]. These works show that

ErrPs can be used to improve performance via reinforcement learning, but have not used error-related potentials in real-time tasks where this is the sole input (i.e. no motor imagery or thought modulation). More recently, the existence of ErrPs and the ability to classify them in an online setting has been demonstrated in a real-life driving task [14], but these signals have not been used in a closed feedback control loop to change the system's behavior.

This work aims to present a new paradigm for human-robot interaction that improves the probability by which ErrPs can be successfully extracted on a single-trial basis and used to modify a robot's behavior in real time. Additionally, in contrast to many Brain Computer Interface (BCI) approaches, our method is applied to subjects who have neither been trained for the task nor used BCIs before.

D. Novel Contributions

The work that we present in this paper is different from existing work in several key ways:

- ErrPs are the sole communication channel from the human to the robot, and are used in a closed-loop task. This enables real-time communication and correction of robot errors, which could then trigger secondary ErrPs.
- The interface is designed around the human; the robot adapts to the human rather than the other way around.
- Human subjects are neither trained for the task nor screened based on EEG performance. This moves one step closer towards bringing robot control to everyday operators who are naive to the system, task, and BCIs.

III. System and Experimental Design

To explore the application of EEG-detected ErrP signals to robotic tasks, a paradigm and feedback system were designed and implemented to allow a human observer to interact with a Baxter robot. The general setup can be seen in Figure 1, where the human observes the robot's actions and influences its behavior using natural thought patterns.

A. Experimental Methodology

An experimental paradigm was designed where a human *passively observes* whether a robot performing a binary-choice reaching task makes a correct or incorrect decision, and resulting EEG signals are used to influence the robot in real time. A typical recording session lasted approximately 1.5 hours including EEG cap preparation, and was separated into 4 blocks (for the closed-loop sessions) or 5 blocks (for the open-loop sessions). Each block contained 50 trials and lasted 9 minutes. This paradigm was implemented at the MIT Distributed Robotics Laboratory.

1) Binary Choice Paradigm: The sequence of events comprising the paradigm is illustrated in Figure 3. During the experiment, a subject wearing an EEG cap is seated 50 cm from Baxter and judges whether Baxter's object selection is correct while a decoder searches for ErrP signals. At the start of each trial, the subject gazes at a fixation point placed directly below Baxter's arm and in the center of two blinking LEDs. One of the two LEDs then cues the desired target (left or right). Baxter then randomly selects a target with a 50/50 or 70/30 bias towards choosing correctly (for open-loop and closed-loop experiments respectively), and performs a two-stage reaching movement. The first stage is a lateral movement that conveys Baxter's intended target and releases a pushbutton switch to initiate the EEG classification system. The human mentally judges whether this choice is correct, and the system informs Baxter to continue toward the intended target or switch to the other target based on whether an ErrP is detected. The second stage of the reaching motion is then a forward reach towards the object.

If a misclassification occurs, a secondary error may be induced as shown in Figure 3b since the robot did not obey the human's feedback. This second ErrP may be used in the future to cause another trajectory change and thereby ultimately choose the correct target.

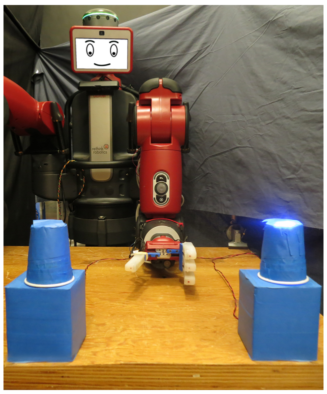
The binary choice paradigm was implemented under two different conditions:

Open-loop sessions: No online classification was running while Baxter performed the target selection, and a one-stage reaching movement was used. The subjects passively evaluated Baxter's performance, and were aware that their EEG signals were not controlling it. In seven out of eight sessions the probability of Baxter choosing the correct target was 50%, and in the remaining session it was 70%. Offline analysis of these experiments confirmed the presence of the error-related potentials and optimized the parameters for the classifier used in the closed-loop sessions.

