Bachelor thesis

Artificial Intelligence

Radboud University

# Exploring code families and event-types for cVEP BCIs

*Author:*
Julia Janssen
j.a.janssen@student.ru.nl
s4859685

*First supervisor:*
drs. J. Thielen
Artificial Intelligence department
jordy.thielen@donders.ru.nl

*Second supervisor:*
dr. M.W. Tangermann
Artificial Intelligence department
michael.tangermann@donders.ru.nl

June 25, 2021

**Acknowledgements**

I would like to thank my supervisors Jordy Thielen and Michael Tangermann for their guidance and feedback. I would like to thank my first supervisor Jordy Thielen especially for his patience and support during a difficult year. I would also like to thank my teammates Elise Lems, David Teulings and Siebe Geurts for their encouragement, help and support.

**Abstract**

A brain-computer interface (BCI) is a system that makes it possible for a user to interact with a computer without muscular control. Code-modulated visual evoked potential (cVEP) BCI can be used to create visual speller systems. To make a step towards practical, real-world cVEP BCI by improving performance and reducing calibration time, this thesis compares three methods of decomposing codes, called event-types, when using the reconvolution method by Thielen et al. (2015) [12] to generate templates. This is done on two different data sets, one using random stimulation codes and one using modulated Gold codes.

It is found that event-types that take into account the nonlinearity of the visual system perform better in terms of classification accuracy, auto explained variance and cross explained variance than event-types that assume linearity. It is concluded that using the edge responses of a flash as an event-type is the most practical approach, as it models only two events and therefore requires less training data than event-types with more events. The contrast event-type is also more versatile, as it does not require events in the validation set to have appeared in the training set.

# Contents


**1 Introduction** **3**
- 1.1 Brain-computer interfaces . . . 3
- 1.2 Reconvolution and events . . . 4
- 1.3 Code families . . . 5
  - 1.3.1 Modulated Gold codes . . . 5
  - 1.3.2 Random stimulation codes . . . 6
- 1.4 Hypothesis and research question . . . 7

**2 Methods** **8**
- 2.1 Data sets . . . 8
  - 2.1.1 Differences and limitations . . . 9
- 2.2 Reconvolution . . . 10
- 2.3 Canonical Correlation Analysis . . . 12
- 2.4 Classification . . . 13
- 2.5 Evaluation . . . 13
- 2.6 Code availability . . . 14

**3 Results** **15**
- 3.1 Modulated Gold codes . . . 15
  - 3.1.1 Classification accuracy . . . 15
  - 3.1.2 Explained variances . . . 16
  - 3.1.3 Transient responses . . . 17
- 3.2 Random stimulation . . . 19
  - 3.2.1 Classification accuracy . . . 19
  - 3.2.2 Explained variances . . . 19
  - 3.2.3 Transient responses . . . 20

**4 Discussion** **23**
- 4.1 Limitations . . . 23
- 4.2 Code families and future work . . . 23
- 4.3 Hypothesis . . . 24

**5 Conclusion** **25**

**A Full results** **28**
A.1 Contrast event-type . . . . . . . . . . . . . . . . . . . . . . . . 28
A.2 Duration event-type . . . . . . . . . . . . . . . . . . . . . . . . 29

**B Alternate preprocessing method** **31**

# Chapter 1

# Introduction

In 2016, approximately 331,000 people were estimated to suffer from motor neuron diseases worldwide (Logroscino et al. [6]). As these diseases progress, patients often develop paralysis. Swallowing, breathing and speaking may become difficult. Speaking may even become impossible. In the same paper, it is also reported that the stress of living with the condition is known to increase risk of developing depression, insomnia and anxiety. So far, motor neuron diseases have no known cure. [11]

Being able to communicate and interact with family and friends or systems such as wheelchairs or other aids is important to increase quality of life for those suffering these symptoms. When a patient is still able to do these things, they may continue to perform some tasks by themselves and have a sense of autonomy in their daily life. [4]

## 1.1 Brain-computer interfaces

Brain-computer interfaces (BCI) can provide a solution for these issues. By recording electrical activity from the brain, systems can be developed that use brain signals for input, bypassing the need for muscle control. One common approach is to use visual evoked potentials (VEP) - brain responses to visual stimulation. Noise tagging BCI uses pseudo-random bit-sequences, called codes, where a '1' represents a white frame and '0' represents a black frame. These complex sequences of flashes evoke code-modulated visual evoked potentials (cVEP).

In BCI visual speller applications, each symbol is associated with a specific code. By attending to a specific symbol, a cVEP related to the associated code is evoked and recorded by the electroencephalogram (EEG), which can be analyzed and matched to a template response to a particular code and thus the associated symbol. Using this, it is possible to build keyboards, wheelchair controls or even games [1]. Matching the EEG response to the right template is an important part of any noise tagging BCI application.

Visual spellers are still only rarely used outside of laboratories. One reason for this are the calibration times. Because brain activity varies between people and even between sessions with the same user (for example due of tiredness), the system needs to be calibrated for each session. This takes time and is not practical for everyday use.

## 1.2 Reconvolution and events

One way to shorten calibration times is to use reconvolution to generate templates instead of obtaining templates by recording many responses from the user. In the reconvolution model by Thielen et al. (2015) [12], templates are built up from the responses to individual events. These are pre-defined subsequences in the code. For example, an event may be a short flash or a long flash. The linear superposition hypothesis states that the response to a sequence of events is the linear summation of the responses to the individual events. In other words, if we can learn the responses to individual events, we can generate predictions of the response to full codes. Exactly how these individual events are defined is therefore an important choice.

This definition of events is called the event-type. For example, when we define our events to be short (010) and long (0110) flashes, this is called the 'duration' event-type. Other examples are 'contrast', which defines every rise (01) and fall (10) in the code as an event, and 'simple', which defines every '1' in the code as an event.

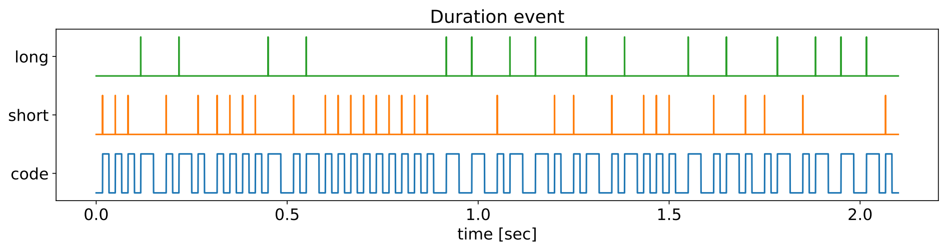



Figure 1.1: Duration event-type example

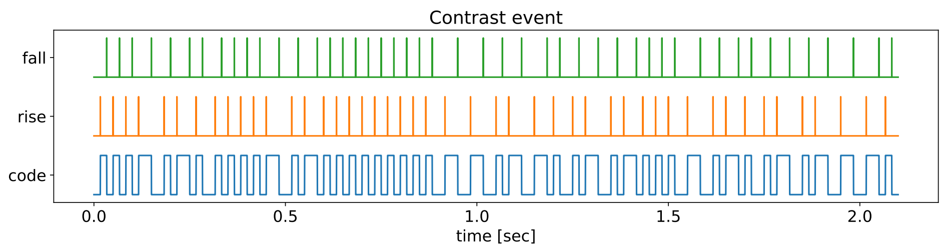



Figure 1.2: Contrast event-type example

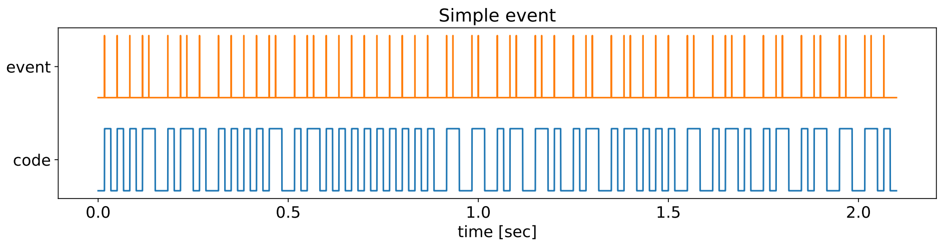



Figure 1.3: Simple event-type example

## 1.3 Code families

Another aspect for consideration is the choice of code family. Isaksen et al. (2016) [5] compared three code families and found that although there was no code family that was more accurate than the others in general, subjects did have an optimal code where their accuracy was significantly higher. Although there may not be a 'best' code family in terms of accuracy, they are interesting to look at for another reason. Since different code families have different properties, like for example the maximum length of a single flash, this influences the events. For example, a code with a maximum flash length of $i$ bits will have $i$ unique duration events. By having certain code properties, other events can be defined. It is possible certain event definitions work best with a certain code family.

