Classification of Error Related Negativity in Brain Computer Interfaces

Brainwaves Research Lab.

Colorado State University
Brain-Computer Interfaces Lab.

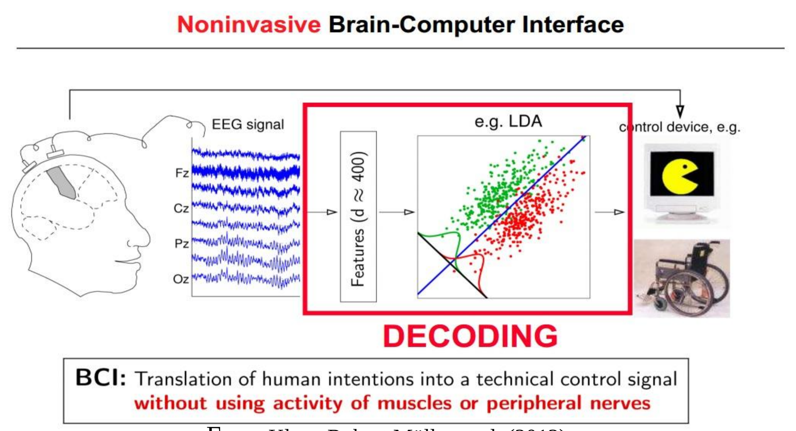
Fereydoon Vafaei, Charles Anderson, William Gavin, Patricia Davies

Non-Invasive Brain-Computer Interfaces

Brain-Computer Interfaces (BCI) are artificial systems that bypass the body's natural communication and control by directly measuring brain activity (typically EEG) associated with the user's intent and interpreting the recorded brain activity into corresponding control signals for various BCI applications.

In many cases, the user's intent may be translated wrongly, and this may generate a specific type of event-related potentials (ERP) named Error Related Negativity (ERN) in the user's brain.

Automatic, fast, and accurate recognition of ERN can improve the performance and robustness of the BCI systems by ignoring or undoing the wrong translations.

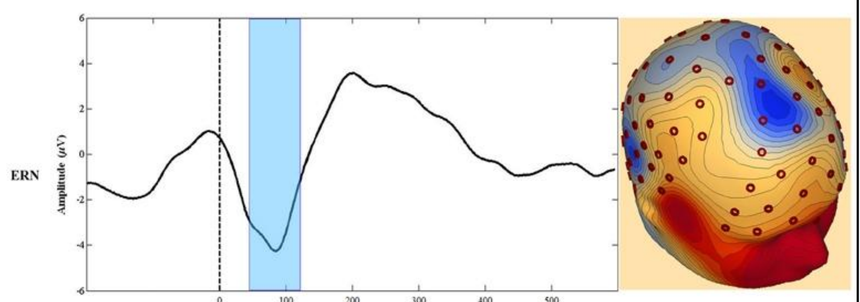


From: Klaus-Robert Müller et al. (2012)

Objective: We study, test and compare various approaches and strategies on how to improve the classification process of ERN. Our study consists of comparing several heuristics for pre-processing of EEG raw data as well as comparison of different classifiers.

Error Related Negativity (ERN)

Error Related Negativity (ERN) is an ERP that peaks 80-100 ms after the response if the user makes an error in performing the tasks. This is generally referred to as response ERN and is believed to be processed in ACC region of the brain.



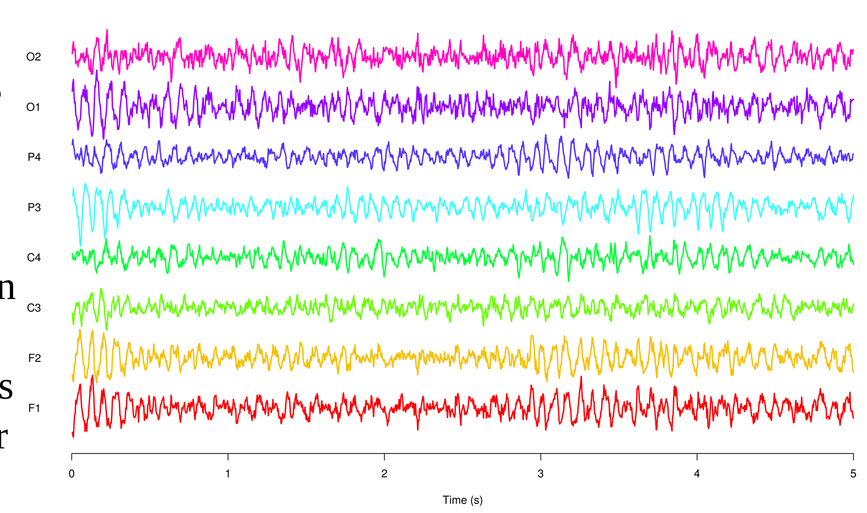
From: Oxford Handbook of ERPs

Similarly, an ERP may be observed 250-300 ms after the presentation of negative feedback stimuli or after observing an unexpected outcome of the response, often referred to as feedback ERN (aka FRN).

