

Error Related Negativity in Observing Interactive Tasks

Chi Thanh Vi, Izdihar Jamil, David Coyle, Sriram Subramanian

Department of Computer Science, University of Bristol, UK.

{Chi.Vi; David.Coyle; Sriram.Subramanian}@bristol.ac.uk

ABSTRACT

Error Related Negativity is triggered when a user either makes a mistake or the application behaves differently from their expectation. It can also appear while observing another user making a mistake. This paper investigates ERN in collaborative settings where observing another user (the executer) perform a task is typical and then explores its applicability to HCI. We first show that ERN can be detected on signals captured by commodity EEG headsets like an Emotiv headset when observing another person perform a typical multiple-choice reaction time task. We then investigate the anticipation effects by detecting ERN in the time interval when an executer is reaching towards an answer. We show that we can detect this signal with both a clinical EEG device and with an Emotiv headset. Our results show that online single trial detection is possible using both headsets during tasks that are typical of collaborative interactive applications. However there is a trade-off between the detection speed and the quality/prices of the headsets. Based on the results, we discuss and present several HCI scenarios for use of ERN in observing tasks and collaborative settings.

Author Keywords

EEG; BCI; Error Related Negativity; ERN, Tabletop

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces - Graphical user interfaces;

INTRODUCTION

Off-the-shelf Electroencephalogram (EEG) headsets are becoming widely available at affordable prices and improving signal quality enabling a growing number of interactive applications for competitive gaming [13] or task classification [16]. Typically, EEG studies focus on Event Related Potential (ERP) which is the time- and phase-locked brain response following an event. A popular example is P300, a positive signal that is elicited about 300ms after the process of decision making. It has been widely used to recognize an intended selection of a speller [15] or an object on a multi-touch surface [27].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.
Copyright © 2014 ACM 978-1-4503-2473-1/14/04..\$15.00.
<http://dx.doi.org/10.1145/2556288.2557015>

Error Related Negativity (ERN) is another form of ERP that can be triggered in the brain when a user is aware of the obvious error(s) or confused about the last decision made in a time-critical task or the application behaves differently from their expectation [4, 8]. For example, ERN would be produced when pressing the LEFT key while intending to press the RIGHT key in a time critical multiple choice task. Previous studies have shown that ERN signals usually appear and peak within 150ms of the committed action [9]. Often ERN is detected offline using an averaging of EEG signals around the event onset from multiple trials (e.g. [4, 18]). However, recent studies demonstrate that ERN can be detected online and on a single-trial with expensive, clinical headsets (e.g. [5, 8]) and commodity headset (e.g. [23]). This opens ERN's use in interactive applications where the system can provide assistance once it can detect that the user is aware of their accidental action (i.e. in gaming or spatial navigation).

To date most examples of ERN have focused on detecting the signal by recording signals from the executer of the task. Recent endeavors [6, 19, 22] suggest that ERN signals even appear while observing another user making errors. These studies confirm the existence of negative potential within 250ms after the event onset and analysis of the signals' origin confirm that they are ERNs. This is useful in scenarios where the executer is not aware of their error but that error is spotted by an observer or supervisor. In current studies, the observer did not have much opportunity to anticipate the executer's actions and often relied on the answer displayed. In HCI scenarios, the executer's actions may become visible to the observer even before the action is committed. For example, in collaborative tabletops, an observer has awareness of an executer's actions through their reaching gesture [21].

We investigate the effect of anticipation in an observing task where the outcome of an action is revealed before that action is committed which has not been investigated with ERN in an observing task before. This can have many applications in pair-programming, collaborative tabletops and emergency response applications. We first investigate these anticipation effects with an expensive clinical EEG system to demonstrate that the signal is present and can be detected. Following this we investigate if an Emotiv headset has the similar detection capability.

We start by repeating an experiment from van Schie et al. [22] with an Emotiv headset to demonstrate that Observer ERN can be detected in a single trial and online basis, with the accuracy up to 64%, using a commodity headset.

After that, we investigated the anticipation effect in observing tasks when one person observes another person committing errors. Our results show that there are ERN-like patterns detected in the observer's EEG about 368ms after the initial movements happened and 55ms before the errors are committed. Following this result, we then show that the Emotiv headset can capture these patterns in the same experiment settings. Finally, we discuss the implications of our results on interactive applications.

The contributions of this paper are: (a) we demonstrate that off-the-shelf EEG devices like an Emotiv EEG headset can capture ERN in an observing task from channels in the frontal-central part of the brain; (b) we investigate the anticipation effects in collaborative settings demonstrated by the ERN detected in an observer's EEG signals before the action is committed; (c) through a final experiment we show that these anticipation effects can still be demonstrated using off-the-shelf EEG devices such as the Emotiv EEG headset.

RELATED WORK

EEG and Passive Brain Computer Interface

In the field of physiological computing, EEG is widely chosen because of its high temporal resolution, ease of use and the reasonably low cost. EEG-based BCI system for healthy users has been gaining interest because of its reliability and usability; opening them for creating new types of applications [29]. Because of this, passive BCI (beside active BCI and reactive BCI) was proposed as one type of new BCI systems [28, 29]. It can be defined as a BCI system that processes neural activity arising from users' involuntary control. An example of this is ERN, an error-potential used to correct a user's erroneous action.

Error Related Negativity (ERN)

ERN is a form of an ERP triggered in the brain when a person makes a mistake or the application behaves differently from their expectation [4]. This pattern is produced in a person's brain when they are aware of the obvious error(s) that they have made; either through system feedback or individual realization [8]. ERN peaks within 150ms after the action onset and has amplitude varied in accordance with the awareness of the mistake. Interestingly, ERN also appears when the user is confused about their last decision [4]. This pattern has been discovered by researchers using both expensive devices [4, 8] and commodity devices (such as Emotiv) [23]. ERN is useful in interactive applications when it is detected immediately following the triggered moment. To this end, researchers have looked at detecting this pattern in real-time, and on a single trial basis with accuracy up to 80% [3, 8, 20]. Vi and Subramanian [23] showed that this detection can be done using a low-cost and off-the-shelf headset like Emotiv with around 65% accuracy. This makes ERN more accessible to game developers and other consumer application designers.