Yasinzai et al. (2020) [16] investigated the brain response to short code sequences by using the contrast event-type and superposition to generate templates. They found the best results using short flashes with enough time between the flashes to ensure responses didn't overlap. They also found that when repeating flashes of the same length, the response becomes smaller with each repetition. It is therefore best to limit these repetitions. Nagel and Spüler (2018) [10] found that the number of up/down changes in a code was another important property and found optimal results for codes containing 7 changes for every 15 bits of code.

### 1.3.1 Modulated Gold codes

Modulated Gold codes have the special property of comprising entirely out of '01' and '10' subsequences. This means the code can only be up or down for at most 2 bits at a time. These codes have the optimal 1 and 2-bit flashes and limited repetitions of the same flash, but not the recommended time between flashes to prevent overlap. Using this code family gives a chance to further investigate these short flashes.

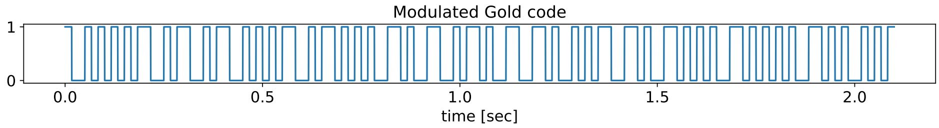



Figure 1.4: Example of a modulated Gold code

### 1.3.2 Random stimulation codes

Random stimulation codes are created by having a random generator pick between 0 and 1 for each bit. These codes can have very long flashes, but they are much rarer than the short ones. This distribution of flash lengths is visualized in figure 1.5. Using this code family gives a chance to investigate the response to very long flashes.

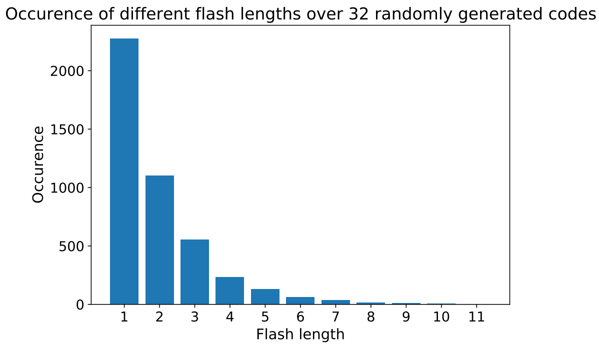



Figure 1.5: Number of flashes by duration over 32 random stimulation codes.

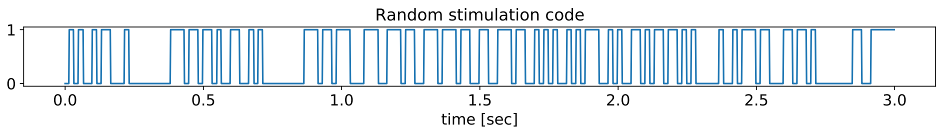



Figure 1.6: Example of a modulated Gold code

So far, little research has been done into this choice of event-type and code family. In this thesis, I will compare three event-types: simple, duration and contrast. These three all represent a different idea of the brain response. Simple assumes linearity of the brain response. Although it is known from effects such as habituation and sensitization that the visual system is nonlinear, assuming linearity has been shown to be sufficient for steady-state VEP [3]. Still, it might be preferable to take nonlinearity into account. Duration allows for some nonlinearity by assuming each flash length to have a different response. Contrast does this by assuming the brain responds to change.

## 1.4 Hypothesis and research question

In this thesis, I will further investigate which event-type/code pairing best explains the brain response, and which produces the highest accuracy. I will do this by comparing the simple, duration and contrast event-types on two data sets, both using different code families: one using modulated Gold codes and one using random stimulation.

I will compare them primarily on the explained variance and will also look at the classification accuracy. The goal is to come to a further understanding of these event-types and the brain response they model to formulate recommendations for future work on cVEP BCI.

This is formalized in my research question:

**How do the 'simple', 'duration' and 'contrast' event-types compare when applied to data sets using modulated Gold codes and random stimulation?**

By taking into account nonlinearity, the contrast and duration event-types will likely outperform the simple event-type. The findings from Yasinzai et al. [16] imply random stimulation might have lower accuracy overall for the duration event-type than modulated Gold codes, since long flashes are not predicted with good accuracy.

In the next chapters I will first discuss the data sets I used in this research. Next I will explain in depth the methods I used, focusing especially on reconvolution. In the results section I will share my results and talk about the conclusions that can be drawn from them. Finally, I will end with a discussion of the results and suggestions for further work.

# Chapter 2

# Methods

In this section, I will describe my methods. I will first talk about the two data sets I used and then describe the analysis pipeline in detail. Finally, I will explain the evaluation methods I used.

## 2.1 Data sets

The data set I used to evaluate modulated Gold codes was produced by Thielen et al. (2021) [13]. Thirty participants participated in their offline experiment. They used pre-defined modulated Gold codes of 126 bits with a minimized cross-correlation. These were presented on a monitor with a 60 Hz refresh rate, meaning presenting one bit-sequence takes 2.1 seconds. Each sequence was repeated 15 times for a trial that lasted 31.5 seconds.

Each participant performed five EEG runs. They tested a calculator application with 20 symbols. One run consisted of 20 trials, one for each target. This gives a total of 100 trials per participant, five per bit-sequence. Their experiment was based on synchronous control, meaning each trial had a set time interval. Each trial lasted 31.5 seconds, followed by an inter-trial interval of 1 second.

Eight channels were recorded, namely Fz, T7, O1, POz, Oz, Iz, O2 and T8. The EEG data was preprocessed with a high-pass filter at 2Hz with a second order Butterworth filter, then with a low-pass filter at 30Hz using a 6th order Butterworth filter. The data was downsampled from 512 Hz to 120 Hz and preprocessed using FieldTrip [2].

The data set I used to evaluate random stimulation codes was produced by Nagel and Spüler (2019) [9]. Nine participants participated in their offline experiment. They used the MT19937 [7] random generator to generate random stimulation codes. Sequences were not repeated and each participant was shown different codes. They were presented on a monitor with a 60 Hz refresh rate.

Each participant performed three runs in the training phase with trial durations of 5s, 4s and 3s. Then they performed 14 runs in the testing phase with a trial duration of 2s. They tested a matrix keyboard with 32 targets. One run consisted of 32 trials, one for each target, selected in lexicographic order. 32 channels were recorded, namely Fz, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P5, P3, P1, Pz, P2, P4, P6, PO9, PO7, PO3, POz, PO4, PO8, PO10, O1, POO1, POO2, O2, OI1h, OI2h, and Iz. The ground electrode was at FCz and the reference electrode at Oz.

The EEG data was preprocessed with a bandpass filter, filtered between 0.1Hz and 60Hz using an 8th order Chebyshev filter, then notch-filtered at 50Hz.

I combined the runs from the training phase with 3 different durations into one set, cutting off the 5s and 4s trials at 3s so all trials were the same length. For subject 1, this made the codes used in the 5s trials identical to the ones used in the 3s trials, causing these trials to be confused with each other. For the other eight subjects, codes did not repeat and this confusion did not occur. For this reason, I decided to exclude subject 1 from my analysis.

