Proceedings Paper:

Wirth, C. orcid.org/0000-0002-1800-0899, Lacey, E., Dockree, P. et al. (1 more author) (2018) Single-trial EEG classification of similar errors. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 17-21 Jul 2018, Honolulu, Hawaii. IEEE , pp. 1919-1922. ISBN 978-1-5386-3646-6

https://doi.org/10.1109/EMBC.2018.8512700

Single-Trial EEG Classification Of Similar Errors

Christopher Wirth*, Eric Lacey, Paul Dockree, and Mahnaz Arvaneh

***Abstract*— When humans recognise errors, either committed by themselves or observed, error-related potentials (ErrP) are produced in the brain. Recently, a few studies have shown that it is possible to differentiate between the ErrPs generated for errors of different direction, severity, or type (e.g. response errors, interaction errors). However, in real-world scenarios, errors cannot always be delineated by these metrics. As such, it is important to consider whether errors that are similar in all of the aforementioned aspects can be classified against each other on a single-trial basis. In this paper, for the first time, we consider two different response errors, which are of equal severity and have no associated direction. This study used electroencephalogram (EEG) data from a sustained-attention based time-critical reaction task, where time pressure caused subjects to commit two different errors. Using data from 16 subjects, we applied time domain EEG features and an ensemble of linear classifiers to separate these two error conditions on a single-trial basis. We achieved a mean balanced accuracy of 63.23% and, for most of these subjects, achieved statistically significant (p < 0.05) separation of the two error conditions. The ability to classify similar error conditions, such as these, increases the scope of possible applications for EEG error detection, and has the potential to improve brain-machine interaction.**

I. Introduction

Error potentials (ErrP) are produced in the brain when a human observes an error, or recognises that they have committed an error themselves [1]. These ErrPs can be detected in electroencephalogram (EEG) signals, and can be utilised as a part of a Brain-Computer Interface (BCI), either for immediate error correction [2], or as a feedback function for a reinforcement learning (RL) strategy [2], [3]. In the case of RL, a system can work effectively as long as classification exceeds chance level [2], [3].

Various different types of error condition are known to elicit ErrPs. For example, ErrPs caused by a human responding incorrectly in a time-critical reaction task have been referred to as "response ErrP" [4], [5]; Other types of ErrP that have been described in literature include "interaction ErrP", elicited when an action is not performed as expected by the computer with which the human is interacting; "feedback ErrP", when a human is told that they committed an error of which they were previously unaware; "observation ErrP", when a human observes an error committed by somebody else; and those elicited by "execution errors", when an action is not performed as expected; or by "outcome errors", when the desired outcome is not achieved [5]–[7].

Recently, a limited number of studies have shown that it is possible to use single trial EEG to differentiate ErrPs evoked by error conditions of different types [6], directions [8], and severities [8], [9]. However, in some tasks, errors will naturally exist that are the same "type" and severity as each other, and are either errors in the same direction, or have no associated direction at all. For example, consider a BCI with integrated image processing. A user wishes to pick up a green apple, but the BCI has to select the apple from an image that also contains some other fruits. First, the system selects an orange, but error classification tells it that this is the wrong colour. It then tries a pear but is told this is the wrong shape. These two errors could not be differentiated using existing metrics, but being able to categorize them could improve the effectiveness of the BCI's learning strategy. As such, we can see that classification of such similar error conditions opens up the possibility of RL being applied to a BCI for the performance of more potential tasks than ever before. There are also other potential applications for this kind of error classification, such as life-logging.

In this study, subjects were given a time-critical reaction task, requiring sustained attention, in which two different error conditions could occur. In one condition, subjects reacted erroneously to the presentation of a blue dot. In the other condition, they reacted erroneously to a dot that was identical to the previous stimulus. These error conditions were both directionless, and were of equal severity. As both conditions were the result of the subject performing the reaction task incorrectly, both would be classed as response errors. To our knowledge, no two error conditions that are similar in all of these aspects have previously been classified against each other using single-trial EEG.

To tackle this challenge, we proposed extracting a small number of highly discriminative time domain features from fronto-central channels, to provide a low-dimensional feature space. We then classified the data using a weighted vote of linear classifiers. The effectiveness of the proposed algorithm in classifying the two error conditions was evaluated using data collected from 16 healthy adults.