Error Related Negativity (ERN) in Observing Task

Recent endeavors [1, 6, 19, 22] suggest that ERN signals even appear while observing another user making errors. For example, Miltner et al. [19] reported the generation of an ERN potential when participants observed errors committed in a choice reaction time (RT) task. Moreover, van Schie et al. [22] reported an ERN pattern elicited in an observer when observing another person performing a modified Eriksen flanker task [7]. Other researches have confirmed these results by showing observed ERN following observations of other's errors [1, 19]. These results suggested that similar neural processes trigger the detection of a person's own error as well as the detection of others. Interestingly, de Bruijn et al. [6] aimed to disentangle the dependency of ERN on error or reward. Their results show that the performance monitoring as reflected by the ERN is error-specific and not directly dependent on reward. As a result, our study focuses on the error awareness of the observers.

To date, all studies of ERN in observation tasks only look at situations where the outcome of the action becomes visible around the moment where the action is committed (e.g. in [22]). Furthermore, these studies focus on detecting the ERN pattern using averaging methods for the purpose of confirming the pattern's existence. We are not aware of studies that investigate the context where the observer can anticipate the executor's actions through their gestures towards the target.

In these contexts, anticipation effects can cause the observer to trigger an ERN before the moment when the executor finishes the action. Successfully detecting this pattern can be useful in many HCI applications such as collaborative tabletops, pair-programming and emergency response scenarios.

EXPERIMENT 1: ERIKSEN FLANKER TASK ON PC

The purpose of this experiment is twofold: first to validate, based on the study in [22] and [6] that we can detect the ERN signal on an observer using off the shelf brain sensing technology; second to verify that the ERN pattern can be detected using the classifier presented in [23] using an off the shelf headset.

Task and Procedure

Each participant performed the experiment paired with an actor. We used an actor in all of our experiments to trigger the ERN signal on the participant. Only the participants wore an Emotiv EEG headset which has the ability to capture EEG signals of 14 channels in the 10-20 international system. Participants were told about the experimental goals to explain why they had to wear the EEG headset and not the actor. Participants were not aware of the actor's role and were led to believe that the actor is another participant. As the window of recognizing the ERN signal is very small, the role of the actor is to perform certain actions that will maximize the ERN detection from the observer during the task period.

The experimental procedure is similar to the task carried out in [4] and [7]. To mimic the experimental conditions for a typical Flanker task experiment and reduce noise in data collection (such as in [4]), both the actor and observer were seated side by side in front of a screen in a dimly lit room. Participants were told to sit comfortably and minimize body, facial, and eye movement as well as to blink as infrequently as possible during the task.



Figure 1. Observation of Flanker Task with the executor (left) and the observer (right)

Trial: Each trial began with a black screen for 3s, followed by a fixation dot in the center of the screen for 200ms. After that, the screen remains clear for 200ms before one of four stimuli was displayed for 300ms. There were two types of arrows, each type had two stimuli: congruent stimuli (<<<<<< and >>>>>>) and incongruent stimuli (<<>><< and >><<>>). All four stimuli were used in our trials in random order. At a viewing distance of about 100cm, the visual angle of the arrow stimuli was 0.4° vertically and 0.6° horizontally, and between them was 0.3° space. The executor was asked to press the direction key as soon as they saw the stimulus to indicate the direction of the middle arrow. The input device used for this key press was a remote control (Genius Media Pointer). The pressed key (< or >) was displayed for 1s afterward. Note that the executor can be either the participant or the actor depending on the trial condition, as explained below.

The observer was then asked to rate the correctness of their last action by choosing one of three options: 1. *Sure correct*, 2. *Do Not Know*, and 3. *Sure Incorrect*. This was input through a mini keypad.

Practice trials: Participants were asked to perform a version of the Flanker task where they had to press one of two keys to specify the direction of a central arrow that was bounded by flanker arrows. This was done to make the participant familiar with the task that will be performed by the actor giving them an idea of what to expect as quick as possible when the stimuli appears.

Each participant performed a practice block of 40 trials where they had to press the direction key and rate the correctness of their action in each trial. During this time, the actor sat beside the participant and played a dormant role. In these trials, the directional arrows, which appeared on the screen after participants pressed a button,

were exactly like the pressed button. We did not collect EEG signals or any data during practice block.

Experimental Trial: After finishing the practice block, the participant and the actor switched places. The actor now had to press the directional keys and the participant (the observer) rated the actor's last performance. However, the actor hid her hands under a box to prevent the observer from seeing which button was physically pressed (see Figure 1). The observer was instructed to focus on the screen to judge the correctness of the actor's answers. As in [1, 19] participants were instructed to silently count the number of incorrect selection. However, the result displayed on the screen was independent of the button press: 40% of the time the computer displayed a wrong answer (the 'pressed button' is opposite of the middle arrow of the stimulus). This allowed us precise control of the experiment settings.

The actor performed 4 blocks of 50 trials each with 3 minutes break after each block. We collected 200 trials from each participants of the experiment. We recruited 6 university students, aged between 21 and 29 years old to give us a total of 1200 trials for this experiment.