### 2.1.1 Differences and limitations

Initially, to adjust for the difference in recording layout and make for a fair comparison between the two data sets, I intended to only take into consideration the eight channels that were used in both data sets. However, since I was unable to find a list of which channels in the data set by Nagel and Spüler correspond to which channel on the EEG cap, it was impossible to identify these eight channels correctly. Therefore, I decided to use all 32 channels after all. Although this negates the risk of selecting the wrong channels, not knowing which channel corresponds to which location is still an issue. Although it does not affect classification, it is impossible to construct the correct spatial filter (described later in this section). For this reason, we will not look at the spatial filters found for this data set in the results section.

For a fair comparison, both data sets should have the same preprocessing applied. However as I do not have access to the raw data for the random modulation data set, my only option was to apply to the modulated Gold codes data a similar preprocessing as was used for the random modulation data. Since the FieldTrip [2] toolbox I used for preprocessing doesn't currently have Chebyshev filters implemented, I instead approximated with a 0.1 Hz 2nd order Buttersworth high-pass filter, 60 Hz 6th order Buttersworth low-pass filter, and a 49-51 Hz bandstop filter. Since the data that was preprocessed this way gave significantly worse results, and because a direct comparison between the results of these two data sets is not an option anyway because of their differences in participants, number of participants, channel layout, etc., I decided to use the authors' original preprocessing method for the modulated Gold data set. I did use the alternate preprocessing method as a way to validate my results. This will be discussed shortly in the next section.

## 2.2 Reconvolution

The reconvolution method is the core of my analysis. In this section, I will explain in-depth how it works.

Reconvolution can be seen as a two step process. The first part of the reconvolution process is called the estimation-step, where full responses are decomposed into responses to individual events. Next is the generation-step, where these decomposed responses are used to generate full template responses. The steps are visualized in figure 2.1.

This decomposition of a single full sequence is expressed in the formula

$$\tilde{x}(t) = \sum_{t=1}^{L} l_1(t)r_1(t-\tau) + \cdots + l_n(t)r_n(t-\tau)$$

where $x(t)$ is the full sequence at time $t$. $n$ is the number of unique events. $l_i(t)$ has value 1 if a particular event occurred at time $t$, where $i$ is in between 1 and the number of events $n$. $r_i(t)$ are the responses to event $i$ at time $t$, where $i$ is in between 1 and $n$. Response vector $r$ is obtained through canonical correlation analysis, which is described in the next section. $L$ is the length of $r_i$. Although here $L$ is the same for all events, the framework allows this to be variable also. This can also be written in matrix notation:

$$\tilde{x} = M_1 r_1 + \cdots + M_n r_n = [M_1 \ldots M_n] \begin{vmatrix} r_1 \\ \vdots \\ r_n \end{vmatrix} = Mr$$

$n$ is the number of unique events. $M_i$ is a structure matrix that has a value 1 in column 1 at all rows $t$ where event $i$ occurred at time $t$. In columns 2 to $L$, ones shift down to $t+L$, resulting in a diagonal line in the matrix. $M_i$ is of size $s \times L$, and the full structure matrix M is size $s \cdot (n \cdot L)$, where $s$ is the number of samples $r_i$ is a column vector of $L$ samples, and $r$ is the size of $(n \cdot L) \cdot 1$. An example of a structure matrix can be seen in figure 2.2.

This model can be generalized to

$$X = [x_1 \ldots x_n] = [M_1 \ldots M_n]r = Mr$$

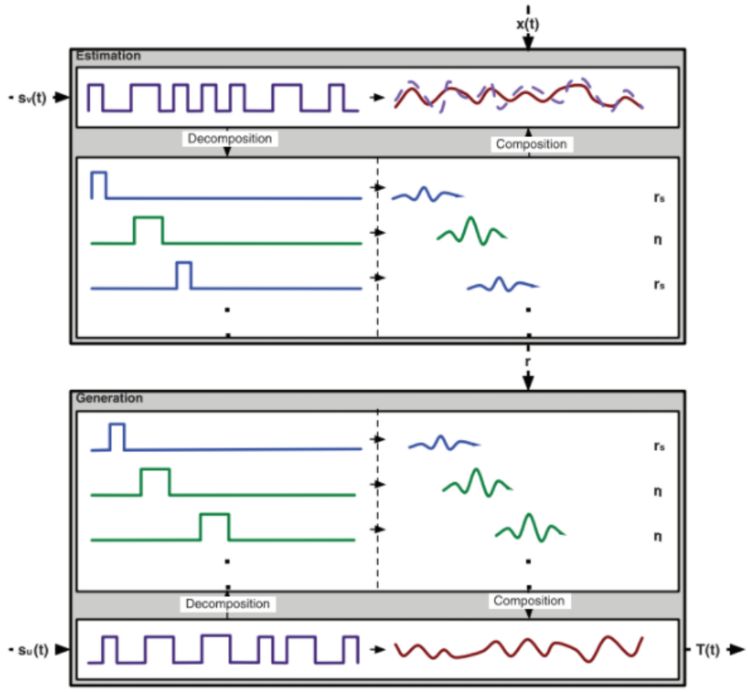



Figure 2.1: In this figure, $x(t)$ is the recorded response. It is decomposed according to the structure of the bit-sequence $s_v(t)$ in the estimation-step. This way, $r_s$ and $r_l$ are obtained, which are the transient response to short and long flashes, respectively. In the generation-step, $r_s$ and $r_l$ are used to compose an estimated response to the new bit-sequence $s_u(t)$. Here the duration event-type was used with two unique durations. Image reproduced with permission, Thielen et al. (2015) [12].

which is a linear regression problem. Here, $X$ is the concatenation of $k$ single-channel, single-trial full sequences, and $M$ is the concatenation of all corresponding structure matrices. $r$ can be found by solving

$$r = M^+X$$

where $M^+$ is the pseudo-inverse of $M$.

In the generation-step, the structure matrix and learned responses to events are combined to generate templates $T$ for unseen bit-sequences. This is done by multiplying the response vector $r$ with the structure matrix $M$.

$$T = Mr$$

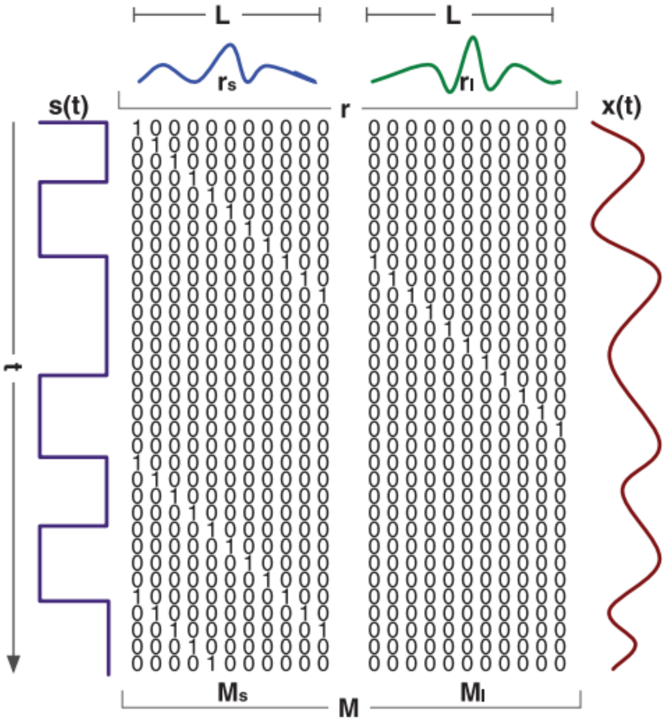



Figure 2.2: An example of a structure matrix $M$. Here the duration event-type was used with two unique durations. $M$ is the concatenation of $M_s$ and $M_l$. $M_s$ and $M_l$ have a value 1 when a short or long pulse, respectively, occurs at time $t$. The value 1 shifts down $L$ rows. Image reproduced with permission, Thielen et al. (2015). [12].