II. Methods

A. Experimental Design

Data for this investigation were taken from an Error Awareness Dot Task carried out at Trinity College Dublin. Subjects were asked to perform a Go/No-Go task, as shown in "Fig. 1". The subjects were shown succession of coloured dots on a screen, and asked to press a button when each

This research was supported by University of Sheffield, United Kingdom, and the Engineering and Physical Sciences Research Council (EPSRC). *Asterisk indicates corresponding author.*

C. Wirth and M. Arvaneh are with the Automatic Control and Systems Engineering Department, University of Sheffield, Sheffield, United Kingdom (emails: cwirth1@sheffield.ac.uk, m.arvaneh@sheffield.ac.uk).

E. Lacey and P. Dockree are with the Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland (emails: elacey@tcd.ie, dockreep@tcd.ie).

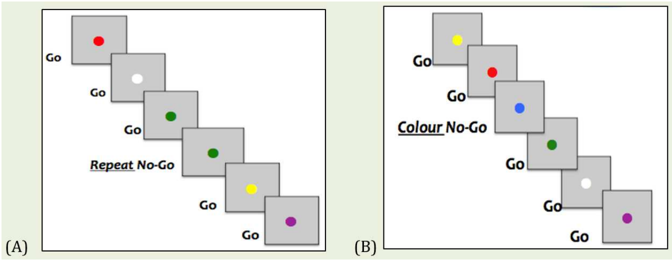

Fig. 1. Go/No-Go task. Subjects should press a button, in a timely manner, in response to each new dot, withholding only in the case of the two “no-go” conditions: (A) a repeat of the previous colour (“repeat condition”) or (B) any blue dot (“colour condition”). Awareness of errors (i.e. pressing the button in the case of a no-go condition) should be acknowledged by a second button press.

new dot appeared. There were two exceptions: for blue dots (colour condition), or dots that were a repeat of previous colour (repeat condition), the subjects should withhold the button press. If they did press the button in either of those scenarios, and realise their error, they should press a second time in order to indicate their awareness of the error.

B. Participants

Data from 28 young participants (aged 18 to 35) were used in this study. All participants reported no history of psychiatric illness, head injury or photosensitive epilepsy, had normal or corrected-to-normal vision, and had no history of color-blindness. Written informed consent was provided before testing began, and all procedures were approved by the Trinity College Dublin ethics committee and in accordance with the Declaration of Helsinki.

C. Data Acquisition and Preprocessing

64 channels of EEG were recorded at 2048Hz. 8 blocks of epochs were collected per subject, with the exception of two subjects, for whom 6 and 5 blocks of epochs were collected, respectively. Each block consists of 200 epochs where 40 were “no-go”. The data were resampled to 64Hz, and bandpass filtered between 4Hz and 32Hz using zero-phase filters. Epochs for both the colour condition (subject pressed the button despite blue dot) and repeat condition (subject pressed the button despite the current dot being the same colour as the previous dot) were extracted from 0.15s to 1s after the subject committed the error. These epochs were then baseline corrected, using an interval from -0.2s to 0s before presentation of the stimulus. Epochs were only retained if the subject pressed the button again to indicate that they were aware of the error. Artefact rejection was then performed, removing any epochs with an amplitude range (highest peak amplitude - lowest peak amplitude) greater than 100μV. Finally, epoch signals were smoothed in the time domain, using a moving mean with a window size of 5 time points (approximately 0.08s).

D. Data Visualisation

Time domain data were plotted for a number of channels in the form of Grand Averages. The data were processed as described in the previous subsection, with the exceptions that the bandpass filter was between 0.4Hz and 32Hz, and the time window was from -0.1s to 1s, relative to the error being committed.

E. Proposed Error Classification Algorithm

1) Feature Extraction: One of the challenges of this study was that the data contained only a small number of epochs per condition. As such, a classifier using a very small number of features was developed, in an attempt to avoid overfitting to noise in the training data due to the so-called curse-of-dimensionality.

EEG signals were taken from 9 fronto-central channels (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, C2). For each channel, a single time point was selected on the basis of providing the best correlation between the training data and their associated class labels (i.e. colour/repeat condition) as the feature representing that channel.