Data Collection and Analysis

The EEG signals were captured from the Emotiv headset. Although Emotiv is an off-the-shelf headset, its ability of capturing EEG signals were validated by previous studies such as in [23, 24]. EEG signals were divided into 2s length epochs; 1s before and 1s after the key press moment. With the sampling frequency at 128Hz the length of each epoch is 256 samples. The first 200ms (about 25 samples) of each epoch were used to remove DC offset following which all epochs were filtered in 3-8Hz to remove components that are not in the ERN frequency bands of that particular epoch. Our chosen band ([3 – 8Hz]) was narrower compared to [22] ([1-10Hz]) but still keeps valuable information about ERN that is at least in part in the θ -freq band (4-7Hz) [17].

We then employed the classifying method described in [3] and validated in [20, 23] to analyze the data. It can be summarized as follow. For each channel, we used the holdout method where half of the trials (100 trials) were used for training via a logistic regression technique and other half were used for testing. We performed a t-test on the classifier output for each channel per user to check if there was a significant difference between two types of output (correct and incorrect epochs). Here, there were two categories to classify: correct or incorrect. Epochs were divided as correct and incorrect trials based on the participants ratings. “Do not know” rated epochs were removed from the analysis (~5% of 1200 trials in total).

The classifier performed the classification on a single trial basis. The result then was compared with the ground truth of each trial. This holdout method has advantages of large training and testing datasets, and fast processing time. It

also mimicked an online working version where a coefficient matrix is extracted for each user through performing a training block (the first consecutive half number of trials), this is used for detecting ERN pattern online (the second consecutive half). To minimize the bias of this method, no samples of the first half was used in the second half. Also in the training set, the number of correct trial was randomly picked so that it equals the number of the incorrect trials. This classification model was trained blindly (with no estimation of the expected accuracy) and test it on an unseen, fixed test set. As this approach was not repeated for a set of parameters, it should not cause any over-fitting effect.

Result

Using the above method, our results reveal significant differences in classification rates (for correct and incorrect epochs) for FC5, F7, and F3 channels. The classification rate for the rest of the channels was not significantly better than chance.

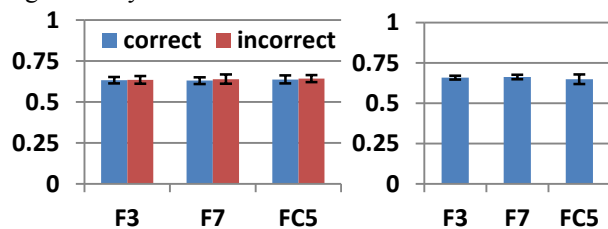


Figure 2. Mean classifier accuracy (left) and AUC (right) for different channels with error bars.

Figure 2 shows the average accuracy of three sensing channels. These three channels correspond to the frontal lobe which is in line with the literature about the origin of ERN (the Anterior Cingulate Cortex - ACC) [26]. Figure 3 illustrates the average EEG signals at F3 and F7 over all epochs belonging to two cases: correct and incorrect. Here, the averaged EEG signals for correct and incorrect over all trials and all participants are displayed in blue and red, respectively. For the incorrect trials (the red curve), we can see an ERN-like pattern (a negative peak followed by a positive peak) at around 250ms after the actor’s committed action (the black vertical line). Note that as a convention, ERPs are usually plotted upside down.

We did a Receiver Operating Characteristic (ROC) Analysis to investigate the efficiency of the classifier by evaluating its discriminating power. ROC analysis uses two distinct inputs: hit rate (or true positive rate) and false alarm rate (or false positive rate) as two separate performance measures. Figure 4 shows ROC curves for F3 and F7 channels. The further the curve is from the diagonal line, the more effective the classifier. The area under the curve (AUC) gives an indication of the performance of the classifier. An AUC of 1 indicates a perfect classifier and 0.5 indicates a random chance of classification. Our classifier achieves an averaged value of 0.66 over these three channels (F3, F7, and FC5). Their AUC are shown in Figure 2 (right).

Discussion

This study demonstrates that Emotiv EEG headset is capable of capturing EEG signals with sufficient quality for a classifier to be able to detect ERN pattern with an accuracy of about 65%. The characteristic of detected ERN patterns is similar to previous studies using more expensive devices [22]. Our results further demonstrate that the classifier described in [3] and validated in [20, 23] can be used to detect ERN in observer tasks. This means an interactive application can have a similar classifier detecting both observer and executer ERNs with minimal modification to the software.

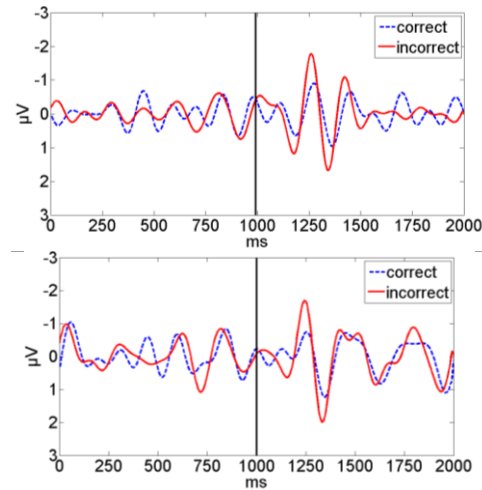


Figure 3. Average EEG signals at channel F3 (top) and F7 (bottom).

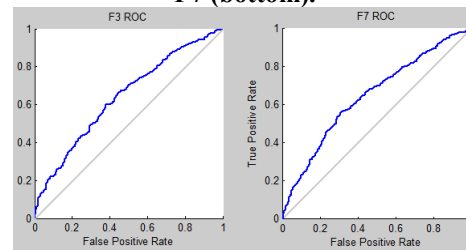


Figure 4. ROC curves for channels F3 and F7

One example of interactive tasks that can benefit from our results is in pair programming where two programmers work together as partner, on the same machine, to complete a programming work with their roles switched frequently [25]. One programmer is the driver, who performs all “on computer” tasks. The other is the observer or navigator who reviews each line of code and points out the errors as it is being written. Usually this error correction process requires the observer to interrupt the programming process, point out the location of the errors in his opinion either using his hand or using the mouse/keyboard that the driver is controlling.