## 2.3 Canonical Correlation Analysis

Reconvolution is performed on the data of one channel. However, in this case we have eight or thirty-two. A way to combine the data channels into one is by using a spatial filter $w$, which can be obtained using Canonical Correlation Analysis (CCA). The response vector $r$ is also obtained by CCA.

The spatial filter $w$ is of size $c \times s$, where $c$ is the amount of channels and $s$ the amount of samples, and is used to reduce input from the (in this case) eight EEG channels into a single channel that can be ran through the reconvolution process. $w$ contains a weight for each EEG channel, and when multiplying $w$ with the data $X$, it gives a single signal of spatially filtered data $x$. By using a spatial filter to obtain this signal, it means the most informative channels (for these visual speller applications typically the ones around the occipital lobe) add more to the filtered data than the less informative channels.

The temporal response vector $r$ contains the associated response for any of the events.

To perform CCA, we again take $M$, the concatenation of all structure matrices from the $k$ training trials. The data from all training trials is concatenated into $S$, which has size $c \times (s \cdot k)$. $w$ and $r$ are then found by solving:

$$\max_{w,r} \rho(w^T S, r^T M) = \frac{w^T S M^T r}{\sqrt{w^T S S^T w \cdot r^T M M^T r}}$$

CCA optimizes $w$ and $r$ to maximize the correlation between $x$ and $t$.

## 2.4 Classification

Having obtained the spatially filtered $x$ with CCA and the templates $T$ with reconvolution, the final step is to find the template that is most similar to $x$. This is done by calculating the Pearson's correlation between $x$ and each individual template $t_i$ and selecting the highest one. This is given by

$$\hat{y} = \underset{i}{\text{argmax}} \frac{x^T t_i}{\sqrt{x^T x \cdot t_i^T t_i}}$$

To prevent overfitting or selection bias, I use 10-fold cross validation. This involved splitting the data set into 10 groups and using each group as a validation set once with the other 9 groups as training sets to fit CCA on. The classification accuracy and explained variances are computed 10 times per subject, once in each fold. The average result over the folds is taken.

## 2.5 Evaluation

Evaluation of the various event-type/code family pairs is done by looking at the classification accuracy and the explained variance of the generated templates.

The classification accuracy is the percentage of correct target selections. This is obtained by solving

$$accuracy = \frac{\text{number of correct trials}}{\text{total number of trials}} \cdot 100$$

The explained variance is the squared Pearson's correlation between the obtained template and single-trial recordings of the same target. This gives an estimate of the amount of variance of the full EEG recording that is explained by the model.

The cross explained variance is the squared Pearson's correlation between a single-trial recording and templates of other labels. This is a way to measure the amount of confusion with incorrect labels.

Pearson's correlation is given by

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

To determine which results are significant, the Wilcoxon signed rank test will be used for the accuracies since these cannot be assumed to be normally distributed as they tend to on the higher end and can't go above 100%. For the explained variances, paired t-tests will be used. With 30 samples, the explained variances for the modulated Gold data set can be plotted in a histogram and seen to be distributed normally. Since the random stimulation data set only has 8 samples, it's not possible to check its distribution. However, it is safe to assume the distribution won't be different that it is for the modulated Gold codes.

## 2.6 Code availability

All code used in this thesis is available at https://gitlab.socsci.ru.nl/J.A.Janssen/.

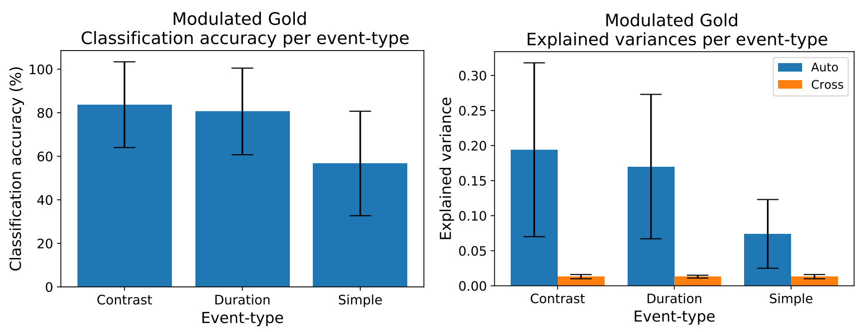
# Chapter 3

# Results

In this section, I will share the results found using the method described in the previous section. For both data sets, I will first show the average classification accuracy and explained variances over subjects. Next, I will show the average transient responses found for each event-type. Finally, I will discuss the conclusions we can draw from these results.

## 3.1 Modulated Gold codes

This section is about the results found for the modulated Gold codes data set.




(a) Classification accuracy (b) Auto and cross explained variance

Figure 3.1: Classification accuracy and explained variances per event-type for modulated Gold codes

### 3.1.1 Classification accuracy

In terms of classification accuracy (figure 3.1a), the contrast and duration event-type both significantly outperform simple. Contrast with a mean of 83.7 and simple with a mean of 56.8, significant with a Wilcoxon rank test

($p < 0.001$, Bonferroni corrected). Duration with a mean of 80.6 against simple is significant with a Wilcoxon rank test ($p < 0.001$, Bonferroni corrected). There is no significant difference between contrast and duration (Wilcoxon rank test $p = 0.459$, Bonferroni corrected).

This means that for classification accuracy when using modulated Gold codes, contrast and duration are both preferred over simple.

### 3.1.2 Explained variances

For cross explained variances (figure 3.1b), we again see contrast and duration outperform simple, with no significant difference between the two. Contrast with a mean of 0.013 and duration with a mean of 0.014 are not significantly different on a paired t-test ($p = 0.054$, Bonferroni corrected). Contrast against simple with a mean of 0.013 is a significant difference with a paired t-test ($p = 0.006$, Bonferroni corrected). Duration against simple is significant with a paired t-test ($p < 0.001$, Bonferroni corrected).

This means that for cross explained variance when using modulated Gold codes, contrast and duration are both preferred over simple.

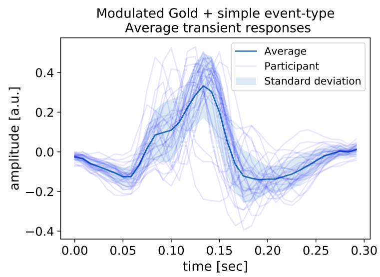
For auto explained variance (figure 3.1b), the difference between all three event-types is significant. Contrast with a mean of 0.194 outperforms duration with a mean of 0.170 on a paired t-test ($p < 0.001$, Bonferroni corrected). Contrast also outperforms simple with a mean of 0.074 (paired t-test $p < 0.001$, Bonferroni corrected). Duration also outperforms simple (paired t-test $p < 0.001$, Bonferroni corrected). This means that in terms of auto explained variance when using modulated Gold codes, contrast is the best choice of these three event-types.

All things considered, contrast is narrowly the best performing event-type when using modulated Gold codes, only significantly outperforming duration in terms of auto explained variance (contrast mean 0.194, duration mean 0.170). Although contrast is preferred, duration is also a good choice since there is no significant difference between the two in terms of classification accuracy and cross explained variance. The simple event-type is outperformed by both contrast and duration on all three metrics and is not a good choice.

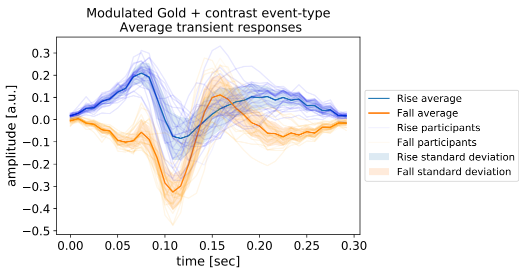
Using the alternate preprocessing method for the modulated Gold data set that was described in the previous section, we can perform the same analysis of event-types. This is a second check that can further validate these findings. Again, contrast and duration are preferred over simple in terms of classification accuracy and cross explained variance. For auto explained variance, contrast is preferred over duration, but there is no significant difference between contrast or duration against simple. All in all, the contrast event-type is the preferred choice and duration is preferred over simple. This is in line with what was found for the authors' original preprocessing method, further validating those findings. The full results for the alternate preprocessing method can be found in the appendix.

