

Bachelor's Thesis in Artificial Intelligence



Radboud University Nijmegen

The effects of color and contrast characteristics in NoiseTag BCI

Author:

J.A. ten Brinke
s4619587
Artificial Intelligence
Radboud University Nijmegen

Supervisors:

prof. dr. P.W.M. Desain
Donders Institute

dr. S. Ahmadi
Donders Institute

dr. J.D.R. Farquhar
Donders Institute

July 2018

Abstract

NoiseTag is a recently developed technology for measuring visually evoked brain stimuli and determining where a user is looking. It utilizes a computer display to present the end-user a set of buttons that will flash on and off, such that each button can be distinguished by a unique flashing pattern. The user focuses on the button to be selected and EEG recordings from the visual cortex are made. Based on the recorded signals, a classifier is used to determine which button the user was focussing on.

NoiseTag is amongst other applications useful for use with BCI-spellers, which allow physically impaired persons to write text by looking at a screen of flashing letters. Another application would for example be home automation, where different electronic devices can be operated using a BCI setup as an interface.

The objective of this research was to determine how the visual appearance of such a BCI-speller can be altered to improve its user experience and what effect such changes have on performance. Changing the colors of the stimuli turned out to have significant effects on both user experience and performance. However these effects appeared to be inversely related, making compromises a necessity.

Contents

1. Introduction	p.3
2. Brain Computer Interfaces	p.5
a. Overview	p.5
b. NoiseTag BCI	p.6
c. The MindAffect setup	p.7
3. Color perception & iso-luminance	p.8
4. Method	p.10
a. Design	p.10
b. Part A: Questionnaire	p.11
c. Part B: BCI EEG experiment	p.12
d. Data Analysis	p.14
i. Part A: Questionnaire data	p.14
1. Interpretation of questionnaire results	p.14
2. Statistical analysis	p.14
ii. Part B: BCI performance	p.14
1. Preprocessing	p.14
2. Calculating Characters Per Minute (CPM)	p.14
3. Statistical analysis	p.14
5. Results	p.15
a. Part A: Questionnaire data	p.15
i. Direct questionnaire results	p.15
ii. Interpretation of results and condition selection	p.17
iii. Statistical analysis	p.19
b. Part B: BCI performance	p.21
i. Preprocessed results	p.21
ii. Statistical analysis	p.22
6. Discussion & Conclusion	p.23
7. Acknowledgements	p.25
8. References	p.25
9. Appendix	p.26
a. Part A: Questionnaire	p.26

1. Introduction

Thanks to steady technological advancements, brain-computer interface systems (BCIs) and their applications are becoming easier to use, more reliable and practical. A traditional application of such systems is the BCI speller: a setup where physically impaired persons (for example, due to the effects of amyotrophic lateral sclerosis (ALS disease)) can communicate in a written fashion by focusing on flashing characters on a computer display. In a BCI setup, a measurement device, such as EEG measures brain activity and a previously trained classifier algorithm processes the measured data in relation to the input presented to the user in real time.

A traditional type of BCI-speller relies on the detection of a p300 signal[3] when the user observes that the desired character is flashing. This technique is reliable, but slow and tedious. A newer approach developed by Desain et al.[2] called "NoiseTag", measures visual EEG signals from the user (rather than p300). Moreover, it utilizes Gold codes[8], which originate in telecommunication for the efficient detection of radio signals, having good correlation properties. By modulating the flashing patterns using Gold codes, each selectable character has its own uniquely distinguishable flashing pattern. An EEG device measures electrical activity around the visual cortex and the data is classified by picking up the gold-code pattern of the character the user is focussing on. This new technique is considerably faster than traditional systems and more reliable.

In light of products based on NoiseTag getting ready to market, it is important to look at practical usability constraints. Anyone involved with BCI spellers can attest the reality that many end users complain about the user experience of BCI spellers: they oftentimes find them annoying, frustrating and tiresome to use. Little research has been done to investigate how the appearance of the flashing signals affects user experience as well as speller performance. The flickering is typically high-contrast in nature (e.g. black/white flashing), even though lowering this contrast may be more comfortable to look at. But simply lowering the contrast could adversely affect system performance. Similarly, using a variety of colors could potentially make for a more pleasant user experience, and perhaps even benefit performance as well.

To investigate the effects that changing the appearance of the flashing patterns has on user experience and system performance, the following research questions are stated:

Main research question: "In what ways is visual NoiseTag-BCI dependent on visual presentation characteristics and can such characteristics be used to further optimise NoiseTag BCI?" .

Sub-questions: "To what extent does altering the contrast between an "on" and an "off" signal affect performance as well as subjective comfort and fatigue of NoiseTag-BCI?", "To what extent does using different colors for the "on" and "off" signals affect performance as well as subjective comfort and fatigue of NoiseTag-BCI?".

Prior to the research, it is hypothesized that altering the contrast will affect performance of the BCI system. More specifically, lowering the contrast is hypothesized to make the EEG signal also

The effects of visual characteristics in NoiseTag BCI

less clear, therefore it becomes harder for the classifier to distinguish the signals which will ultimately result in longer classification times and/or lower overall accuracy. Based on personal experience, it is hypothesized that a lower contrast will generally be easier and more comfortable to look at, while extremely contrasting signals are fatiguing and unpleasant to look at. Changing the color of the signals is hypothesized to have an effect on the performance of the BCI system, where some color combinations for the on/off signals may improve performance, while others may deteriorate it. Some color combinations are likely more comfortable than others, and it would seem logical that comfort and performance be inversely related.

Importantly, different colors can produce different amounts of (perceived) luminance [7]. Therefore, it is important to separate the effects of luminance from the effects of hue on induced fatigue and comfort levels.