However, our experiment results suggest that monitoring the appearance of ERN in the observer’s EEG can speed up this error correction process. Each action of the driver is considered a trial of a multiple choice RT task. If the

ERN is detected in the observer's brain, it can be used to pinpoint the whereabouts of the error (e.g. highlighting the code section) which is 250ms after the action's committed moment. The system can also provide suitable suggestions based on the context of the location where the error was triggered. This can happen either with or without the error awareness of the executer.

In addition, we believe the detection accuracy can be enhanced by improving the visibility and awareness of the performer's actions through better visualization. This will elicit higher amplitude of ERN which can improve the classification rate. More detailed implications of this study to HCI are discussed at the end of this paper.

OBERSEVER ERN IN A TABLETOP TASK

In the first study the participant did not have any opportunity to see the actions of the actor and often relied on the answer displayed. In collaborative HCI scenarios like around a tabletop, a participant can usually anticipate the executer's actions through their gestures which reveal the outcome of an action before that action is committed.

The awareness caused by anticipation gives the observer more time to form an opinion on the action. This could potentially reduce the time-critical aspect of the ERN leading to a low signal quality. Alternatively, the observer may reach an opinion of the executer's action as soon as they see the initial cues. In this case, we may detect a good quality ERN in the observer well before the executer has even completed the action.

We are not aware of any experimental investigation of the effect of anticipation in an observing task where the outcome of an action is revealed before that action is committed. Therefore we extended the duration between start and committed moments of the actor's action to investigate this effect. We also aimed to determine the moment when the observer elicits an ERN. Successfully detecting this opens a rich design space of interest to HCI such as collaborative tabletops and pair-programming.

Experiment Setup

Each experimental session involved two users – an executer and an observer seated around a rear-projected FTIR interactive table of height 76cm. The executer was an actor trained to do this study while the observer was our experiment participant. The executer and observer sat opposite to each other so that they were aware of each other's movements and actions (similar to the setup in [22]). The projection area of the table was 72cm x 48cm (resolution 1024x768 pixels) and touch detection was done through a Point-Grey Dragonfly 2 camera.

All participants performed the task with the same actor (the person at the top in Figure 5). However, participants were led to believe that the actor is just another participant like themselves. In order to study the effect of anticipation, we had two experimental conditions – a *close layout* and *far layout*.

In the *close layout* the actor's buttons (size 100x100 pixels each) were placed right under their hands and next to each other. The participant would struggle to see the actors hand movement limiting their ability to anticipate the outcome before the actor commits to the action. Any ERN elicited in the participant (observer) in this condition would be because of the committed action of the actor.

In the *far layout* the two confirmation buttons were placed close to the left and right edges of the table (200 pixel gaps to left and right edges with 524 pixels distance between them; 80 pixels gap from two buttons' centers to the edge of the actor's table side). The actor was asked to keep his hands touching two touch sensors placed on his side of the table (see the table's top edge in Figure 5, right). In order to select a response the actor had to lift one hand from the rest position and move towards the button. This action took approximately 400ms giving the participant sufficient time to anticipate the actor's actions. The actor's other hand was left in the rest position touching the sensor during this action.

Task and Procedure

Each participant wore the EEG cap of the BE-MRI System from EbNeuro. In this experiment, 19 electrodes were used (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2). This system has higher sampling frequency, larger coverage, and has access to Cz channel compare to Emotiv EEG headset. Participants observed and gave feedback about the correctness of the actor's performance in an Eriksen flanker task.

Each trial began with a black screen for 3s, followed by a fixation dot in the center of the screen for 200ms. After that, the screen remains clear for 200ms before one of four stimuli was displayed at the center of the table for 300ms. All four stimuli were used in a random order.

In both *close* and *far layouts* (Figure 5 left and right), the actor then touched one of two buttons to indicate the direction of the middle arrow. In *close layout* there is no noticeable hand movement whereas in the *far layout* the actor has to reach the button to make a selection.

After the answer button was selected, it was highlighted in red for 1s before turning back to the initial color. This is to re-enforce to the observer the actor's selection. The participant was then asked to rate the correctness of their last action by choosing one of three options: 1. Sure correct, 2. Do Not Know, and 3. Sure Incorrect. During the experiment, the observers were asked to minimize their body and facial movements as well as blink as infrequent as possible. They were also instructed to guess the outcome of each answer as quick as possible although they could input that answer at their own pace. As in [1, 19] participants were instructed to silently count the number of incorrect selection. Since it was not possible to control the ratio of correct and incorrect trials to the extent done in Experiment 1, the actor was asked to keep

the incorrect selection rate to between 25 and 30%. The actor received practice to minimize difference in action between correct and incorrect gestures.

Each participant performed 2 blocks of 80 trials each per layout. The order of presentation of the blocks was controlled using a Latin square to reduce order effects. Participants received 3mins break between blocks and the whole experiment took about 90mins per participants including about 40mins of setup time.

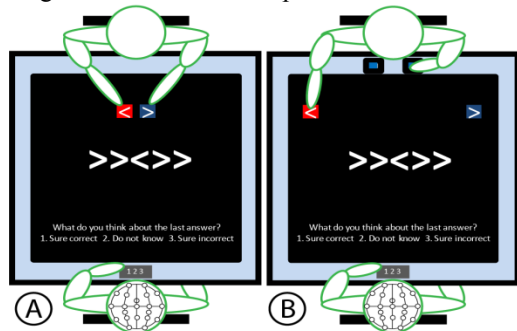


Figure 5. Tablet Flanker task a) close, b) far layout.