### 3.1.3 Transient responses

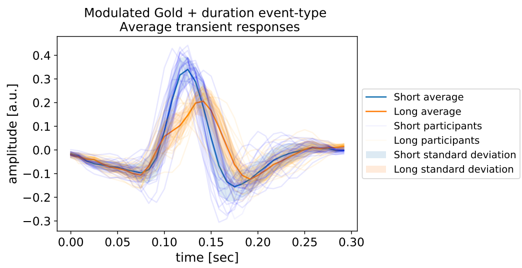
These are the average transient responses found for each event-type. They were found by taking the average of all participant responses.




(a) Simple




(b) Contrast

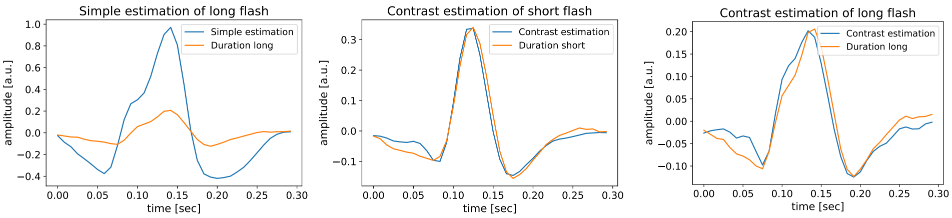



(c) Duration

Figure 3.2: Transient responses per event-type for modulated Gold codes

Looking at these transients, we can notice some interesting things. Firstly, the simple event-type assumes the brain response is linear, meaning that flashes of more than 1 bit long are modelled as two pulses shortly after each other, doubling the amplitude. Looking at the responses found by the duration event-type, we can see this is not appropriate. The peak of the response for the long event is shifted a bit to the right, but is not doubled in amplitude. Using the response found by the simple event, we can shift it to the right by one bit and add this to the original response to find the response the simple event-type estimates for long flashes.

We can do something similar with the contrast event-type. The rise response, added to the fall response that's been shifted by $n$ bits, should be equal to the response found for flashes of duration $n$. We can verify how well this event-type models the brain response by comparing these responses.



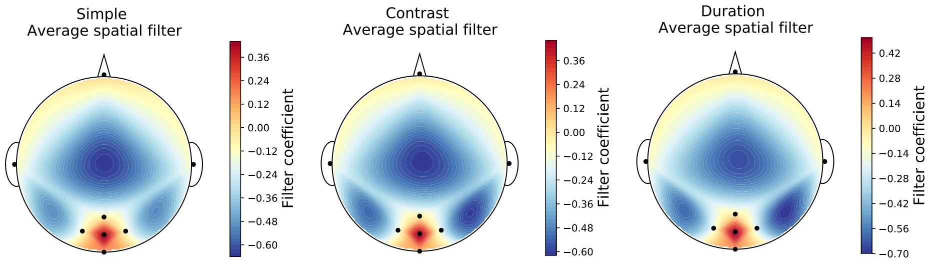

(a) Simple - long (b) Contrast - short (c) Contrast - long

Figure 3.3: These three figures show how closely the simple and contrast event-types can model the duration response.

Figure 3.3a shows the transient response the simple event-type finds for long flashes and the transient response the duration event-type finds for long flashes. It is clear that the simple event-type does not have an accurate response for these flashes, which explains why the simple event-type is outperformed by both contrast and duration. We can conclude that assuming linearity of the brain response is not appropriate for modelling these events.

Figure 3.3b shows the contrast event-type estimating the duration response to a short flash and figure 3.3c shows the estimation for a long flash. These are very close to the transient response found by the duration event-type, meaning the contrast event-type explains the brain response about as well as the duration event-type does, further validating the conclusion that taking nonlinearity into account gives better modelled responses.

Finally, these are the spatial filters found for each event-type.



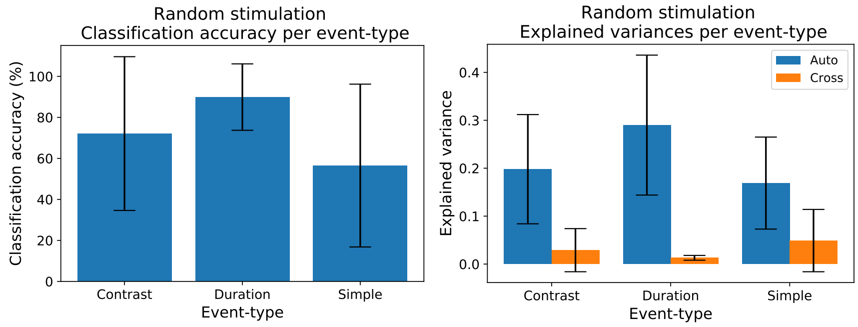

(a) Simple (b) Contrast (c) Duration

Figure 3.4: Spatial filters per event-type for modulated Gold codes

As expected, there is little difference between these spatial filters. Since visual activity is processed in the occipital lobe, the channels around this area are expected to be the most interesting regarding the response to visual stimulation. As expected, the difference between the event-types is in the transient responses and not the spatial filter.

## 3.2 Random stimulation

This section is about the results found for the random stimulation data sets.




(a) Classification accuracy (b) Auto and cross explained variance

Figure 3.5: Classification accuracy and explained variances per event-type for random stimulation codes

### 3.2.1 Classification accuracy

In terms of classification accuracy, there is no significant difference found between any of the three event-types. The difference between contrast and duration is not significant with a Wilcoxon rank test (means of 72.1 and 89.9, $p = 0.282$, Bonferroni corrected). With a mean of 56.5, the difference between contrast and simple is not significant with a Wilcoxon rank test ($p = 0.339$, Bonferroni corrected). Finally the difference between duration and simple is not significant with a Wilcoxon rank test ($p = 0.06$, Bonferroni corrected). This means there is no optimal event-type found in terms of classification accuracy.

### 3.2.2 Explained variances

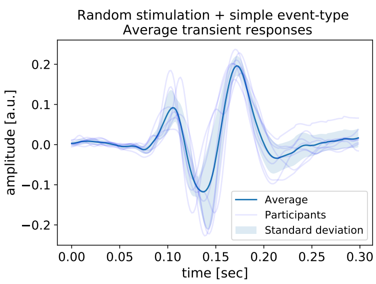
In terms of auto explained variance, the difference between all three event-types is significant. Duration with a mean of 0.146 significantly outperforms contrast with a mean of 0.114 (paired t-test $p = 0.006$, Bonferroni corrected). Contrast outperforms simple with a mean of 0.096 (paired t-test $p = 0.036$, Bonferroni corrected). Duration also outperforms contrast (paired t-test $p = 0.003$, Bonferroni corrected). In terms of auto explained variance for random stimulation, duration is the optimal choice of event-type.

In terms of cross explained variance, there is no significant difference found between any of the three event-types. The difference between contrast and duration is not significant with a paired t-test (means of 0.029 and 0.013 respectively, $p = 1$, Bonferroni corrected). With a mean of 0.049, the difference between contrast and simple is not significant with a paired t-test ($p = 0.891$, Bonferroni corrected). Finally the difference between duration and simple is not significant with a paired t-test ($p = 0.492$, Bonferroni corrected). This means there is no optimal event-type found in terms of cross explained variance.

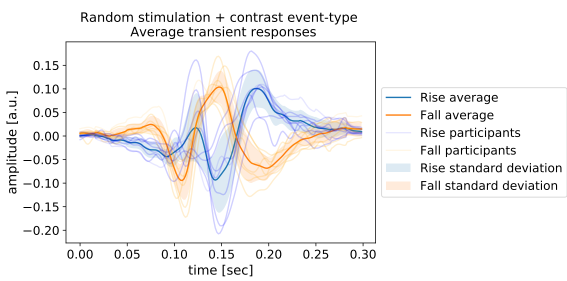
This means that for random stimulation codes, duration is somewhat preferred out of these three, as its auto explained variance is significantly better than the other two. Although the difference is not significant, it also has the best means in terms of classification accuracy and cross explained variance. This is an indication that with more participants, the duration event-type might be found to be preferred on those two measures as well.