2. Brain Computer Interfaces

a. Overview

Brain computer interfaces (BCIs) allow living beings to directly communicate with a computer system solely by using their brains. This principle was first described by J.J. Vidal[11]. Today, many different types of BCI-systems exist, but they all consist of the same basic components.

The user is attached to a brain signal measurement system. One such system is Electroencephalography (EEG), which provides excellent temporal resolution and is non-invasive to the user. Other systems exist, such as an implanted electrode array, which provides more spatial resolution compared to EEG but is more invasive to install. In the case of EEG, electrical signals are measured on the outside of the skull using a number of electrodes, which are individually connected to a data acquisition device. This device makes a digital recording of the signals that is subsequently sent towards the computer system. The computer system is tasked with fitting a classifier to the raw signals received from the data it receives, and subsequently using the classifier to recognize the input signals in a meaningful way (for example, detecting where a subject is looking on a monitor). This is made possible through a feedback loop, such as a visual feedback loop by means of a monitor. The computer presents a visual stimulus on the monitor, and subsequently detects how the data generated by the test subject responds. This basic operation of a BCI system is visually presented by the figure below:

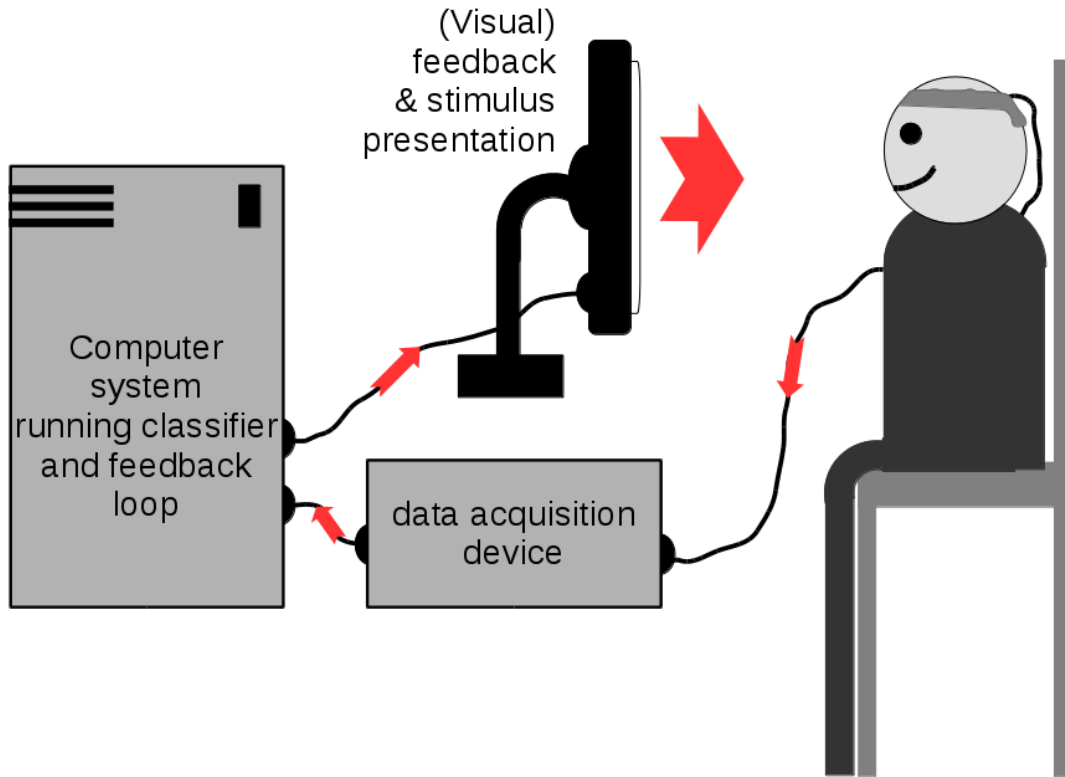


Figure 1: Basic components of a BCI system, arrows describe the flow of information

b. NoiseTag BCI

BCI-systems need to modulate the stimuli presented to the user in some way, for example by flashing an object on a computer display. One of the most common traditional ways of modulating the stimuli is through frequency tagging[9]. This modulation technique uses a fixed frequency for each stimulus (ie. by flashing an object on the screen at a particular frequency), and detection of a user's attention to the signal is therefore also dependent on this frequency-specific signal. This is a very straightforward and relatively simple system, but it is also very prone to noise as described by Farquhar et al.[2].

NoiseTag is a modulation system originally described and developed by Desain et al.[10]. Rather than relying on the production and detection of frequency-specific signals, it takes somewhat the other approach. By applying a spread spectrum-type of signal, using a wider band of modulation frequencies for stimuli, this system is more robust to noise that obscures a part of the frequency spectrum. It uses so-called Gold codes[8], originally developed for use in mobile communications, to modulate the stimuli signals in such a way that different stimuli are maximally uncorrelated to one another and optimally detectable.

The NoiseTag system is not only more robust to noise, but it is also significantly faster in terms of information transferring whilst maintaining excellent classification accuracy[10].

The effects of visual characteristics in NoiseTag BCI

c. The MindAffect Setup

This research project is based on a particular implementation of NoiseTag BCI, developed by the MindAffect company based in Nijmegen. MindAffect has developed a speller system using visual NoiseTag BCI. The setup consists of an Apple iPad which serves as the stimulus presentation and visual feedback device, a specially developed EEG-cap that utilizes 8 electrodes in conjunction with the TMSi Porti and a PC running the classifier and feedback software. In addition, an optical sensor is attached to a corner of the iPad's display and connected to the Porti, for use as a triggering and visual delay detection device.