Ten participants (7 males) between the age of 19 and 31 volunteered for the study. All were from the local university and did not participate in the earlier experiment. All participants had normal or correct-to-normal vision, and none of them were color blind. Also no participant had undergone brain surgery or had any known neurological disorders. Participants wore EEG cap during the experiment. Participants received a financial compensation for their participation in the study.

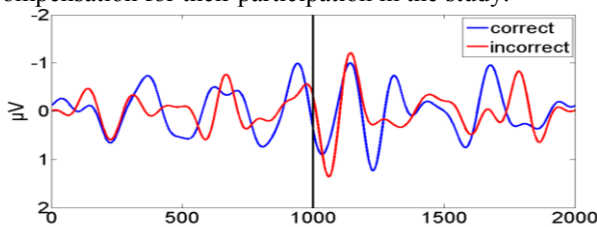


Figure 6. Averaged signals of the ‘close’ layout at Cz

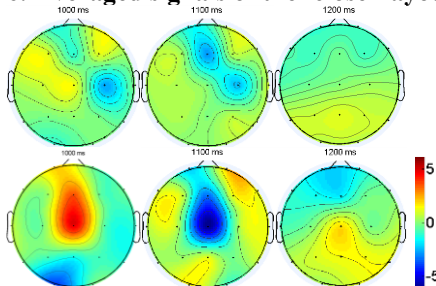


Figure 7. Topographic distribution of correct trials (top) and incorrect trials (bottom) in the ‘close’ layout at the intervals: key press (left), 100ms after (middle), and 200ms after (right).

Data Collection and Analysis

EEG signals were collected using the BE-MRI System from EbNeuro. The sampling frequency was set at 4 KHz

and electrode impedance was below 5 KΩ. EEG signals were then down sampled to 1 KHz before being divided into 2s epochs around the touching moment, which is also the moment that the results were displayed or the touched button was highlighted in red. It is also assumed to be the moment when participants became aware that the actor had or had not made a mistake. Epochs were averaged, filtered, and divided into groups as in prior experiment.

Result

We collected 3200 trials, 1600 each for close and far layout. The number of incorrect trials was 27% in the close layout and 30% in the far layout. 7 trials in the close layout and 30 in the far layout were rated as “Do Not Know” by the participant.

Close Layout Analysis: Figure 6 shows the averaged signals in Cz channel in the close layout. It can be seen that there is a difference, however small, between incorrect trials (red curve) and correct trials (blue curve). Latency of the highest peak after the touch button was pressed was 140ms. However, the topographic distribution (Figure 7) shows a clear difference between correct and incorrect trials around the time of the peak. It can be seen that there is a high density of negative EEG at the central and frontal area around the latency of the highest peak after the key press onset (Figure 7, middle).

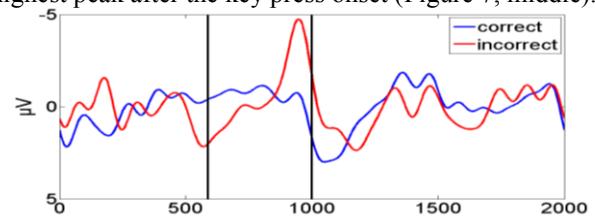


Figure 8. Averaged signals of the ‘far’ layout at Cz

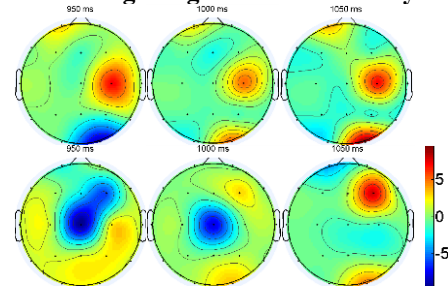


Figure 9. Topographic distribution of correct trials (top) and incorrect trials (bottom) in the ‘far’ layout at the intervals: 50ms before key press (left), key press (middle), and 50ms after key press (right).

Far Layout Analysis: In the far layout, the differences can be seen clearer with the averaged EEG signals in Figure 8. The left vertical line depicts the averaged moment that the actor lifted up his hand and started to reach to the answer button. The right vertical line depicts the moment that the actor pressed the answer button. It took the actor averagely 428ms (correct trials) and 423ms (incorrect trials) to touch the answer. There was no significant difference found between these two durations ($p > 0.05$).

The signals peak at 55ms before the answer buttons were touched. The topographic distribution (Figure 9) shows a clear difference (in the density of negative EEG at the frontal and central area) between correct and incorrect trials at the time of the peak (~ 50ms before button touch, left figures) and at the time of the touch (middle figures).

Using the same classifying method described in the previous experiment, we found that several channels in the frontal-central area of the brain yield successful classification. More details are in depicted in Figure 10. The highest classification rates are from channel Cz for both *close layout* (correct/incorrect = 70.43%/68.71%) and *far layout* (correct/incorrect = 73.02%/72.40%).

We also calculated the AUC for the above channels. Their values are plotted in Figure 11. Overall, the classifier achieved averaged AUC values of 0.6733 for the *close layout* and 0.6602 for the *far layout*.

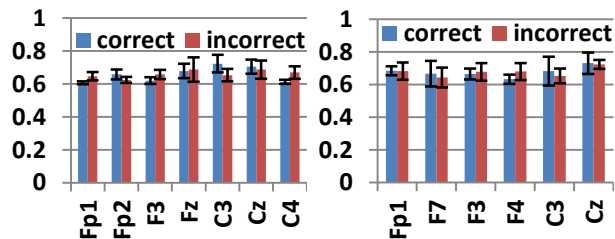


Figure 10. Classification rates of *close layout* (left) and *far layout* (right) with error bars

Discussion

The result of this experiment demonstrates that ERN can be detected in an observer in both hand layouts. It also shows that there is an anticipation effect where the ERN pattern in the observer’s mind appears about 55ms before the touching action is committed.