### 3.2.3 Transient responses

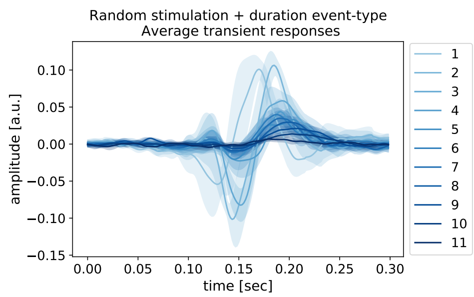
These are the average transient responses found for each event-type. They were found by taking the average of all participant responses.




(a) Simple



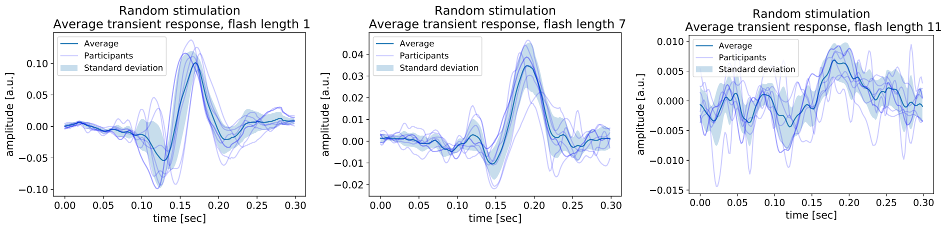

(b) Contrast




(c) Duration

Figure 3.6: Transient responses per event-type for random stimulation. Note that for the duration event-type it was impractical to include the response for each participant in the figure. Instead, these figures are included in the appendix.

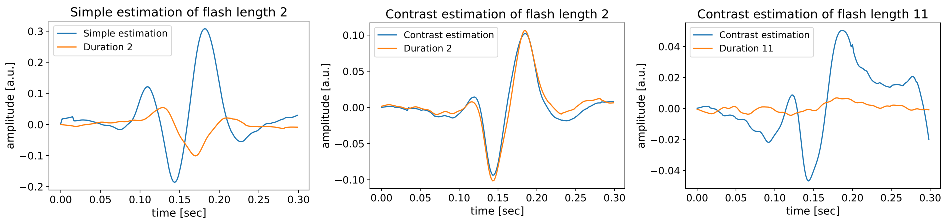
The first thing to note is the duration event-type, which looks very different from the duration event-type for modulated Gold codes, since it has nine additional events. The amplitude becomes lower for each longer event after 3, presumably because each longer event is rarer and therefore has less data to be modelled on, resulting in a less optimal fit.




(a) Length 1 (b) Length 7 (c) Length 11

Figure 3.7: Average transient responses for the duration event-type, flashes of length 1, 7 and 11. Note how the response becomes less pronounced for longer events, since there is less data for them.

Next, like with the modulated Gold codes, it is possible to use the simple and contrast event-types to estimate the duration responses.



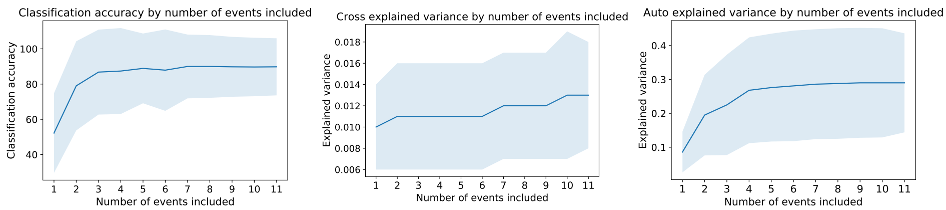

(a) Simple - length 2 (b) Contrast - length 2 (c) Contrast - length 11

Figure 3.8: These three figures show how closely the simple and contrast event-types can estimate the duration response.

Figure 3.8c reveals an advantage the contrast event-type has over the duration event-type that is not visible when only looking at the performance scores, namely that it is the more practical of the two. The contrast event-type always consists of two events: rise and fall. But the duration event-type may have more events, as can be seen with the random stimulation codes. Contrast requires less data since it has fewer events it needs to fit. For flashes of length 11, although the duration event-type struggles to find a good fit, the contrast event-type can be used to produce a reasonable response. The reverse is also true: since contrast has less events, it is likely to fit its events better as it has more data per event.

Another advantage of the contrast event-type is that while duration needs the same flash lengths to occur in the training and validation set, contrast does not. Since the duration event-type learns the responses to unique flash lengths, if an event occurs in the validation set that did not occur in the training set, it does not have a fitted response for this event. The contrast event-type does not have this limitation, as any flash that occurs will have a rise at the start and a fall at the end. This makes contrast the more flexible of the two, as it can be used even if the training and validation set contain different events.

Finally, since longer events are rarer in random stimulation codes, we can ask ourselves how useful it is to take these long flashes into consideration. Finding a transient response for these rare events adds to the computing time, and an excessively long computing time is no good for real-world BCI systems. To investigate the information gained by including each additional event, I plotted the classification accuracy and explained variances against the number of events included. This gave the following results:




(a) Classification accuracy (b) Cross (c) Auto

Figure 3.9: Classification accuracy and explained variances plotted against the number of events included.

Including events past 4 does not significantly improve performance anymore. The main reason for this is that in random stimulation codes, short flashes are much more common than long ones. Because they occur less, less information is gained by including them. Another reason is that because of their relative rarity, there is less data available for these longer flashes and it is harder to estimate the transient response well. This further reduces the information gained by including these events.

With this in mind, when using the duration event-type for random stimulation codes in a situation where reducing computing time is of interest, it is sufficient to only include flashes of length 1, 2, 3 and 4 when generating templates. Note that this is only true for random stimulation codes since longer events are rarer and it may not be true when these events occur more.

As mentioned before in the methods section, because it is unknown for this data set which channel corresponds to which channel on the EEG cap, I was unable to obtain the correct spatial filters and they are not included in this thesis.

# Chapter 4

# Discussion

In the previous section, it was found that the contrast event-type gives the best results for modulated Gold codes and that the duration event-type gives the best results for random stimulation codes in terms of auto explained variance. In this section, I will talk about the limitations of this research and discuss my findings in relation to existing literature. I will come back to my hypothesis and conclude with a recommendation of things to look into in future works.

## 4.1 Limitations

The main limiting factor of this thesis is the small sample size ($n = 8$) for the random stimulation data set. This makes it difficult to come to meaningful conclusions on this data set. Although this thesis provides an indication that the duration event-type might gives the best results on all three scores, it is only statistically significant for the auto explained variance. Further research with a larger sample size would be needed to validate these findings.

## 4.2 Code families and future work

For various reasons, it is impossible to conclude from my analysis whether one code family is better than the other. This is because the difference between these two data sets is too big. Differences include the aforementioned channel layout, preprocessing and sampling rate, but also the number of participants, the participants themselves, the recording hardware, the trial length and the laboratory conditions. Ideally the same participants would perform the same task, just with a different code family.

Although it is not possible to conclude from this thesis whether modulated Gold codes or random stimulation codes are better, we can look at the existing literature to see which code families are commonly used in noise tagging BCI and how they compare.

Although m-sequences are most commonly used, there is no definite optimal choice of code family for BCI applications. Numerous less common code families have been used in research, including modulated Gold codes ([12], [13]), Almost Perfect Autocorrelation sequences ([15]) and random stimulation codes ([8], [9]). Isaksen et al. (2016) [5] suggests a true optimal code family may not exist, rather the best code is subject-dependent.