Classification of ERN on a single trial basis is a hard problem due to the variability in amplitude and time latency from trial to trial and from subject to subject in BCI experiments. We try to improve the accuracy of the ERN classification in this study.

Participants and Data Collection

Data was recorded from 30 adult participants (16 females), aged between 19 to 27 (μ = 23)by the Brainwaves Research Lab for offline analysis at a later time.



Each participant performed Eriksen Flanker tasks in two consecutive sessions with a short break. The stimuli was congruent and incongruent letter strings and the participants were propmted to right-click or left-click based on the stimuli.

Participants made errors in the range of (2-20)% of the whole number of trials. Correct and incorrect trials were marked during the experiments.

The EEG equipment was "Active Two BioSemi" including 41 channels (6 non-EEG channels which were removed later) and with a sampling rate of 1024 Hz.

Preprocessing the Data

We have tried different configurations in the preprocessing phase. We believe that proper selections of preprocessing parameters has a significant impact on the resulted accuracy of the classification. Our results confirmed this.

The 4 midline EEG channes {FCz, Fz, Cz, Pz}were selected as spatial features. This selection was based on the distribution of ERN signal strength over the scalp, which has been shown to be maximum on the midfrontal channels. Different configurations of the 4 channels were tested for classification. Data coming from each channel was concatenated to data of other channels to construct the input vector.

Table 1: Preprocessing Parameters

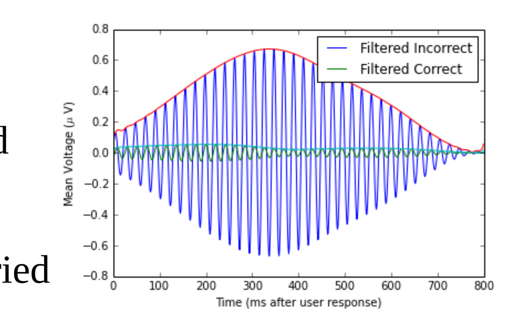
Pre-processing Parameter	Tested Configurations	Selected Settings
EEG channels	Different configurations of the 4 midline channels {FCz, Fz, Cz, Pz}	Concatenated data from the 4 channels on the midline
Bandwidth of bandpass filter (Hz)	0.1-10, 0.1-30, 0.5-10, 0.1-7, 4-7 (theta), 46-50 (gamma)	0.1-10 Hz
Segmentation length (ms)	200, 400, 800	400 ms
Downsampling (Hz)	512, 256, 128, 64, 32, and no downsampling (1024 Hz)	No downsampling (1024 Hz)

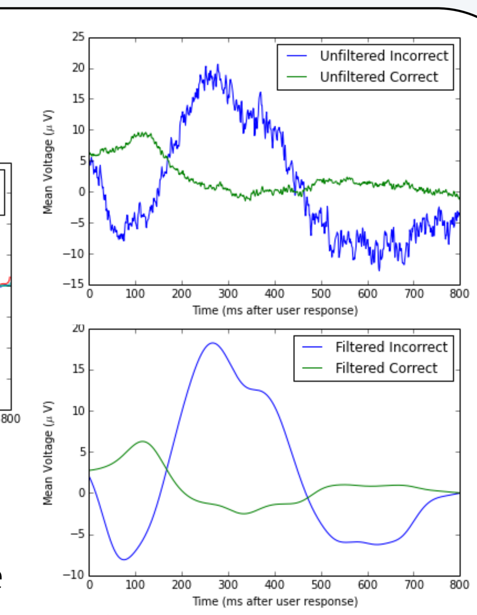
EEG data was referenced by the average of the two earlobe channels, and baseline corrected by 200 ms before the response. Data was also segmented to check the presence or absence of ERN within each segment. Different window length was tested to determine proper segmentation. Finally, data was downsampled with different sampling rates.

Table 1 shows different tested configurations for preprocessing as well as the selected settings.

Bandpass Filtering

As Table 1 shows, we also tested different bandwidth ranges for bandpass filtering to remove the noise. Low-pass filter with the bandwidth (0.1-10) Hz turned out to give the best results (right-bottom plot).





ERN is believed to be a theta oscillation, so we also tried theta range (4-7) Hz versus gamma range (46-50) Hz.

As the left plot shows, there is a noticable distinction between correct and incorrect trials (ERN) in the gamma range. However, on average, gamma range did not give better results compared to lower frequency bandwidth. We also tried the envelope of the waves, but the results were even lower. We think this should be further investigated to see how we can take benefit of the distinction between correct and incorrect trials in the gamma range (although not for all subjects).

Classifiers and Parameters

We tried four classifiers: LDA with shrinkage, Linear SVM, Neural Network, and Ridge Regression, and for each classifier, we tried different parameter settings.