The experiment mimicked a classic Flanker task to keep our experimental settings in line with other research in this area. Interactive applications can trigger ERN using a similar paradigm. It may be possible to detect the ERN pattern earlier by making the actor’s movement clearer to the observers that can trigger the awareness sooner.

It is worth noting that the classifier used is a light-weight module which can provide an output in less than 1ms for a 1kHz sample once the β coefficient matrix is known for that participant (tested in Matlab @R2013a 64 bit, Intel® Core™ i3 CPU 3.10GHz, 8GB RAM). This means any interactive application could act on an executor’s action based on the ERN of an observer. For example, in collaborative tabletops a user could be asked for a stronger confirmation if the system detects an ERN in the observer. This could also benefit peer-learning activities where the observer and executor can constantly switch roles to learn from each other.

A drawback of this study is that the EEG cap is an expensive clinical system that is cumbersome to wear, tethered and does not lend itself to applications that

require the observer to move freely within the experimental environment. This further limits the possibility of both the observer and the executor wearing the cap in a realistic environment. The following experiment aims to address this drawback.

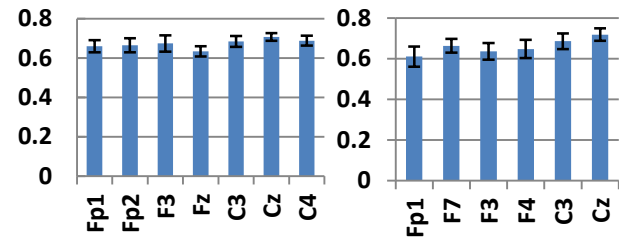


Figure 11. AUC for *close* (left) and *far layouts* (right) with error bars

TABLETOP OBSERVER ERN WITH EMOTIV

The purpose of this experiment is to see if the Emotiv EEG headset can be used to detect the same pattern as in experiment 2. This would then open up the use of observer ERN to a wide range of interactive applications by detecting and predicting errors in observers’ mind.

Task and Procedure

The experimental setup, task and procedure were identical to the previous experiment. We recruited 11 participants (8 males), aged between 20 and 31 years old. None of them took part in previous studies. The experiment took about 60mins per participant. This shorter experiment time, compared to the previous experiment, was due to the shorter setup time needed for Emotiv headset.

As with experiment 2, we collected 320 trials per participants yielding a total of 3520 trials from all participants of the experiment for both *close* and *far layouts* with 1760 trials each. The number of incorrect trials was 26% in the *close layout* and 28% in the *far layout*. 19 trials in the *close layout* and 34 in the *far layout* were rated as “Do Not Know” by the participant.

Data Collection and Analysis

Signals were collected at a sampling frequency of 128 Hz using the Emotiv EEG Neuroheadset. They were then divided into 2s epochs around the touching moments, which is also the moment that the results were displayed. The analysis was the same as in previous experiment.

Result

Close layout: Figure 12 shows average signals at channel AF3 in the *close layout* over all participants and all trials. It can be seen that there is a difference between incorrect trials (red curve) and correct trials (blue curve) at the latency of 188ms after the moments where the button answers were pressed. This pattern has a negative peak appeared within the time period of 250ms after the event onset, similar to previous studies such as in [6, 22].

Far layout: Figure 13 shows the averaged EEG signals for the *far layout* over all participants and all trials. There is also an ERN-like pattern which peaks at about 3ms

after the answer button was touched. This indicates that if this is the ERN pattern, the observer must have the judgment about the actor's action's correctness about 250ms before which also is an indication that the observers have made the decisions before the answer buttons were touched.

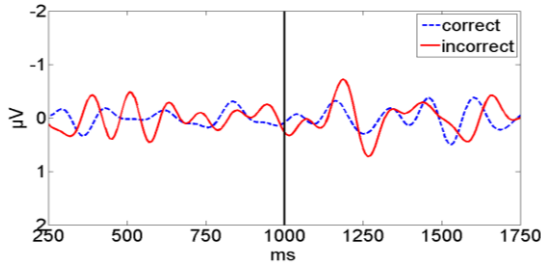


Figure 12. Averaged signals of the close layout at AF3. The vertical line is the button touched moments

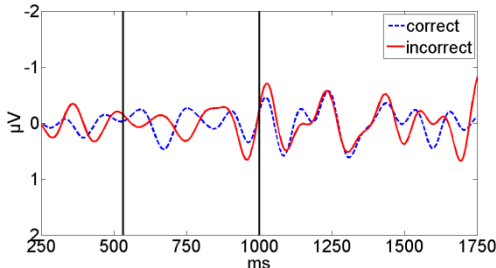


Figure 13. Average signals of the far layout at F3. The left vertical line is the averaged hand lift-off moments, the right vertical line is the button touched moments.

Between Figure 6 & Figure 12 and Figure 8 & Figure 13, we observe similar ERN-like patterns (negative peaks 140ms after touch moments in Figure 6; 188ms after touch moments in Figure 12; 50ms before touch moment in Figure 8; 3ms after touch moment in Figure 13).

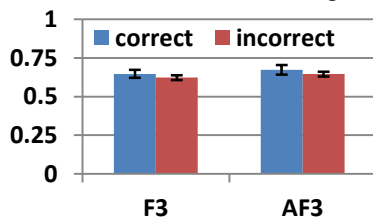


Figure 14. Classification rates with error bars

Using the same classifying method described in the first experiment on the collected signals from Emotiv device, we found that in close layout, channel AF3 yields the classifying results with 67.32% of correct trials were classified as correct and 64.57% of incorrect trials were classified as incorrect. Additionally, in far layout, channel F3 yields the successful classification with classification rates for correct and incorrect trials are 65.03% and 62.30% respectively (see Figure 14).