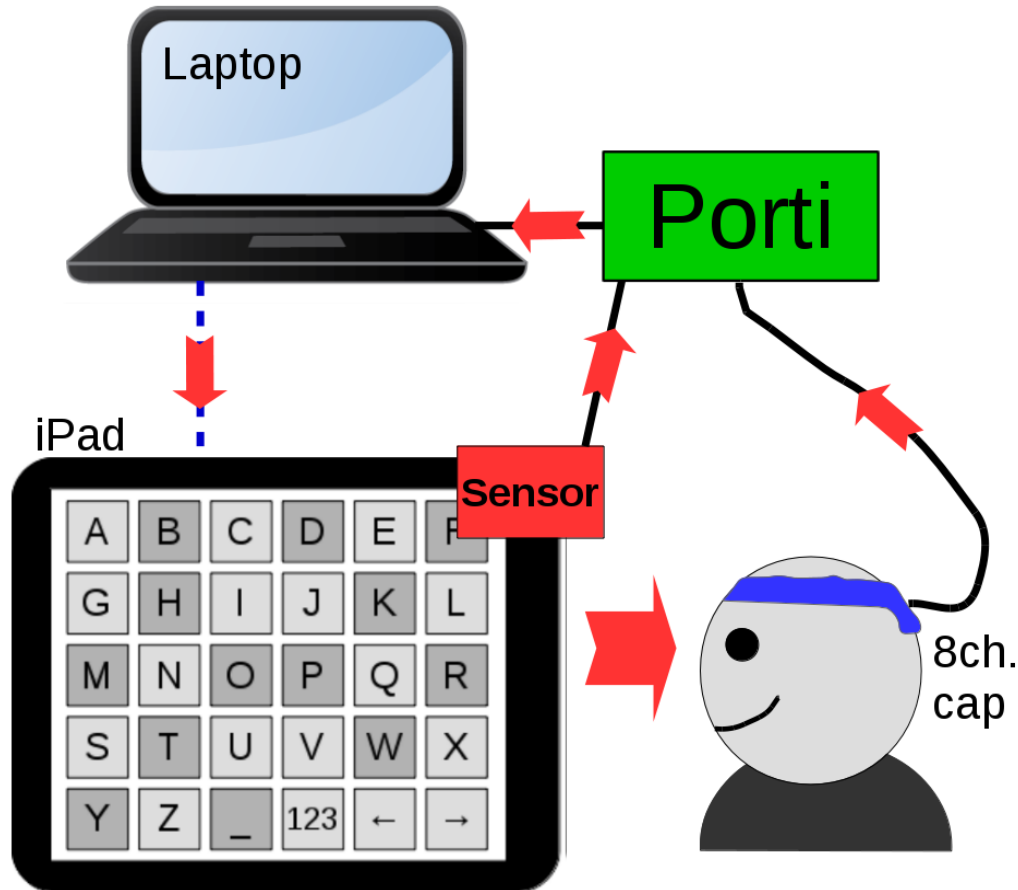


Figure 2: Diagram of the MindAffect Setup, arrows describe information flow

The iPad shows a graphical keyboard and communicates wirelessly with the PC over Bluetooth. During operation, the PC sends sequences of NoiseTag stimuli to the iPad, which will display them in an accurately timed manner, along with a constant trigger pulse which is recognized by the optical sensor. This results in rapidly flickering keys on the graphical keyboard. The user sits directly in front of the iPad's display, wearing the EEG cap. During the training phase, the user focuses on characters as requested by the system (buttons will temporarily turn green and "wiggle" to indicate they are the next target). The EEG signals are recorded by the Porti and the classifier is subsequently fitted based on this training data. After this training cycle, the system is ready to classify any character the user is looking at.

The effects of visual characteristics in NoiseTag BCI

On the PC-side, the incoming data is preprocessed prior to classification. First, a low-pass filter is applied that attenuates frequencies above 50 Hz by 50 dB. Next, a high-pass filter is applied that cuts off any frequencies below 2 Hz. After filtering, the signal is downsampled to 360 Hz. The classifier compares the input data to EEG-templates that describe the EEG response to particular visual stimuli. These EEG-templates are generated through reconvolution[1], a generative modelling process that is applied during the training phase of the BCI-setup.

3. Color perception & iso-luminance

Whenever the effects of different colors are studied, using the primary RGB colors (pure red, green and blue) may seem the most obvious path to take. However, matters get a bit more complicated than that. First of, as described by [4], one should be wary of using a pure red color in a fast flickering application, as it tends to produce odd responses in test subjects and "can even elicit epileptic responses". Furthermore, the color yellow is suggested by [4] as a substitute for red. However, there is another caveat. As the intent is to be able to study the effects of color and contrast individually, it is therefore necessary that all the colors are presented to the test subject at the same brightness, or to be more precise: the different colors must appear identical in light strength to the test subject, in order to isolate from the effects of contrast difference[7].

As described by [6] and [7], color perception by human beings is irregular. Some colors are perceived stronger than others, even though they have exact equal light strength. Furthermore, the type of display used also changes the light frequency response[7] (e.g. CRT display versus LCD). It is therefore necessary to use "isoluminant" colors[5]. Such colors are at the same perceived light strength, taking into account human perception characteristics as well as display device characteristics.

The following formula is used to calculate the perceived signal strength (measured as "luma (Y)") [7] for sRGB monitors (such as the iPad's display) given an RGB value:

$$Y'_{601} = 0.299R + 0.587G + 0.114B$$

The following table is constructed using this formula:

Color	RGB*	Luma
Pure red	1R 0G 0B	0.299
Pure blue	0R 0G 1B	0.114
Pure green	0R 1G 0B	0.587
Pure yellow	1R 1G 0B	0.886
Pure magenta	1R 0G 1B	0.343

*Note: the notation "1R" means "100% red", "0G" means "0% green", etc.