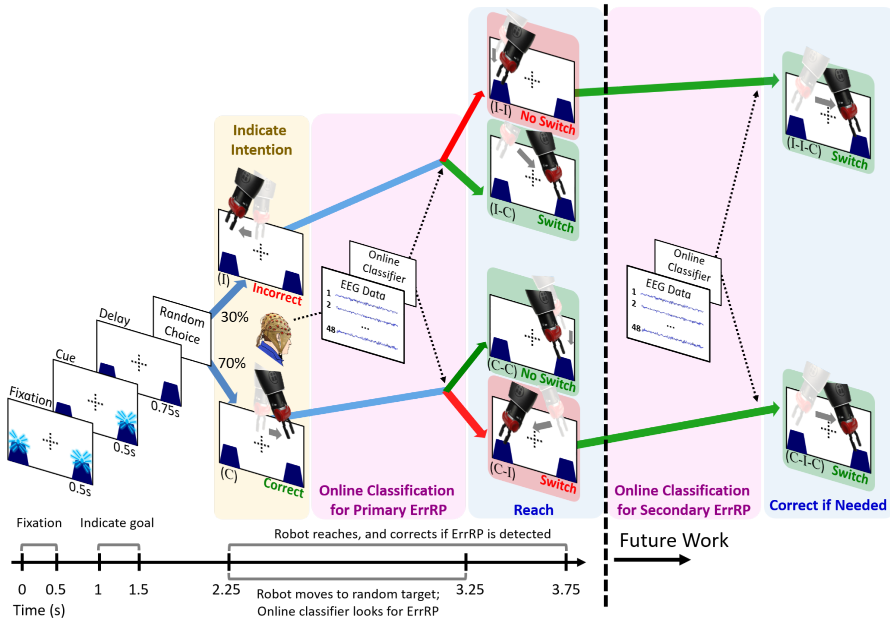
Closed-loop sessions: The subject's EEG signals were used to control Baxter's behavior in real time. A full session included four blocks of trials. The first block was used to collect training data; no classification was performed, but the controller randomly decided whether to inform Baxter of an error to induce a secondary ErrP in the subjects. At the end of each block, a new classifier was trained with the data from the current session. Online classification was performed and used as closed-loop feedback for all blocks after the first.

2) Subject Selection: All subjects provided informed consent for the study, which was approved by the Internal Review Board of Boston University and the Committee on the Use of Humans as Experimental Subjects of MIT. For all EEG recordings, participants were recruited through community advertisements at Boston University and MIT, were selected from the general population and did not undergo any training sessions. Subjects were not screened based on their innate ability to produce the desired error-related potentials or their experience with EEG systems or BCI.

A total of twelve individuals (91.67% right-handed and 83.33% male) participated in the experiments, seven in the open-loop condition (85.71% right-handed, 71.43% male), and five in the closed-loop paradigm (100% male and 100% right-handed). From the five subjects participating in the closed-loop paradigm, only data from four of them is presented. The fifth subject performed the task while being in a meditative state, and for consistency we exclude that subject's data from the analysis.



(a) The subject is seated directly between two possible targets with indicator LEDs - here, the right LED indicates the correct target. Curtains surround the test area and subject to minimize distractions.



(b) Two LEDs first flash to cue the subject's attention, then one LED indicates the randomly chosen correct target. Baxter randomly chooses a target, biased towards choosing correctly, and makes an initial lateral movement to indicate its choice which is correct (C) or incorrect (I). Online classification of EEG data determines if a trajectory switch is required; this leads to four possible outcomes based on whether the initial movement and classification outcome were correct or incorrect. In the future, secondary error-related potentials can be used to correct misclassifications in the I-I or C-I cases.

Fig. 3: The experimental paradigm implements a binary reaching task; a human operator mentally judges whether the robot chooses the correct target, and online EEG classification is performed to immediately correct the robot if a mistake is made.