2) Training The Classifier: Each feature was then used as the input of a one-dimensional threshold-based classifier. The threshold was defined on the basis of providing the best possible separation of the training data. To determine the “best” separation, a maximin algorithm was employed, minimizing the possible loss for the worst performing condition. In other words, the threshold was defined such that the minimum sensitivity achieved between the two error conditions was maximized. This decision was taken in order to encourage good performance for both conditions, as they were considered to be equally important.

To find all potential thresholds, the midpoints between each adjacent pair of unique training data points was computed. For each of the candidate thresholds, the percentage of colour condition points below the threshold, and repeat condition points above the threshold, were computed. The lowest of these two percentages was counted as the threshold’s score in the “colour-condition-below-threshold” orientation. The process was repeated with the threshold orientation reversed (i.e. repeat condition below the threshold, colour condition above). The threshold with the highest score (in either orientation) would be selected, and its orientation noted for the classification of future epochs. Specifically, the algorithm for choosing a channel’s classification threshold, based on the training data, was as follows:

$x \leftarrow trainingEpochValues_{channel}$
$y \leftarrow trainingEpochClassLabels$
$x_{col} \leftarrow x_{selectedFeature,colourCondition}$
$x_{rep} \leftarrow x_{selectedFeature,repeatCondition}$
for each *threshold* **do**

$$colPct_{below,threshold} \leftarrow \frac{sum(x_{col} < threshold)}{length(x_{col})}$$

$$repPct_{above,threshold} \leftarrow \frac{sum(x_{rep} > threshold)}{length(x_{rep})}$$

$$colPct_{above,threshold} \leftarrow \frac{sum(x_{col} > threshold)}{length(x_{col})}$$

$$repPct_{below,threshold} \leftarrow \frac{sum(x_{rep} < threshold)}{length(x_{rep})}$$

end for
$mm1 \leftarrow max(min(colPct_{below}, repPct_{above})$
$mm2 \leftarrow max(min(colPct_{above}, repPct_{below})$
if $mm1 > mm2$ **then**
  $maximin_{channel} \leftarrow mm1$
  $selectedThreshold_{channel} \leftarrow threshold_{mm1}$
  $colourConditionBelowThreshold_{channel} \leftarrow T$
else
  $maximin_{channel} \leftarrow mm2$
  $selectedThreshold_{channel} \leftarrow threshold_{mm2}$
  $colourConditionBelowThreshold_{channel} \leftarrow F$
end if

To construct the ensemble classifier, the 9 trained single-channel classifiers were then given a weighted vote. Weightings were based on the maximin scores of each channel's classifier. The weight for a given channel was calculated according to equation 1:

$$w_{channel} = max((maximin_{channel} - 0.5), 0)^4 \quad (1)$$

Scores were first reduced by 0.5, capped with a lower bound of 0, to ensure that only channels with > 50% correct classification of both conditions in the training data received a vote. These scores were raised to the fourth power in order to sufficiently accentuate the votes of better performing channels, while only giving veto power to a single channel if it had substantially outperformed all others.

3) Classifying New Data: With the ensemble classifier trained, voting was carried out in order to classify any new epoch. For each channel, the time domain data would be extracted for the channel's single selected time point. The channel would then cast a vote: 0 for the colour condition or 1 for the repeat condition. Note that which condition fell below the threshold, and which was above, was decided in the earlier threshold training. The votes were then multiplied by each channel's weight, and added together to provide an overall score. If this score were greater than half the sum of all weights, the epoch was classified as being from the repeat condition. Otherwise, it was classified as being from the colour condition.

This strategy was tested on all epochs, using leave-one-out cross-validation. All analysis was carried out in MATLAB, version R2017b.

III. Results and Discussion

A. Condition Separability in Grand Average EEG signals

The signals of the two conditions were seen to be distinguishable in some channels, especially fronto-central ones. "Fig. 2" shows the Grand Average time domain data for channel Cz. The patterns for both conditions - an early negativity, followed by positivity, resemble those seen in other ErrP processing studies [4], [7]. After the error is committed (t=0), the two conditions display similar error-related negativities (ERN), but the following positive peaks are of markedly greater amplitude in the colour condition than the repeat condition. The offset between the two conditions continues until the end of the epoch.