The NN has one hidden layer and performs a logistic regression approach for classification and uses Scaled Conjugate Gradient (SCG) as a parabolic approximation for optimization instead of the traditional approach of gradient descent for optimization.

Table 2: Classifier Parameters

Classifier	Parameter	Tested Values	Selected Value
LDA	Regularization parameter (shrinkage)	1, 0.1, 0.01, 0.001	0.1
Linear SVM	Soft margin parameter C	0.1, 1, 10, 100, 1000	1
Neural Network	Number of neurons per hidden layer	1 - 30	5
	Number of iterations for SCG to converge	1 - 1000	300
Ridge Regression	Regularization parameter	1, 0.1, 0.01, 0.001	1

To evaluate the accuracy of the classifiers, we used Balanced Success Rate (BSR) given the unbalanced nature of our datasets. BSR is computed as the average of true positive rate and true negative rate. We used the first session data as training set, and the second session data as testing set. Parameters are fine-tuned by Leave-One-Out protocol.

Table 2 shows different parameter settings that we have tested and compared.

Results

Table 3 shows the results of the four classifiers applied on the preprocessed data of 10 selected subjects. These 10 subjects had the highest accuracy among all.

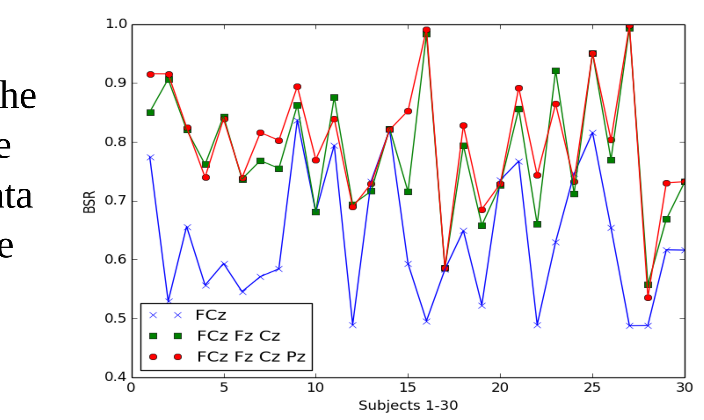
For each of these 10 subjects, there is at least one classifier with the accuracy of higher than 90%. Over these 10 subjects, linear SVM outperforms others. However, on average across all 30 subjects, LDA with shrinkage outperformed others with 80% accuracy.

The ratio of incorrect trials over the whole number of trials is given in the table for each of the 10 subjects.

Table 3: Classification Results

	Training Set	LDA with Shrinkage	Linear SVM	Neural Network	Ridge Regression
Subject	**(Trials/Errors)**	**(BSR)**	**(BSR)**	**(BSR)**	**(BSR)**
S_a	223/18 (8%)	91.6	92.0	**98.3**	80.3
S_b	224/19 (8%)	91.6	94.2	85.1	**97.1**
S_c	234/23 (10%)	89.4	87.4	84.5	**94.9**
S_d	239/5 (2%)	82.0	**90.2**	82.7	72.4
S_e	237/9 (4%)	**99.2**	97.9	98.7	92.7
S_f	229/18 (8%)	82.8	83.0	**90.5**	76.4
S_g	239/8 (3%)	73.3	71.8	49.4	**96.6**
S_h	236/45 (19%)	**95.1**	90.2	91.2	84.8
S_i	239/16 (7%)	80.4	**90.0**	76.1	68.5
S_j	239/5 (2%)	**99.8**	**99.8**	99.6	45.4
	Mean BSR	88.5	**89.7**	85.6	80.9

The plot shows the accuracy for all 30 subjects using the data of only FCz channel versus three and four midline channels. For most subjects, using the concatenated data of 4 channels gives better results. We have not used the data coming from other channels in this study.



Conclusions

Results show that the classification accuracy significantly varies from subject to subject. This is due to high variability in the shape, amplitude and time latency of ERN across different trials and subjects.

We have shown the importance of the pre-processing applied on the raw EEG data for acquiring better results. We have argued that classification studies should include a comprehensive analysis of preprocessing configurations.

Our results indicate that applying low-pass filters for smoothing the waves in raw EEG data performs better than high-pass filters, and there are important features in low frequencies ≤ 0.5 Hz that contribute for a better recognition of ERN.

We have also shown that the positivity Pe which follows Ne component of ERN can be included in the temporal features to better recognize ERN. Our results showed that a 400 ms window length for segments (that includes Pe component) improves the accuracy compared to using shorter segments.

Given the subject-dependent features of ERN, it might not be feasible to formulate a general heuristic that can be applied for all subjects and experiment settings. However, one can use training session to calibrate the classifiers to improve the results of single trial classification of ERN on a real-time basis.

Future Work

Online classification for BCI applications on a real-time basis is required to confirm the results.

Classification of FRN data should be performed and compared to ERN results.

Other classifiers should be tried. Also, other spatio-temporal features should be constructed and tested.