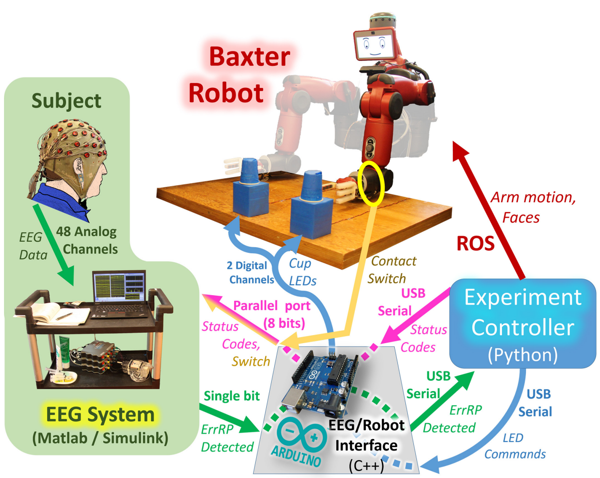


Fig. 4: The system includes a main experiment controller, the Baxter robot, and an EEG acquisition and classification system. An Arduino relays messages between the controller and the EEG system. A mechanical contact switch detects arm motion initiation.

B. Robot Control and EEG Acquisition

The control and classification system for the experiment is divided into four major subsystems, which interact with each other as shown in Figure 4.

The **experiment controller**, written in Python, oversees the experiment and implements the chosen paradigm. For example, it decides the correct target for each trial, tells Baxter where to reach, and coordinates all event timing.

The **Baxter robot** communicates directly with the experiment controller via the Robot Operating System (ROS). The controller provides joint angle trajectories for Baxter's left 7 degree-of-freedom arm in order to indicate an object choice to the human observer and to complete a reaching motion once EEG classification finishes. The controller also projects images onto Baxter's screen, normally showing a neutral face but switching to an embarrassed face upon detection of an ErrP; see Figure 3a and Figure 1, respectively.

The **EEG system** acquires real-time EEG signals from the human operator via 48 passive electrodes, located according to the extended 10/20 international system and sampled at 256 Hz using the g.USBamp EEG system [15]. Based on previously studied ErrP characteristics, this configuration ensures that the dominant traits of the desired signal will be captured. It will also capture other higher cognitive processes and task-related EEG activity. A dedicated computer uses Matlab and Simulink to capture, process, and classify these signals. The system outputs a single bit on each trial that indicates whether a primary error-related potential was detected after the initial movement made by Baxter's arm.

The **EEG/Robot interface** uses an Arduino Uno that controls the indicator LEDs and forwards messages from the experiment controller to the EEG system. It sends status codes to the EEG system using 7 pins of an 8-bit parallel port connected to extra channels of the acquisition amplifier. A pushbutton switch, discussed in detail below, is connected directly to the 8^{th} bit of the port to inform the EEG system of

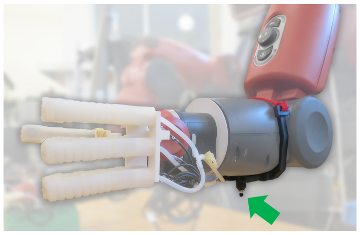


Fig. 5: A mechanical momentary switch affixed to the bottom of Baxter's arm precisely determines when the arm begins moving. This feedback onset time is vital for reliable EEG classification.

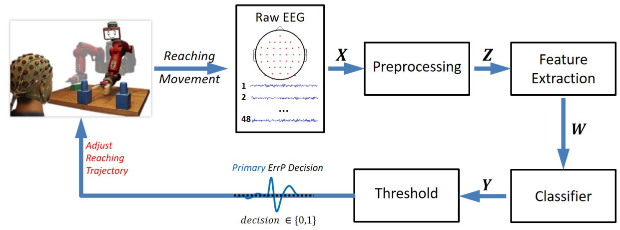


Fig. 6: Various pre-processing and classification stages identify ErrPs in a buffer of EEG data. This decision immediately affects robot behavior, which affects EEG signals and closes the feedback loop.

robot motion. The EEG system uses a single 9^{th} bit to send ErrP detections to the Arduino. The Arduino communicates with the experiment controller via USB serial.