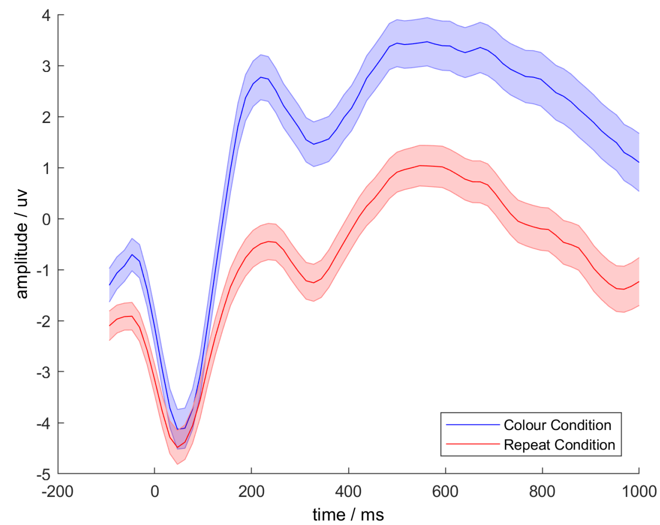

Fig. 2. Grand Average time domain data, channel Cz, bandpass filtered at 0.4Hz to 32Hz, smoothed using a moving mean with a window size of 5 time points (approximately 0.08s)

B. Single-Trial Classification

Not all subjects produced enough epochs to properly gauge the success of any attempted classification. In fact, after artefact rejection, a small number of subjects had produced only 1 trial in each condition. Initially, however, classification was attempted for all subjects with more than 3 artefact-free epochs per error condition. While classification rates of > 50% were achieved for each condition in the majority of subjects, a clear trend was found that there was a greater likelihood of achieving this goal when a subject had produced more epochs, as shown in "Fig. 3".

As such, it was decided that it would be reasonable to focus on subjects with more than 20 epochs per condition. This left 16 subjects. Classification accuracy of greater than 50% was achieved in each condition for 15 of these subjects (94%). The best performance was found for subject 1, with accuracy of 82.61% in the colour condition and 80.00% in the repeat condition, giving a balanced accuracy of 81.30%. Mean classification rates across these 16 subjects were 62.28% in the colour condition, 64.17% in the repeat condition. Mean balanced accuracy was 63.23%. Classification accuracies for the colour condition and repeat condition, as well as balanced accuracies, are reported for the 16 subjects in Table I.

Right-tailed Fisher's exact tests were performed on the confusion matrices for each of these 16 subjects, to further judge the statistical separability of the conditions. For 10 of the subjects (1, 2, 3, 4, 6, 7, 8, 10, 13, 14), the p-value was < 0.05.

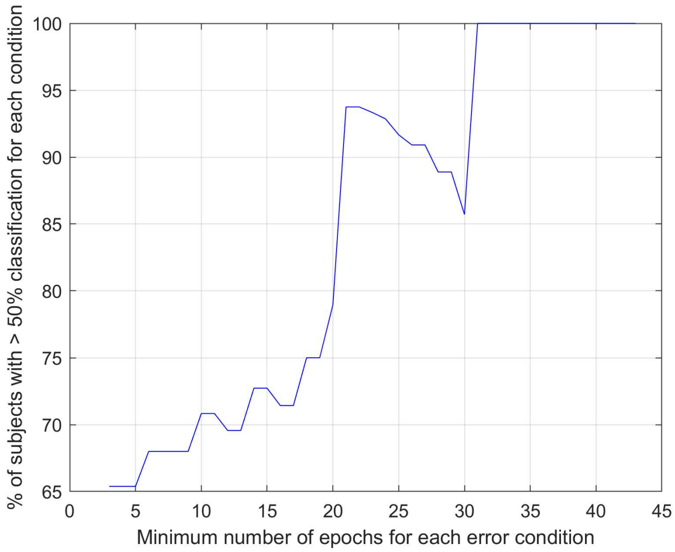

Fig. 3. Rising "successful classification" rate (percentage of subjects for whom > 50% classification was achieved for each condition) with an increasing minimum-epochs-per-condition cutoff employed (correlation coefficient 0.9418, p-value 4.5894e-20)