Table 1: Luma values of pure colors

The effects of visual characteristics in NoiseTag BCI

It follows from these luma values that, in order to achieve isoluminance between these colors, it is necessary to lower the luminance of green, yellow and magenta (red is not used from here on) to the same level as blue (blue can't be increased beyond 100%, after all). Therefore, it is necessary to decrease the luma values of green, yellow and magenta to 0.114, whilst retaining their hue and saturation. To achieve this, the RGB values must be lowered in such a way that the relative strength of the three color components remains the same. An example calculation for the iso-luminance conversion of the color magenta is given below.

Modifying pure magenta to become iso-luminant is done as follows:

$$0.114 = 0.299a + 0.114a$$

(Here the R and B components are both equal to the value of a)

Therefore the RGB values for iso-luminant magenta are:

$$0.114 = 0.413a$$

$$R = B = a = \sim 0.2760$$

$$G = 0$$

Using these calculations, the following table is constructed:

Color	RGB (decimal)	RGB (8-bit)	Luma
iso-lum. red	0.3813R, 0G, 1B	97R, 0G, 0B	0.114
iso-lum. blue	0R, 0G, 1B	0R, 0G, 255B	0.114
iso-lum. green	0R, 0.1942G, 0B	0R, 50G, 0B	0.114
iso-lum. yellow	0.1287R, 0.1287G, 0B	33R, 33G, 0B	0.114
iso-lum. magenta	0.2760R, 0G, 0.2760B	70R, 0G, 70B	0.114

Table 2: iso-luminant colors

Below are the resulting adjusted colors:



Figure 3: From left to right: blue, red, green yellow, magenta (all iso-luminant)

Because the yellow color clearly does not scale down in luma without turning brownish, magenta is used instead for the experiment.

4. Method

a. Design

For the experiments, conditions are defined as certain combinations of on/off colors used by the speller-application. These on/off colors "fill" the buttons on the screen except for the key-characters which are always displayed on top of these colors in grey. Initially, there are a total of 10 conditions, including a baseline. The experimental (non-baseline) conditions are divided into three sets.:

Default/baseline condition:

White 100%/white 0% (i.e. "white/black")

Color-color contrast set:

1. Green/blue
2. Green/magenta
3. Blue/magenta

Color-black contrast set:

4. Green/black
5. Blue/black
6. Magenta/black

Grayscale contrast set*:

- 8.** White 85%/white 15%
9. White 70%/white 30%
10. White 55%/white 45%

* After running experiment part A for the first time (section 4.b.), it was decided to change the color of the characters displayed on the buttons (ie. the letters themselves) from the default grey to yellow for the set of grayscale-conditions to improve legibility. It became apparent that the characters became unreadable during these conditions, therefore it became necessary to make this adjustment. To compensate for any influence of this change on the results, a new baseline was presented to the test subjects prior to running the set of grayscale conditions, which was identical to the original baseline except the character color was changed to yellow. The data from the first test-subject in this part of the experiment was discarded.

**Originally, there was a condition 7, but this was later (before the experiments commenced) removed. To avoid confusion with the condition numbering, the original numbering scheme has remained in place.

It should be noted that in case of the colors green, blue and magenta, isoluminant RGB values are used, rather than "pure" values, to compensate for the irregular color perception characteristic of human beings. This is discussed in section 3.

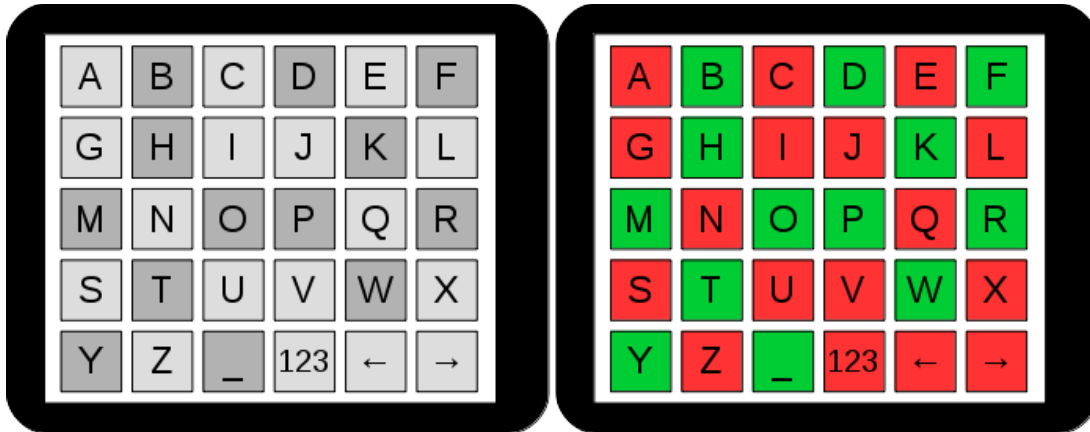


Figure 4: Illustration of the iPad display during two experiments: grayscale contrast to the left, different colors to the right.

The experiment is divided into two parts, which occur at different points in time. The purpose of the first part (part A) is to determine how changing from the baseline condition to any other condition affects perceived comfort and fatigue in users, as well as to compile a subset of conditions that are subsequently applied in the second part (part B). Part A involves data collection through a questionnaire. In part A no BCI interface is connected. In part B, a full BCI-setup is used to process EEG-data and run the classifier. Both parts of the experiment will be discussed in detail in the next subsections.

b. Part A: Questionnaire

In this part of the experiment, a number (5) of test subjects are presented with all of the aforementioned conditions. That is, they are watching the iPad's display as the speller application is running, and they are tasked to focus on letters as if they were trying to use the application. No actual EEG measurements are taken at this time, only behavioural data is acquired. For each condition, starting with the base condition, the subject focuses on 5 (random) characters in sequence, as told by the researcher, taking approximately 4 seconds per letter. For the base condition, the subject is told to assume comfort and fatigue scores of 6 out of 10 and 5 out of 10 respectively. For the remaining conditions, the subject rates induced fatigue and level of comfort relative to the baseline on scales of 0 to 10 on a supplied questionnaire. In addition, any supplementary feedback is also recorded, should the subject have any remarks on a condition or the experiment in general. After each set of conditions, the subject takes a short break of approximately 3 minutes.