Experiment codes that describe events such as stimulus onset and robot motion are sent from the experiment controller to the Arduino, which forwards these 7-bit codes to the EEG system by setting the appropriate bits of the parallel port. All bits of the port are set simultaneously using low-level Arduino port manipulation to avoid synchronization issues during setting and reading the pins. Codes are held on the port for 50 ms before the lines are reset. The EEG system thus learns about the experiment status and timing via these codes, and uses this information for classification. In turn, it outputs a single bit to the Arduino to indicate whether an error-related potential is detected. This bit triggers an interrupt on the Arduino, which then informs the experiment controller so that Baxter's trajectory can be corrected. The experiment controller listens for this signal throughout the entirety of Baxter's reaching motion.

Knowing the exact moment of arm motion initiation, i.e. the feedback onset time, is vital for reliable EEG classification. This is achieved via a hardware switch mounted to the bottom of Baxter's arm as shown in Figure 5. The switch contacts the table when the arm is in its neutral position, and it is immediately released when motion begins. One pole of the switch is wired to 5 V, while the other is passively pulled low by a resistor and wired directly to the 8^{th} pin of the aforementioned parallel port. This allows the EEG system to detect the instant at which the arm motion begins, much more precisely and reliably than if a software trigger was used via the Arduino parallel port.

While the above switch robustly detects Baxter's initial motion and is used for the classification of a primary ErrP, there is no analogous switch for detecting secondary errors. Detection of the secondary error, which indicates whether the first classification was incorrect, should begin as soon as the human observer processes that the robot chose to continue towards the initial target or switches to the opposite target. Since the arm is in free space when this motion begins, and it is unclear when the human interprets the robot's target destination, it is difficult to add a hardware switch. Instead, many trials were run to determine when the arm begins its second motion relative to the initial switch liftoff, and this offset was used as the second feedback onset time.

IV. Training and ErrP Classification

The classification pipeline and feedback control loop are illustrated in Figure 6. The robot triggers the pipeline by moving its arm to indicate object selection; this moment is the feedback onset time. A window of EEG data is then collected and passed through various pre-processing and classification stages. The result is a determination of whether an ErrP signal is present, and thus whether Baxter committed an error. The implemented online system uses this pipeline to detect a primary error in response to Baxter's initial movement; offline analysis also indicates that a similar pipeline can be applied to secondary errors to boost performance.

This optimized pipeline achieves online decoding of ErrP signals and thereby enables closed-loop robotic control. A single block of 50 closed-loop trials is used to train the classifier, after which the subject immediately begins controlling Baxter without prior EEG experience.

A. Signal Classification Pipeline

The classification pipeline used to analyze the data has a pre-processing step where the raw signals are filtered and features are extracted. This is followed by a classification step where a learned classifier is applied to the processed EEG signals, yielding linear regression values. These regression values are subjected to a threshold, which was learned offline from the training data. The resulting binary decision is used to control the final reach of the Baxter robot. The same signal classification pipeline was applied to all experiment scenarios, including both open-loop and closed-loop settings. During offline analysis (using data from the open-loop condition) the frequency range, XDAWN filter order, number of electrodes, type of features, cost function, and decoder were optimized via 10 iterations of 10-fold cross-validation. The following subsections describe the steps for closed-loop online classification in more detail.

1) Pre-Processing: Starting at the feedback onset time, a buffer of 800 ms from all 48 EEG channels is filtered using a 4^{th} order Butterworth zero-phase filter with a passband of 1-80 Hz. The mean of all channels is then subtracted from each channel to remove noise common to all electrodes. Since the error-related activity is linked to central cortical areas, dimensionality reduction is then implemented by selecting only the 9 central electrode channels (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2). This channel selection procedure also helps remove electrodes likely to be contaminated with eye blink and muscle artifacts.

2) Feature Extraction: Features are extracted from the reduced 9 channels using two different pipelines.

Covariance and XDAWN filter: The mean of all correct (mean-correct) and of all incorrect (mean-incorrect) training trials were computed and then spatially filtered using a 5^{th} order XDAWN filter [16], [17] (this order provided the best performance in offline analysis). The two filtered mean trials are then appended to each individual training or testing trial to create augmented trials [18], [19]. The covariance of each augmented trial is computed and vectorized using a tangent space projection, yielding 190 features [18].