Additionally, we calculated the area under the curve (AUC) values for channels AF3 and F3 to justify how well the classifier performs on these two channels. Our obtained results show that AUC for channels AF3 and F3 were 0.6483 and 0.6486 respectively.

Discussion

The results from this experiment show that it is possible that ERN can be observed in collaborative tasks from the observers' EEG. Although the channels that show the detection are limited to AF3 (*close layout*) and F3 (*far layout*), it is an indication that an off-the-shelf EEG headset such as Emotiv is capable of detecting this pattern. As a result, other commodity headsets which have access to these channels can also harness the advantage of detecting ERN pattern in observing tasks.

There is a shift in latency of the ERN peaks between experiments 2 and 3. The latency differences of about 40ms (*close layout*) and 60ms (*far layout*) are due to the devices used in each experiment; EbNeuro BE-MRI system (~USD 10k) and Emotiv headset (~USD 300). We can postulate many reasons for the difference:

- The buffer used in each device is different. While Emotiv headset waits for the buffer to be filled before pushing signals toward the receiving Bluetooth dongle, EbNeuro pushes the signals continuously for each collected data sample. As the process of continuously pushing signals requires much more expensive Digital Signal Processing (DSP) module, this reflects the price difference between the two devices. Furthermore using Bluetooth to transfer might add to the latency.
- While EbNeuro pushes signals continuously at the original sampling frequency (4KHz), Emotiv samples EEG signals internally at 2KHz then down samples it to 128Hz. This can introduce latency.

From the experiment's results, it can be seen that there is a trade-off between how early the ERN pattern in observing tasks can be observed vs. the prices and quality of EEG headsets. Although the differences are small (from about 40ms to 60ms), they should be considered carefully to fit with the goals of each interactive task.

DISCUSSION

In addition to the benefit of correcting one's own errors, ERN has the potential to enrich interactive applications in the collaborative working setting. The results from our studies can provide guidance on how best to begin harnessing ERN for such interactive experiences. From our results, HCI designers can employ ERN in interactive tasks to pinpoint the executor's error's whereabouts, which can be from 50ms before to 250ms following the triggered action, depend on the interaction technique.

The usefulness of executor' and observer' ERNs in interactive tasks depends on the usage context and the designer's creativity in making use of it. Here we suggest some applications to highlight design possibilities.

Collaborative tabletop settings

Groupware applications use embodiments to help people stay aware of the presence, location and movement of users in collaborative tasks [11]. They can either be real embodiments which make use of the actual physical body

of the user or virtual embodiments which are digital representations of users (e.g. Telepointers [10]).

In these settings, real and virtual embodiments both provide obvious information of awareness to other group users. Therefore, an observer can judge quickly about the correctness of the action being performed. However, this information is not accessible unless the observers provide feedback to the performer or the system. This is a time consuming process and can interrupt the performance.

Our results show that an integrated ERN detection module can be used to monitor an ERN's appearance in the observers. As the classification time is about 1ms with the coefficient matrix known, the detected ERN can be used to trace back the time period where the observer thinks the performing user made a mistake or is confused about their action. Although these time periods range from 50ms before to 250ms following the committed action, careful task design can help to shorten this range to specific time points. This is due to different interaction techniques providing different awareness to the observers leading to different ERN triggered moments. Therefore, HCI designers should design their applications to maximize the in order to trigger clearer ERNs.

As a simple example, an action of a performer leading to an observer ERN can result in an ignorable pop-up menu that can help revert or discuss based on the source of ERN in the observer. Rather than merely automate the action's reversion, ERN can serve as an information point for discussion and clarification between the collaborators. The pop-up menu can then provide contextually relevant information including an "undo" item.

The goal is to maximize group awareness of individual's actions without disrupting the work-flow of the task. Further contextual studies can explore this trade-off.

War-rooms / Emergency response

In emergency scenarios, it is crucial to give instructions and make decisions. ERN can be integrated into the Emergency Management Information System [14] so that each decision that has been made can be crossed checked by the observers/ supervisors in the same team. If the executer is made aware that other teammates thinks a mistake is about to be made by the performing action, it can be avoided before it is made. Moreover, if the mistake is already committed, the system can sense if an ERN appears in an observer's brain to speed up the correction.

Gaming scenarios

Usually in shooting games and MMORPGs gamers team up to act against other teams. Here, every decision needs to be precise and made in a timely manner. ERN could be integrated to provide a new tool to team members to observe and react to each other's actions. Detected ERNs from teammates can be combined and displayed as feedback to every continuous action say in the form of an overlay window. This can give gamers an idea of where

and when they committed a mistake, which could add to the game tactics a new dimension. This ERN communicating channel, beside verbal and chatting ones, can be used to improve the teamwork and strategic analysis skills during the game hence make it more challenging between teams. However, the mechanism to trigger error-events has to follow an oddball paradigm in the context of a reaction time task.

Limitations of the classification method

Our classifier is based on a linear regression method described in [3] and validated in [20, 23] for single trial ERN detection. Applying this method provides classifying rates of up to 73% accuracy. [23] evaluated the usefulness of different ERN detection accuracies and showed that when integrating an ERN detection module in an interactive application a classifier with 65% accuracy is as good as a classifier with 80% accuracy. Moreover, if a system has very high accuracy (less error rate), it may promote hasty commitment to selections [12] and thereby increase the cost of recovery from an error.

In addition, our classifier requires input EEG signals $\pm 1s$ around the time-locked event which means that it can detect the ERN pattern 1s after the action. However it can be used to pinpoint the ERN's triggered moment which can be up to 50ms before the action. Despite the delay, this form of ERN detection can still be useful in many applications like in the earlier outlined case of Pair programming. A 1s delay will not affect the benefit of ERN in such activities especially when we can highlight the code-section that is in question. Additionally, other classifier methods can improve accuracy and reduce the window size of input signals leading to earlier detection of the ERN after it is triggered. For instance, a BCI competition on customized classifier increased classification rates of P300 up to 96.5% and motor imaginary up to 94.2% [2].