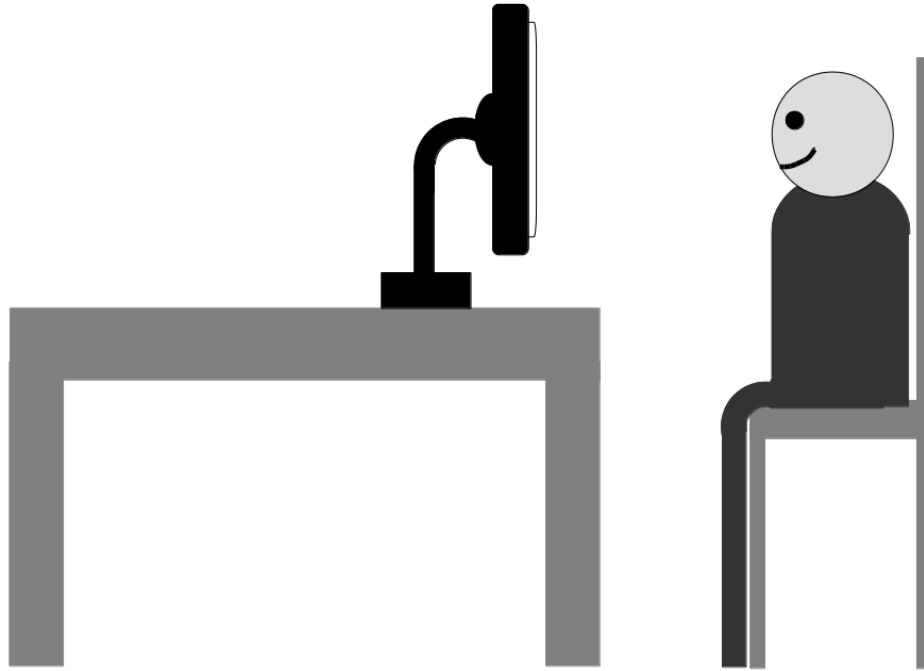


Figure 5: Experiment Part A setup.

c. Part B: BCI EEG Experiment

Even though ideally all conditions are tested for performance, due to time constraints a limited selection is to be made. For part B of the experiment, the three most *interesting* conditions are selected (interesting in the sense that they have the most favourable comfort and fatigue levels). In this case, it was concluded that conditions C6 and C9 were the most interesting to pursue in part B, along with the baseline condition (see section 5.ii). Next, 5 test subjects are subjected to these conditions, and their BCI performance is measured. To do so, the test subjects are connected to the BCI system as described in section 2.c. For each condition, they first undergo a training session to fit and optimize the classifier, and then they do two verification (performance) tests. This will now be described in more detail.

During training, the BCI system presents a random set of characters on the screen that the user subsequently has to focus on. Each character first lights up green and wiggles, giving the user time to locate it. Then, the NoiseTag sequence starts running: the buttons start flashing. After training has completed, the optimization process is performed, in which the system fits the classifier according to the data acquired through the training cycle.

For the verification tests, the subjects are tasked to write out a sentence of 20 characters. This sentence is always the same: "the brown fox jumped". The target buttons are again indicated by turning green and wiggling. During this test, it is recorded for every character in the sentence whether it was correctly recognized and how long it took the classifier to recognize the character from the moment of stimulus onset ("classification time"). The verification step is performed twice per condition. Each condition (training + optimization + verification*2) is followed by a 3 minute break. For each condition, if large discrepancies in performance between two related verification tests are encountered, a third test is run to compensate for measurement errors (due to bad electrodes, other physical/software issues, etc.) during data analysis.

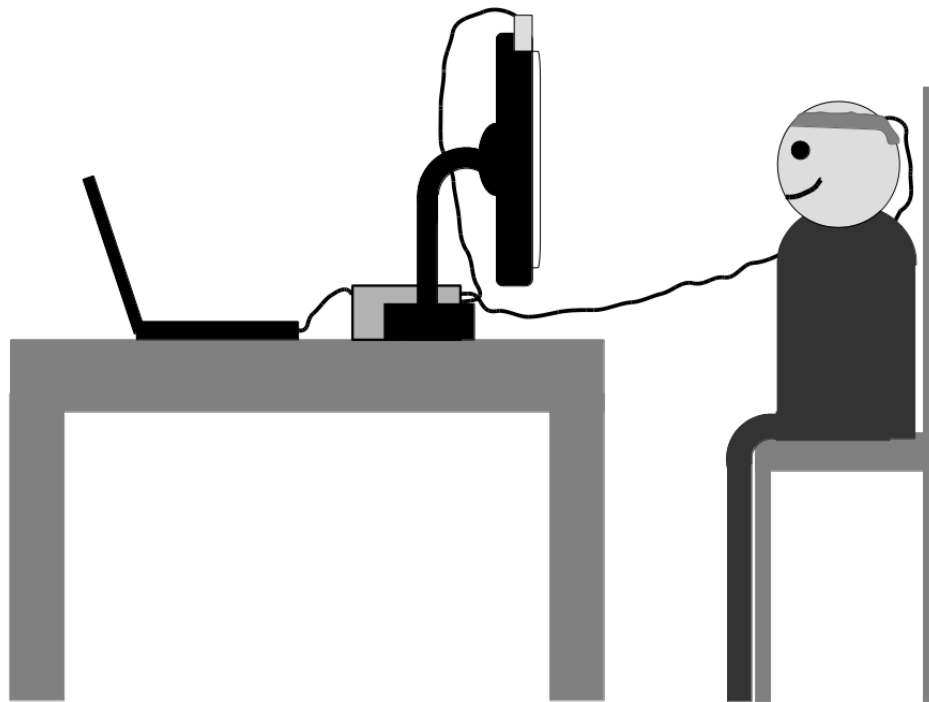


Figure 6: Experiment Part B setup.

d. Data Analysis

i. Part A: Questionnaire data

1. Interpretation of questionnaire results

Based on the data gathered from the questionnaires, figures are constructed that describe how comfort and fatigue vary between conditions for each subject and on average. These are then interpreted to determine which conditions should be applied in the second part of the experiment.