Correlation: Correlation indexes are computed on a per-channel basis between the 9 chosen electrode channels and the mean-correct and mean-incorrect trial traces obtained from training. For each channel and trial, the difference between the mean-incorrect and mean-correct correlation indexes is computed to yield 9 features per trial.

Classification is done using a vector of all 199 features. Different feature selection approaches were evaluated in offline analysis (such as Fisher Score, one-way ANOVA, and PCA) but no significant performance increase was found.

3) Classifier: During offline analysis, an Elastic Net with $\alpha = 0.5$ and $l1_{\text{ratio}} = 0.0002$ was implemented and trained. This trained classifier was then used during online classification to output linear regression values.

4) Threshold: The desired output is a binary indication of the ErrP presence, so a threshold is chosen for the regression values. A threshold was computed during the offline training that minimized a biased cost function: $Cost = \sqrt{0.7\,(1 - sensit.)^2 + 0.3\,(1 - specif.)^2}$. More emphasis is placed on detecting ErrPs since missing an error in a robotic task could break a process or injure a person.

5) Decision: Once the regression value is subjected to the threshold, the classifier outputs the final binary decision. A 0 indicates that no ErrP is present and the robot should maintain its initial choice, while a 1 indicates an ErrP is present and the robot should switch targets.

V. Results: Primary and Secondary Errors

Analysis of the data demonstrates existence of ErrP signals for the chosen robotic task in both closed-loop and open-loop scenarios. The classification pipeline is evaluated using both online performance and offline enhancements.

A. Existence of Error-Related Potentials for HRI Task

Figure 7 shows a representative ErrP signal detected in the closed-loop experiments. The mean of FCz electrode traces are shown when there was an error (red), and when there was no error (green). The dark black trace is the difference between them, and displays a negative peak occurring around 250 ms after feedback onset that agrees with theory [7] followed by a positive peak. Figure 8a shows detailed time and spatial evolution of the primary ErrP for one of the subjects, where the plots are time-locked to the feedback onset when Baxter first moves to indicate object selection. The top plot shows the presence of the signal across all 48 EEG electrodes. The bottom plots are a time-lapse of the signal location on the scalp, created by interpolating between electrode locations. The signal is concentrated at the center of the head, which agrees with ErrP theory [7].

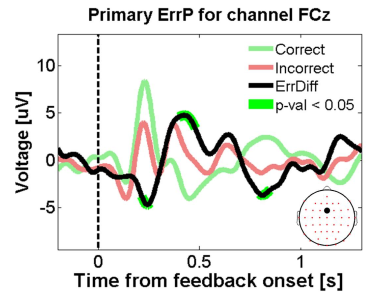


Fig. 7: This example ErrP was recorded during a closed-loop human-robot interaction. FCz location is shown as the black dot on the scalp, and green highlights show where correct and incorrect trials differ significantly (p-val $<$ 0.05, Wilcoxon signed rank test, corrected)

B. Closed-Loop ErrP: The Secondary Error

An important result is that allowing the human to influence the robot in real time via their brain activity enables secondary ErrPs, or *secondary errors*. These arise in response to system misclassifications, i.e. when Baxter does not correctly interpret the human's feedback. This occurs when the initial reach was correct but the system detects an ErrP, or when the initial reach was incorrect but the system fails to detect an ErrP - see cases C-I and I-I in Figure 3b, respectively. Since the human is actively engaged and aware that the robot should obey their feedback, secondary ErrPs are often stronger and easier to detect than the primary ErrP.

Figure 8b demonstrates the existence of the secondary ErrP in an analogous format to that described for the primary ErrP. Since no hardware switch is available for secondary feedback onset time, however, the signals are aligned using status codes sent from the experiment controller and a measured offset time from the primary switch liftoff. EEG activity is again centered towards the middle of the scalp.