Torres and Daly (2021) [14] compared several analysis methods on six different code families, namely Almost Perfect Autocorrelation sequences, deBruijn sequences, Golay sequences, m-sequences, Gold codes and Kasami sequences. They found promising results for Almost Perfect Autocorrelation, deBruijn and Golay sequences, while m-sequences, Gold codes and Kasami sequences performed poorly in comparison. M-sequences did generally perform better than Gold codes. Based on their findings, in future work it would be worthwhile to investigate Almost Perfect Autocorrelation, deBruijn and Golay sequences as I investigated modulated Gold codes and random stimulation codes. In any case, it would be good for future work to move from m-sequences to one of these three promising code families.

## 4.3 Hypothesis

In my hypothesis, I stated the contrast and duration event-types would likely outperform the simple event-type, since they take into account nonlinearity while the simple event-type does not. This was found to be true for the modulated Gold code data set, but cannot be confirmed for the random stimulation code data set due to its small sample size. There is however indication this may be true for this code family as well.

This is also confirmed by looking at the transient responses. Contrast can be used to estimate the duration events well, while the simple event-type assumes linearity of the brain response and cannot estimate the duration event well.

I also expected random stimulation to have a lower accuracy overall for the duration event-type than modulated Gold codes, since Yasinzai et al. [16] found long flashes are not predicted with good accuracy. Because of the differences between the data sets, it is not appropriate to compare accuracies against each other. Although it was found the longer events added less relevant information to the template, it is likely this is because the longer events also occur less in random stimulation codes.

# Chapter 5

# Conclusion

In conclusion, I have shown that event-types that take nonlinearity into account explain the brain response better and give better performance than event-types that assume linearity. I have also provided an indication that the optimal event-type may differ per code family (contrast for modulated Gold codes and duration for random stimulation codes), which future research may wish to look into.

Furthermore, I have demonstrated that the contrast event-type requires less data to find a good fit while still being able to sufficiently model the response to long flashes, making it more practical than the duration event-type for codes that contain flashes of length 3 or more.

For this reason and because of its good performance in terms of classification accuracy and explained variances, it is the recommendation of this thesis that future research considers the use of the contrast event-type when designing cVEP BCI.

# Bibliography


[1] *Mindaffect releases open-source brain computer interface*, 2020, `https://www.mindaffect.nl/2020/11/23/mindaffect-releases-open-source-brain-computer-interface/`.

[2] Sylvain Baillet, Karl Friston, and Robert Oostenveld, *Academic software applications for electromagnetic brain mapping using meg and eeg*, Computational Intelligence and Neuroscience **2011** (2011), 972050.

[3] Almudena Capilla, Paula Pazo-Alvarez, Álvaro Darriba, Pablo Campo, and Joachim Gross, *Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses*, PloS one **6** (2011), e14543.

[4] Jane Huggins, Patricia Wren, and Kirsten Gruis, *What would brain–computer interface users want? opinions and priorities of potential users with amyotrophic lateral sclerosis*, Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases **12** (2011), 318–24.

[5] Jonas Isaksen, Ali Mohebbi, and Sadasivan Puthusserypady, *A comparative study of pseudorandom sequences used in a c-vep based bci for online wheelchair control*, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 1512–1515.

[6] Giancarlo Logroscino, Marco Piccininni, Benoît Marin, Emma Nichols, Foad Abd-Allah, Ahmed Abdelalim, Fares Alahdab, Solomon Asgedom, Ashish Awasthi, Yazan Chaiah, Ahmad Daryani, Phuc Huyen Do, Manisha Dubey, Alexis Elbaz, Sharareh Eskandarieh, Farzaneh Farhadi, Farshad Farzadfar, Seyed-Mohammad Fereshtehnejad, Eduarda Fernandes, and Stefan Lorkowski, *Global, regional, and national burden of motor neuron diseases 1990–2016: a systematic analysis for the global burden of disease study 2016*, The Lancet Neurology **17** (2018).

[7] Makoto Matsumoto and Takuji Nishimura, *Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number gen-*

*erator*, ACM Transactions on Modeling and Computer Simulation (TOMACS) **8** (1998), no. 1, 3–30.

[8] Sebastian Nagel, Wolfgang Rosenstiel, and Martin Spüler, *Random visual evoked potentials (rvep) for brain-computer interface (bci) control*, 09 2017.

[9] Sebastian Nagel and Martin Spüler, *World's fastest brain-computer interface: Combining eeg2code with deep learning*, bioRxiv (2019).

[10] Sebastian Nagel and Martin Spüler, *Modelling the brain response to arbitrary visual stimulation patterns for a flexible high-speed brain-computer interface*, PLOS ONE **13** (2018), e0206107.

[11] S. A. Patten, Gary A B Armstrong, Alexandra Lissouba, E. Kabashi, J. A. Parker, and P. Drapeau, *Fishing for causes and cures of motor neuron disorders*, Disease Models & Mechanisms **7** (2014), 799 – 809.

[12] Jordy Thielen, Philip Broek, J. Farquhar, and Peter Desain, *Broad-band visually evoked potentials: Re (con)volution in brain-computer interfacing*, PLoS ONE **10** (2015).

[13] Jordy Thielen, Pieter Marsman, Jason Farquhar, and Peter Desain, *From full calibration to zero training for a code-modulated visual evoked potentials brain computer interface*, Journal of Neural Engineering (2021).

[14] Juan Antonio Ramírez Torres and Ian Daly, *How to build a fast and accurate code-modulated brain-computer interface*, Journal of Neural Engineering **18** (2021), no. 4, 046052.

[15] Qingguo Wei, Yonghui Liu, Xiaorong Gao, Yijun Wang, Chen Yang, Zongwu Lu, and Huayuan Gong, *A novel c-vep bci paradigm for increasing the number of stimulus targets based on grouping modulation with different codes*, IEEE Transactions on Neural Systems and Rehabilitation Engineering **PP** (2018), 1–1.

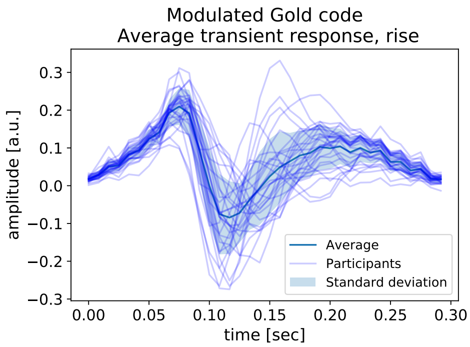
[16] M. Yasinzai and Y.Z. Ider, *New approach for designing cvep bci stimuli based on superposition of edge responses*, (2020).

# Appendix A

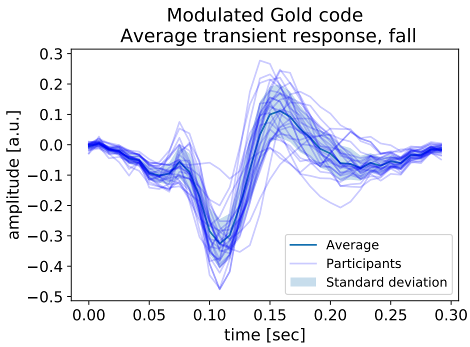
# Full results

This appendix contains the average transient response by event for each event-type and data set combination. Please note that the simple event-type has been excluded, since it only has one event which has already been shown entirely in the results section.