2. Statistical analysis

The data from the questionnaires are analyzed using a Double Multivariate Repeated Measures ANOVA using condition as the factor, comfort and fatigue scores as the dependent variables.

ii. Part B: BCI performance

1. Pre-processing

Based on the data recorded during the experiment, for each subject, for each condition, it is determined how many characters (out of 40 in total) are recognized correctly. The total time taken for each condition is also calculated by summing all the classification times for all characters and during both verification runs. Note that when no character is recognized, the classifier times out after 4.2 seconds, so every time no character is recognized, the classification time equals 4.2 seconds.

2. Calculating Characters Per Minute (CPM).

Characters Per Minute (CPM) is defined as the number of correctly classified characters per minute. This measure takes into account the time wasted by incorrect and failed classifications. CPM is calculated using the following formula:

$$CPM = 60 \times \frac{c}{t}$$

Where c is the number of correctly classified characters and t the total classification time for the condition.

For each subject and each condition, the corresponding CPM value is calculated.

3. Statistical analysis

Since the experimental design has a single dependent variable (CPM) and focuses on a within subject factor (condition), a 1-factor repeated measures ANOVA is chosen for determining significant effects. The results are used to evaluate if any effects of changing from the baseline condition to a different condition are statistically significant.

The effects of visual characteristics in NoiseTag BCI

5. Results

a. Part A: Questionnaire data

i. Direct questionnaire results

The following tables entail the raw comfort and fatigue scores given by the test subjects in part A of the experiment, for all conditions. It should be noted that in case of subject number 1, the data from the set of grayscale conditions (c8 through c10) is discarded. This is because issues with legibility caused erratic performance: for the remaining experiments, the character color was changed to yellow during these conditions (section 4.a).

Subject	BASE	C1	C2	C3	C4	C5	C6	C8	C9	C10
1	6	7	8	7	8	7	7			
2	6	8	9	7	7	7	8	7	7	8
3	6	9	8	8	8	7	9	7	8	9
4	6	8	9	9	8	7	9	8	9	10
5	6	8	9	7	7	7	9	7	8	9
Average	6	8	8.6	7.6	7.6	7	8.4	7.25	8	9

Table 3: Comfort scores

Subject	BASE	C1	C2	C3	C4	C5	C6	C8	C9	C10
1	5	4	3	4	3	4	4			
2	5	4	2	4	4	4	3	5	4	4
3	5	4	3	4	3	4	2	6	4	2
4	5	3	2	2	2	4	2	3	2	2
5	5	2	1	3	3	4	1	4	3	3
Average	5	3.4	2.2	3.4	3	4	2.4	4.5	3.25	2.75

Table 4: Fatigue scores

The scores are also represented by the following graphs:

Comfort scores

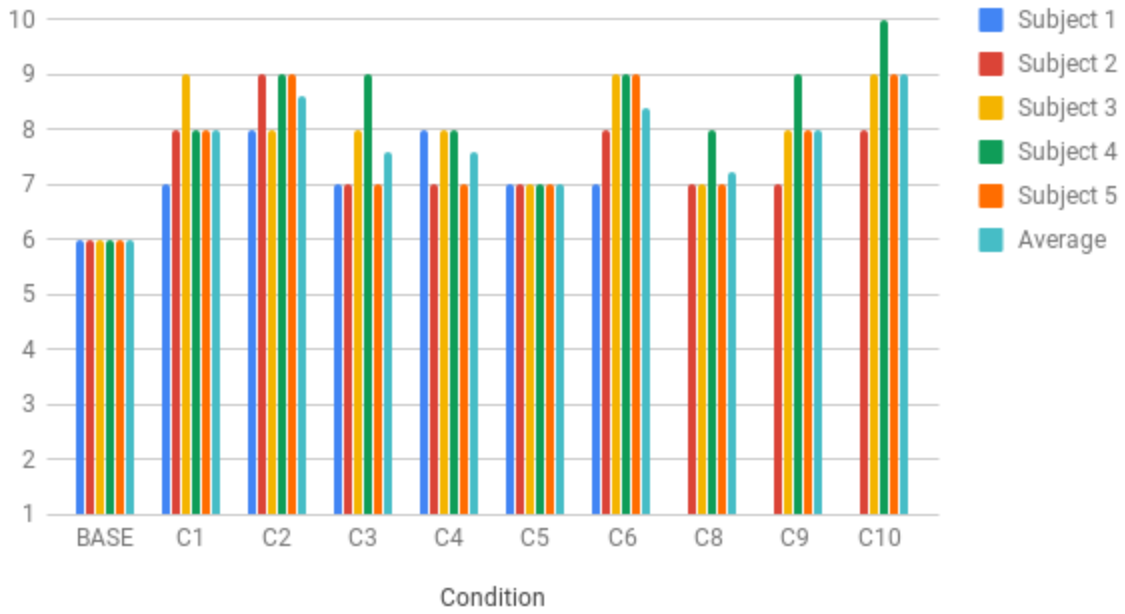


Figure 7: Comfort scores

Fatigue scores

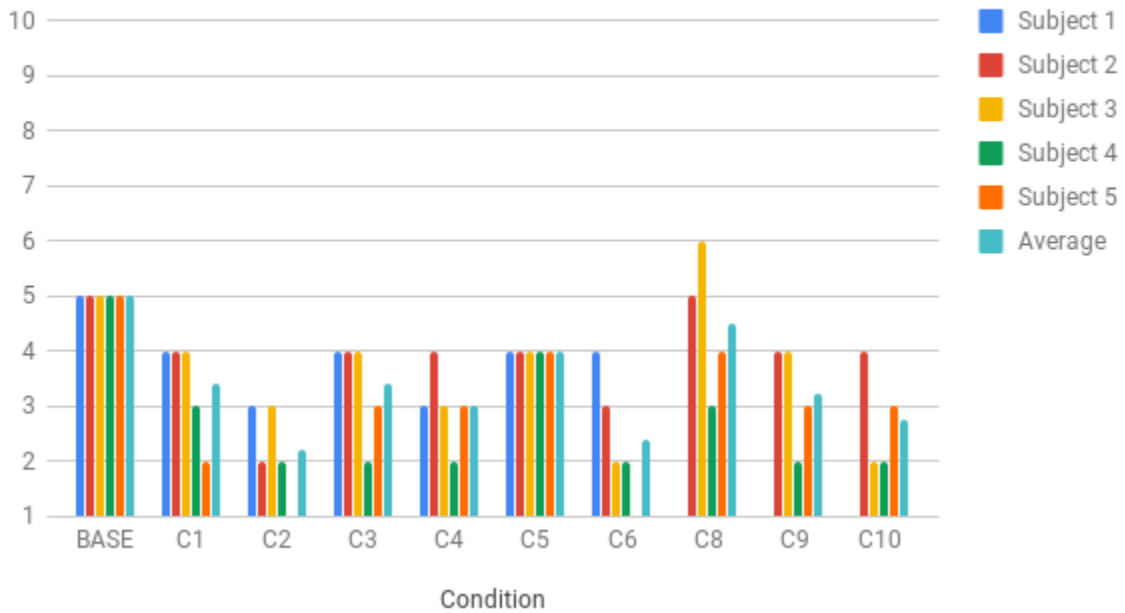


Figure 8: Fatigue scores

The effects of visual characteristics in NoiseTag BCI

In addition to these data, the following remarks were given by the test subjects during the experiment:

Subject	Remark
1	<ul style="list-style-type: none"> - During color trials: conditions containing blue are less comfortable compared to other colors - During grayscale trials: the flickering became less severe, but text became unreadable
2	<ul style="list-style-type: none"> - During color trials: C2 was particularly comfortable: it reminded of "parelmoer" of a conch held in the sun, while looking through the water
3	<ul style="list-style-type: none"> - During color trials: the blue color was less comfortable, it is relatively bright. Less contrast is preferred. - During grayscale trials: yellow letters are awesome
4	<ul style="list-style-type: none"> - The second baseline: using yellow letters is very fatigue-inducing (even though it is set as a baseline at 6 for comfort and 5 for fatigue) - Please use a qwerty layout instead.
5	<ul style="list-style-type: none"> - Lower contrast between colors is more comfortable to look at.

Table 5: Remarks by test subjects

ii. Interpretation of questionnaire results and condition selection

This section discusses describes how the final selection of condition used during part B of the experiment is compiled. Based on the data gathered, the following conclusions are drawn with respect to comfort and fatigue-levels:

- The most comfortable color-color condition is C2: green/magenta.
- The most comfortable color-black condition is C6: magenta/black.
- From the color conditions, the baseline (black/white) is universally rated least comfortable.
- From the color conditions, the baseline (black/white) is universally rated most fatiguing.
- From the color conditions, C2: green/magenta appears to be the least fatiguing condition.
- Clearly, subject 1's grayscale data is skewed by the effect of the text being gray, whereas the text color for other subjects during these conditions was yellow.
- Ignoring subject 1, in the grayscale conditions there is a clear correlation between contrast and comfort levels: a lower contrast is more comfortable to attend. As a result, C10: 55%w/45%w is universally rated the most comfortable grayscale condition, whereas the baseline (black/white) is the least comfortable.
- Similarly (once again ignoring subject 1), in the grayscale conditions there is a clear correlation between contrast and fatigue: a lower contrast is less fatiguing. As a result, C10: 55%w/45%w is the least fatiguing grayscale condition, whereas the baseline (black/white) is the most fatiguing.
- In general: comfort and fatigue appear to be correlated: they seem inversely proportional.

The following scatterplot displays the inverse relationship between comfort and fatigue levels more clearly. It is constructed based on the comfort score/fatigue score couples from tables 3 and 4 (except for averages). In addition, a trend line is fitted.