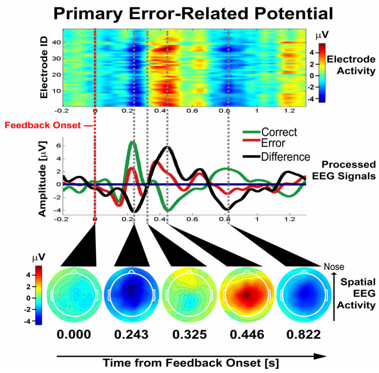
1) Impact on Classification Accuracy: Two stark differences between the primary and secondary errors are illustrated in Figure 9: the amplitude of the mean secondary ErrP (black) is about double that of the primary ErrP, and the shape of the secondary ErrP is more consistent across subjects. These differences can facilitate robust classification and lead to better overall performance and generalization.

C. Performance

This section presents the online closed-loop classification performance, as well as offline analysis of primary and secondary errors in open-loop and closed scenarios, to evaluate the success of the system in the chosen collaboration task.

Online closed-loop: EEG signals from the human observer are classified in real time and communicated to the robot. If an error is detected, the controller commands Baxter to switch targets before completing the reach. The classification decision, using a pre-trained classifier, is made within 10-30 ms after acquiring 800 ms of EEG data.

Offline closed-loop: The classifier is trained using EEG data from an entire online closed-loop session (4 blocks of

(a) 

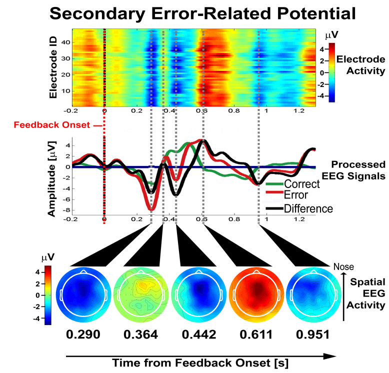
(b) 

Fig. 8: Primary (a) and secondary (b) ErrP signals evoked by the task from the same subject are shown evolving over time and location. The top row shows activity of all 48 electrodes, the middle row demonstrates mean FCz EEG activity, and the bottom row interpolates activity across the scalp. The ErrP signal is seen as early as 200 ms from the robot's movement.

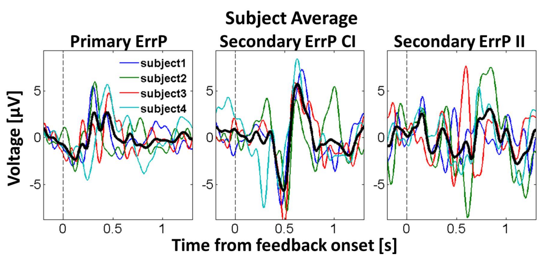


Fig. 9: Compared to primary ErrPs (left), secondary ErrPs after Baxter's correct selection (CI, center) demonstrate increased amplitude and decreased shape variability across subjects. Secondary ErrPs after Baxter's incorrect selection (II, right) are less common and therefore more variable. Black traces represent the mean trace across individual subject traces from the FCz electrode.

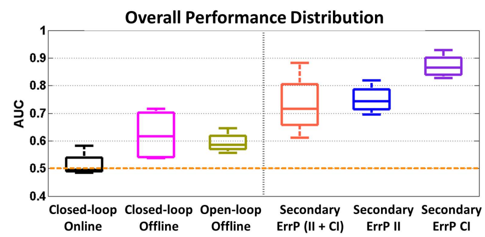


Fig. 10: Performance is summarized by the AUC score for different session types, where 1 is perfect classification.