TABLE I
CLASSIFICATION ACCURACY FOR SUBJECTS WITH > 20 EPOCHS PER CONDITION

Subject	Colour Cond.	Repeat Cond.	Balanced Accuracy
1	82.61%	80.00%	81.30%
2	75.00%	80.95%	77.98%
3	61.11%	64.91%	63.01%
4	66.67%	73.08%	69.87%
5	40.00%	42.11%	41.05%
6	63.41%	63.46%	63.44%
7	66.67%	65.71%	66.19%
8	68.18%	73.33%	70.76%
9	52.00%	56.41%	54.21%
10	62.50%	61.22%	61.86%
11	55.56%	56.76%	56.16%
12	63.16%	53.13%	58.14%
13	72.41%	70.00%	71.21%
14	60.47%	63.64%	62.05%
15	51.52%	62.07%	56.79%
16	55.17%	60.00%	57.59%
Mean	**62.28%**	**64.17%**	**63.23%**

IV. CONCLUSIONS AND FURTHER WORK

For the first time, we have attempted single-trial EEG classification of error conditions that could not have been differentiated by error "type", direction, or severity. Our strategy was to select a small set of time-domain features and use a weighted vote of threshold-based classifiers. Interestingly, our results showed that, for most of the subjects who generated enough epochs per condition, we were able to achieve statistically significant separation of the error conditions.

It is encouraging to see that, even with as few as 6 epochs per condition, classification rates of greater than 50% were achieved for both conditions in two thirds of subjects. This indicates that the small feature set may allow a degree of robustness, even when very few training epochs are available. However, the general trend appears to be that a higher number of training epochs implies a higher likelihood of successful classification. As such, we believe that this work would benefit from further investigation, with more data being generated per subject. Traditional methods such as common spatial patterns (CSP), and linear discriminant analysis (LDA) or support vector machines (SVM) were found to be susceptible to overfitting to the existing data. However, if enough epochs were generated, it may no longer be necessary to use such a small feature set. It may, then, be possible to achieve higher, or more consistently high, classification accuracies with such methods.

Reinforcement learning – a useful application of error detection in BCI – can effectively converge on optimal solutions as long as classification rates are greater than chance level [2], [3]. Thus, for many of our subjects, classification of these very similar errors could be used in a learning system to help improve the performance of a BCI.

This study opens a new door in enhancing brain-computer interactions by requiring less mental workload from the user, leading to a more intuitive and intelligent interaction.

REFERENCES

[1] W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, and E. Donchin, "A neural system for error detection and compensation," *Psychological Science*, vol. 4, pp. 385–390, Dec. 1993.

[2] R. Chavarriaga, A. Sobolewski, and J. del R Millán, "Errare machinale est: the use of error-related potentials in brain-machine interfaces," *Frontiers in Neuroscience*, vol. 8, p. 208, Jul. 2014.

[3] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. del R. Millán, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control," *Scientific Reports*, vol. 5, p. 13893, Sep. 2015.

[4] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "Erp components on reaction errors and their functional significance: a tutorial," *Biological Psychology*, vol. 51, pp. 87–107, Jan. 2000.

[5] S. K. Kim, E. A. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning – using error-related potentials for real world human- robot interaction," *Scientific Reports*, vol. 7, p. 17562, Dec. 2017.

[6] M. Spüler and C. Niethammer, "Error-related potentials during continuous feedback: using eeg to detect errors of different type and severity," *Frontiers in Human Neuroscience*, vol. 9, p. 155, Mar. 2015.

[7] A. Buttfield, P. W. Ferrez, and J. del R. Millán, "Towards a robust bci: Error potentials and online learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 2, pp. 164–168, Jun. 2006.

[8] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using eeg-based reward signals," in *Proc. IEEE International Conference on Robotics and Automation (ICRA'10)*, Anchorage, AK, USA, May 2010, pp. 4822–4829.

[9] P. S. Bernstein, M. K. Scheffers, and M. G. H. Coles, "Where did i go wrong? a psychophysiological analysis of error detection," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21(6), pp. 1312–1322, Dec. 1995.