## A.1 Contrast event-type

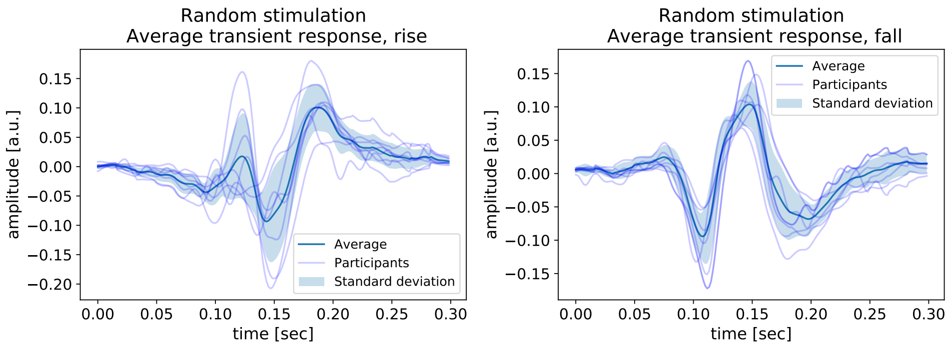



(a) Rise




(b) Fall

Figure A.1: Average transient responses per event for the contrast event-type with modulated Gold codes

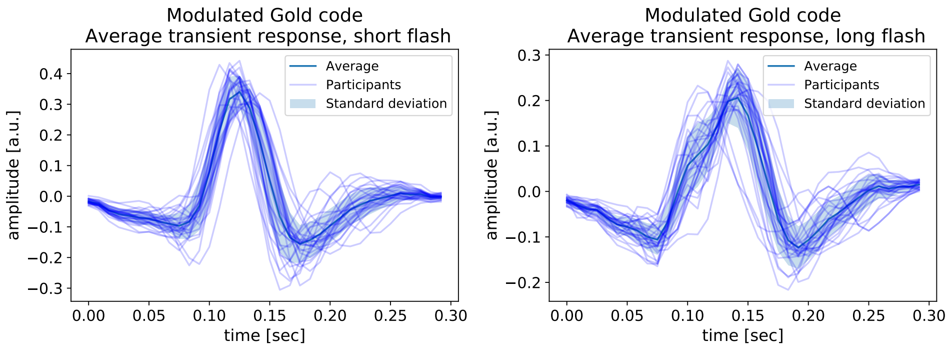



(a) Rise

(b) Fall

Figure A.2: Average transient responses per event for the contrast event-type with random stimulation codes

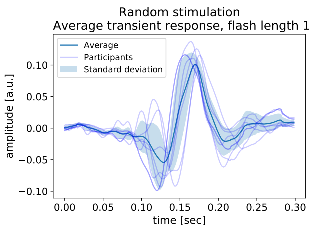
## A.2 Duration event-type



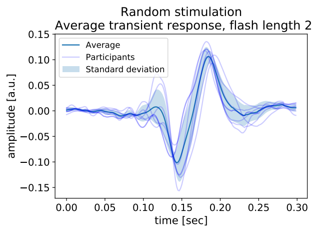

(a) Short flash

(b) Long flash

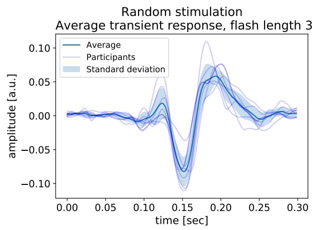
Figure A.3: Average transient responses per event for the duration event-type with modulated Gold codes



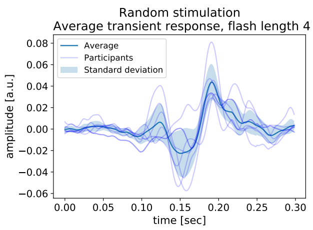

(a) Length 1



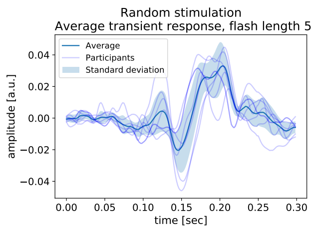

(b) Length 2



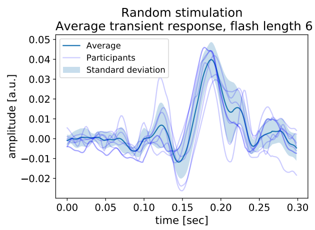

(c) Length 3



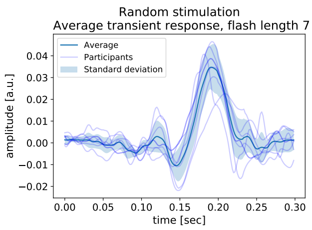

(d) Length 4



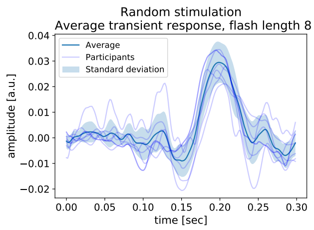

(e) Length 5



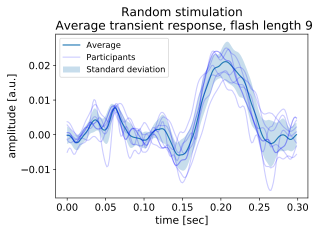

(f) Length 6



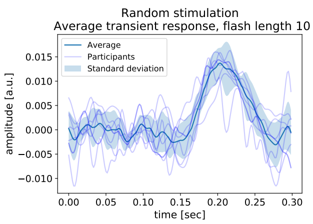

(g) Length 7



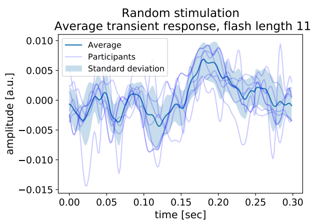

(h) Length 8




(i) Length 9




(j) Length 10



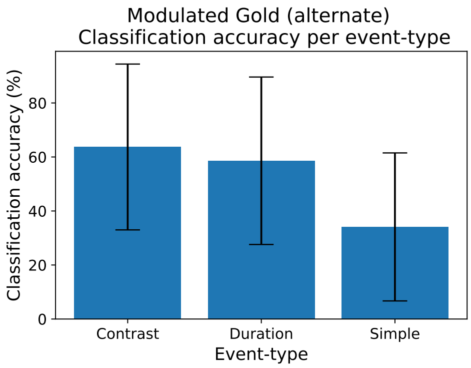

(k) Length 11

Figure A.4: Average transient responses per event for the duration event-type with random stimulation codes

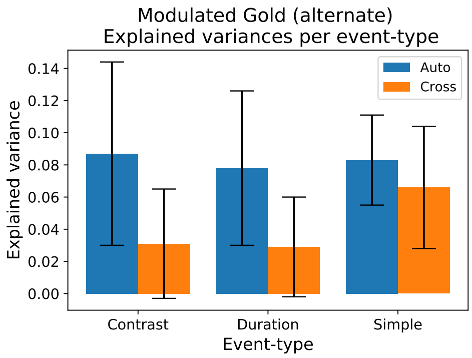
# Appendix B

# Alternate preprocessing method

This section contains the full results for the alternate preprocessing method for the modulated Gold codes data set.




(a) Classification accuracy




(b) Auto and cross explained variance

Figure B.1: Classification accuracy and explained variances per event-type for the alternate preprocessing method for modulated Gold codes

For the classification accuracy, the Wilcoxon rank test is used. Results are Bonferroni corrected. Here there is no statistical difference between contrast and duration ($p = 0.696, n = 30$). Their means are 63.7 and 58.6, respectively. Contrast significantly outperforms simple ($p < 0.001$), as does duration ($p = 0.009$). Simple has a mean of 34.7.

For the explained variances, paired t-tests are used. Results are Bonferroni corrected. For cross explained variance, contrast and duration again outperform simple. There is no significant difference between contrast and duration ($p = 0.78$). Their means are 0.031 and 0.029, respectively. Contrast significantly outperforms simple ($p < 0.001$), as does duration ($p = 0.009$). Simple has a mean of 0.064.

For auto explained variance, contrast significantly outperforms duration ($p = 0.003$). Their means are 0.087 and 0.078, respectively. There is no significant difference between contrast and simple ($p = 1$) or duration and simple ($p = 1$).

All in all, as contrast significantly outperforms duration in terms of auto explained variance while they are not significantly different on the other two scores, contrast is the preferred event-type. Duration is also a decent choice. Simple is significantly worse than the other two in terms of classification accuracy and cross explained variance, so it is not recommended.

These findings are the same as for the authors' original preprocessing method.

Since the scores for this preprocessing method are generally worse than for the authors' original method, it is likely the signal-to-noise ratio is worse for this method. I will not show the average transient responses or spatial filters for this method, as they will be less informative than the ones shown in the Results section.