TABLE I: Classification Performance

Session Type	Accuracy Mean	Accuracy Std. Dev.	Chance	% Above Chance
Closed-loop Offline	64.17	06.56	56.49	13.59
Open-loop Offline	65.06	01.75	58.91	10.43
Second. ErrP (II+CI)	73.99	07.64	58.16	27.21
Second. ErrP (II)	83.49	01.64	73.19	14.07
Second. ErrP (CI)	86.51	05.03	58.41	48.10

9 minutes each, totaling 190 trials on average). Features and thresholds are optimized, and the classifier is tested on a single-trial basis via 10 iterations of 10-fold cross-validation.
Offline open-loop: The classifier is trained using EEG data from an open-loop session (5 blocks of 9 minutes each); the human was a passive observer, and Baxter did not receive real-time feedback. The classifier is tested on a single-trial basis via 10 iterations of 10-fold cross-validation.
Offline secondary error: EEG data is pre-collected during an online closed-loop task. All EEG signal traces over five 9-minute blocks are used to train a classifier for the secondary ErrP that occurs after the first ErrP classification. The performance reported is of the classifier applied on a single-trial basis using 10-fold cross-validation.

Figure 10 presents the ROC AUC performance for each session type, which is an unbiased metric with 0.5 as chance. As expected, offline cases have better performance than online cases since the latter can only use data from previous blocks and thus have fewer training trials. However, even using all data from a subject (190 trials) offline represents a difficult task and achieves an AUC of 0.65.

The right three boxes of Figure 10 demonstrate the overall AUC performance gain achieved by using secondary errors, and Figure 11 separates the AUC results into individual subjects. The secondary errors significantly increase performance, to over 0.8 in CI cases. Since Baxter's initial choice is correct in 70% of trials, there are many more trials in this case than the II case so it is expected that the trained classifier can perform better for this case.

Table I summarizes performance in terms of accuracy, which measures how often Baxter ultimately makes the correct selection. Chance is determined by randomly shuffling trial labels. Notably, classification using secondary ErrPs boosts performance by over 20% in the CI case (the most common case). Figure 12 further breaks down the performance into True Positive Rates and False Positive Rates for each of the four session types. Offline analysis shows that classifying secondary ErrPs greatly increases true positive and true negative rates. Thus, secondary errors are promising for improving feedback between humans and robots.

VI. Conclusion and Future Work

By focusing on the detection of naturally occurring error-related potential signals, an online closed-loop EEG system has been developed that enables intuitive human-robot interaction even for general population subjects that have not been

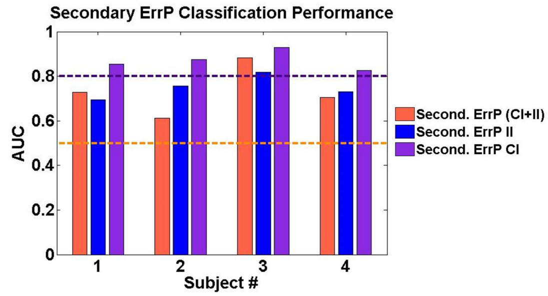


Fig. 11: Classifying secondary errors significantly raises AUC performance. All subjects perform well above chance, and AUC values over 0.8 are attained in CI cases (the most common case).

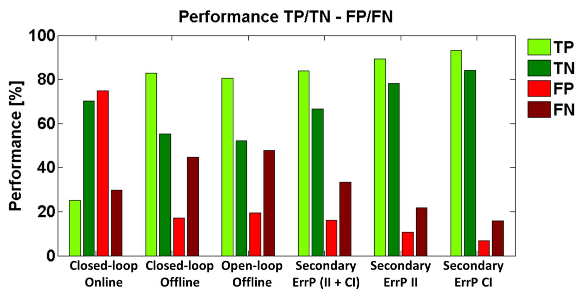


Fig. 12: Using secondary ErrPs in the classification loop greatly increases true positive and true negative classification rates.

previously trained on the task or EEG systems. The existence of ErrPs for a real-world robotic application is demonstrated, and a classification pipeline is developed to decode the EEG activity. Once trained offline using a small sample of open-loop trials, the pipeline can decode brain signals fast enough to be used online.

Offline analysis also demonstrates the existence of a secondary ErrP when the human observes that the robot has incorrectly interpreted their feedback. This signal is typically easier to classify than the original error, and can thus be used to improve the performance accuracy. In the future, this signal can be incorporated into the online scenario to boost closed-loop performance as well. This also suggests that new paradigms can be designed that exploit the secondary error. For example, the robot can perform motions that are designed to elicit error potentials from the human user to acquire feedback at crucial decision points, when choosing between many options, or even during a continuous trajectory.