Moreover, the selection of channels is based on a set of t-tests on the output of the training set. Then, the test set was used to remove under-the-chance classification rate channels as a beta-correction method. A more robust method (e.g. bootstrapping) can be used here to reduce the chance of selecting randomly significant channels.

CONCLUSION

The experiments described in this paper offer some valuable guidelines for HCI designers. We show that ERN patterns can be detected in an observing task using an off-the-shelf EEG headset on a single trial basis. Moreover, we show the anticipation effects in collaborating work where ERN can be detected before a committed action in the observer's mind. We then extended our finding to a commodity headset to show that it can also detect the anticipation effect through the existence of ERN. In our discussions, we suggested novel ways in which HCI applications can benefit from ERN in collaborative and observing environments.

ACKNOWLEDGEMENTS

We thank Dr. Jade Thai (CRICBristol, University of Bristol) and Dr. Rafal Bogacz (University of Oxford) for their help in providing access to the EbNeuro equipment.

REFERENCES

1. Bates, A. T., Patel, T. P. and Liddle, P. F. External Behavior Monitoring Mirrors Internal Behavior Monitoring. *J. of Psychophysiology* 281-288. 2005.
2. Blankertz, B., Muller, K., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlogl, A., Pfurtscheller, G., Millan, J. R., Schroder, M. and Birbaumer, N. The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2006.
3. Christoforos, A., Dimitris, T., Niall, A. and David, H. Temporally adaptive estimation of logistic classifiers on data streams. *ADAC*, 3, 3 243-261. 2009.
4. Coles, M. K. S. a. M. G. H. Performance Monitoring in a Confusing World: Error-Related Brain Activity, Judgments of Response Accuracy, and Types of Errors. *Proc. of JEPHPP*, 26, 1 141-151. 2000.
5. Dal Seno, B., Matteucci, M. and Mainardi, L. Online detection of P300 and error potentials in a BCI speller. *Computational Intelligence and Neuroscience*. 2010.
6. De Bruijn, E. R. A. and von Rhein, D. T. Is your error my concern? An event-related potential study on own and observed error detection in cooperation and competition. *Frontiers in Neuroscience*, 6. 2012.
7. Eriksen, C. and Schultz, D. Information processing in visual search: A continuous flow conception and experimental results. *Attention, Perception & Psychophysics*, 25, 4 249-263. 1979.
8. Ferrez, P. W. and Millán, J. D. R. You are wrong!: automatic detection of interaction errors from brain waves. *Proc. of IJCAI*. 2005.
9. Ganushchak, L. Y. and Schiller, N. O. Brain error monitoring activity is affected by semantic relatedness: An event-related brain potentials study. *J. Cognitive Neuroscience*, 20, 5 927-940. 2008.
10. Greenberg, S., Gutwin, C. and Roseman, M. Semantic Telepointers for Groupware. *OZCHI*. 1996.
11. Gutwin, C. and Greenberg, S. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *CSCW*, 11, 3 411-446. 2002.
12. Gutwin, C. and Cockburn, A. Improving list revisitation with ListMaps. *AVI. ACM*. 2006.
13. Hjelm, S. I. and Browall, C. Brainball - using brain activity for cool competition. *NordiCHI*. 2000.
14. Kristensen, M., Kyng, M. and Palen, L. Participatory design in emergency medical service: designing for future practice. *CHI. ACM*. 2006.
15. Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M. and Wolpaw, J. R. Toward enhanced P300 speller performance. *Neuroscience Methods* 2008
16. Lee, J. C. and Tan, D. S. Using a low-cost electroencephalograph for task classification in HCI research. *UIST*. 2006.
17. Luu, P., Tucker, D. M. and Makeig, S. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *J. Clin Neurophysiol*, 115, 8 1821-1835. 2004.
18. Miltner, W. H. R., Braun, C. H. and Coles, M. G. H. Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a neural system for error detection. *Cognitive Neuroscience*. 1997.
19. Miltner, W. H. R., Brauer, J., Hecht, H., Trippe, R. and Coles, M. Parallel brain activity for self-generated and observed errors. 2004.
20. Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A. and Sajda, P. Linear Spatial Integration for Single-Trial Detection in Encephalography. *NeuroImage*, 17, 1 223-230. 2002.
21. Pinelle, D., Nacenta, M., Gutwin, C. and Stach, T. The effects of co-present embodiments on awareness and collaboration in tabletop groupware. *GI*. 2008.
22. van Schie, H. T., Mars, R. B., Coles, M. G. H. and Bekkering, H. Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, 7, 5 549-554. 2004.
23. Vi, C. and Subramanian, S. Detecting error-related negativity for interaction design. *CHI. ACM*. 2012.
24. Vi, C., Takashima, K., Yokoyama, H., Liu, G., Itoh, Y., Subramanian, S. and Kitamura, Y. D-FLIP: Dynamic and Flexible Interactive PhotoShow. *Springer*. 2013.
25. Williams, L. and Kessler, R. *Pair Programming Illuminated*. Addison-Wesley Longman. 2002.
26. Yeung, N. and Cohen, J. D. The Impact of Cognitive Deficits on Conflict Monitoring: Predictable Dissociations Between the Error-Related Negativity and N2. *Psychological Science*, 17164-171. 2006.
27. Yuksel, B. F., Donnerer, M., Tompkin, J. and Steed, A. A novel brain-computer interface using a multi-touch surface. *CHI*. 2010.
28. Zander, T., Kothe, C., Jatzev, S. and Gaertner, M. Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces. 2010.
29. Zander, T. O. and Kothe, C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8, 2. 2011.