In this way, the presented system moves closer towards the goal of creating a framework for intuitive human-robot interaction in real-world tasks.

ACKNOWLEDGMENTS

This work was funded in part by the Boeing Company, the NSF Graduate Research Fellowship number 1122374, and CELEST, a NSF Science of Learning Center (NSF SMA-0835976), for which the authors express thanks. Gratitude is also expressed towards Bianca Homberg, whose time and effort while laying a foundation for the Baxter interface are much appreciated.

REFERENCES

[1] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based braincomputer interface," *Journal of Neural Engineering*, vol. 10, no. 4, p. 046003, 2013.

[2] I. Iturrate, J. M. Antelis, A. Kubler, and J. Minguez, "A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation," *IEEE Transactions on Robotics*, vol. 25, no. 3, pp. 614–627, June 2009.

[3] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based braincomputer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, 2008, brain-Computer Interfaces (BCIs).

[4] L. Acqualagna, M. S. Treder, M. Schreuder, and B. Blankertz, "A novel brain-computer interface based on the rapid serial visual presentation paradigm," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug 2010, pp. 2686–2689.

[5] C. Guger, B. Allison, B. Grosswindhager, R. Prckl, C. Hintermller, C. Kapeller, M. Bruckner, G. Krausz, and G. Edlinger, "How many people could use an SSVEP BCI?" *Frontiers in Neuroscience*, vol. 6, p. 169, 2012.

[6] R. Ramli, H. Arof, F. Ibrahim, N. Mokhtar, and M. Y. I. Idris, "Using finite state machine and a hybrid of EEG signal and EOG artifacts for an asynchronous wheelchair navigation," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2451–2463, 2015.

[7] 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, Aug 2010.

[8] R. Chavarriaga, P. W. Ferrez, and J. d. R. Millán, "To err is human: Learning from error potentials in brain-computer interfaces," in *Proceedings of the International Conference on Cognitive Neurodynamics. ICCN 2007. Advances in Cognitive Neurodynamics*, 2008, pp. 777–782.

[9] P. W. Ferrez and J. d. R. Millán, "You are wrong!: Automatic detection of interaction errors from brain waves," in *Proceedings of the 19th International Joint Conference on Artificial intelligence*, 2005.

[10] 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, Aug 2010.

[11] S. K. Kim and E. A. Kirchner, "Classifier transferability in the detection of error-related potentials from observation to interaction," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2013, pp. 3360–3365.

[12] I. Iturrate, L. Montesano, and J. Minguez, "Single trial recognition of error-related potentials during observation of robot operation," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug 2010, pp. 4181–4184.

[13] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. d. R. Millán, "Latency correction of error potentials between different experiments reduces calibration time for single-trial classification," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2012, pp. 3288–3291.

[14] H. Zhang, R. Chavarriaga, Z. Khaliliardali, L. Gheorghe, I. Iturrate, and J. d R Millán, "EEG-based decoding of error-related brain activity in a real-world driving task," *Journal of Neural Engineering*, vol. 12, no. 6, p. 066028, 2015.

[15] Guger technologies. [Online]. Available: http://www.gtec.at/

[16] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: Application to brain-computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, Aug 2009.

[17] B. Rivet, H. Cecotti, A. Souloumiac, E. Maby, and J. Mattout, "Theoretical analysis of xDAWN algorithm: Application to an efficient sensor selection in a P300 BCI," in *2011 19th European Signal Processing Conference*, Aug 2011, pp. 1382–1386.

[18] "Classification of covariance matrices using a riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013, Advances in artificial neural networks, machine learning, and computational intelligence - Selected papers from the 20th European Symposium on Artificial Neural Networks (ESANN 2012).

[19] A. Barachant and S. Bonnet, "Channel selection procedure using riemannian distance for BCI applications," in *2011 5th Int. IEEE/EMBS Conference on Neural Engineering*, April 2011, pp. 348–351.