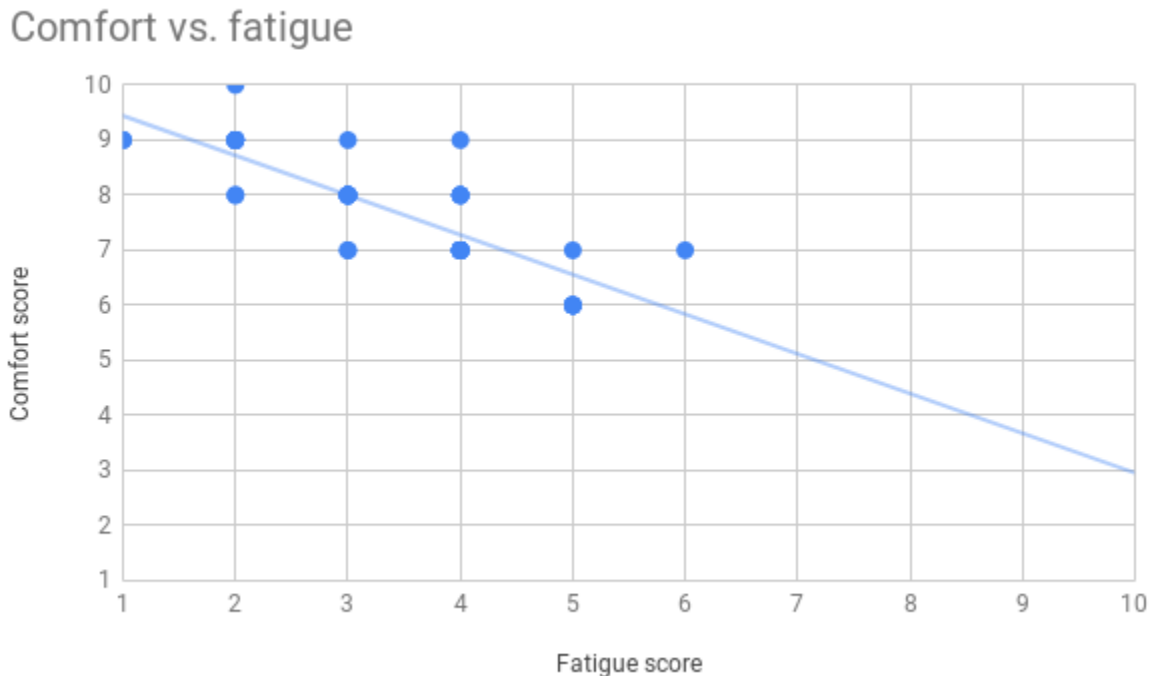


Figure 9: Scatterplot comfort and fatigue scores.

The following set of conditions is selected to be further studied in part B (rationale below):

- C6 (magenta/black), C9 (white 70%/white 30%), Base (white 100%/white 0%, with gray letters)

C2 (green/magenta) and C6 (magenta/black) are both rated highly comfortable color conditions, yet C2 is slightly less fatiguing. Therefore, it is interesting to measure the performance of C2. C10 (55% white/45% white) is clearly the most comfortable/least fatiguing grayscale condition, so it is interesting to see how it performs relative to the baseline condition. However, after evaluating the first test-subjects performance during part B of the experiments, it became clear that C2 and C10 yielded near-zero performance (unusable and uninformative). Therefore, C6 and C9 were used instead for the remaining test-subjects, and the first subjects data was replaced by that of a new subject. Finally, a baseline was necessary to compare the other conditions to, hence the inclusion of the base-condition.

iii. Statistical analysis

A Double Multivariate Repeated Measures ANOVA is performed, with the condition as factor and comfort, fatigue as the dependent variables. Subject number 1's data pertaining to condition 8 through 10 are missing as described before (sections 4.a and 5.a.i).

	Mean	Std. Deviation	N		Mean	Std. Deviation	N
baseComfort	6.00	.000	5	baseFatigue	5.00	.000	5
c1Comfort	8.00	.707	5	c1Fatigue	3.40	.894	5
c2Comfort	8.60	.548	5	c2Fatigue	2.20	.837	5
c3Comfort	7.60	.894	5	c3Fatigue	3.40	.894	5
c4Comfort	7.60	.548	5	c4Fatigue	3.00	.707	5
c5Comfort	7.00	.000	5	c5Fatigue	4.00	.000	5
c6Comfort	8.40	.894	5	c6Fatigue	2.40	1.140	5
c8Comfort	7.25	.500	4	c8Fatigue	4.50	1.291	4
c9Comfort	8.00	.816	4	c9Fatigue	3.25	.957	4
c10Comfort	9.00	.816	4	c10Fatigue	2.75	.957	4

Tables 6, 7: Descriptive statistics

Measur e	Condition	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Comfort	C1 vs. base	20.000	1	20.000	40.000	.003	.909
Comfort	C2 vs. base	33.800	1	33.800	112.667	.000	.966
Comfort	C3 vs. base	12.800	1	12.800	16.000	.016	.800
Comfort	C4 vs. base	12.800	1	12.800	42.667	.003	.914
Comfort	C5 vs. base	5.000	1	5.000	0	0	1.000
Comfort	C6 vs. base	28.800	1	28.800	36.000	.004	.900
Comfort	C8 vs base	6.250	1	6.250	25.000	.015	.893
Comfort	C9 vs base	16.000	1	16.000	24.000	.016	.889
Comfort	C10 vs base	36.000	1	36.000	54.000	.005	.947
Fatigue	C1 vs. base	12.800	1	12.800	16.000	.016	.800

The effects of visual characteristics in NoiseTag BCI

Fatigue	C2 vs. base	39.200	1	39.200	56.000	.002	.933
Fatigue	C3 vs. base	12.800	1	12.800	16.000	.016	.800
Fatigue	C4 vs. base	20.000	1	20.000	40.000	.003	.909
Fatigue	C5 vs. base	5.000	1	5.000	0	0	1.000
Fatigue	C6 vs. base	33.800	1	33.800	26.000	.007	.867
Fatigue	C8 vs base	1.000	1	1.000	.600	.495	.167
Fatigue	C9 vs base	12.250	1	12.250	13.364	.035	.817
Fatigue	C10 vs base	20.250	1	20.250	22.091	.018	.880

Table 8: Tests of Within-Subject Contrasts

Measure	Condition	Mean	Std. Error	Measure	Condition	Mean	Std. Error
Comfort	Base	6.000	.000	Fatigue	Base	5.000	.000
Comfort	C1	8.000	.316	Fatigue	C1	3.400	.400
Comfort	C2	8.600	.245	Fatigue	C2	2.200	.374
Comfort	C3	7.600	.400	Fatigue	C3	3.400	.400
Comfort	C4	7.600	.245	Fatigue	C4	3.000	.316
Comfort	C5	7.000	.000	Fatigue	C5	4.000	.000
Comfort	C6	8.400	.400	Fatigue	C6	2.400	.510
Comfort	C8	7.250	.250	Fatigue	C8	4.500	.645
Comfort	C9	8.000	.408	Fatigue	C9	3.250	.479
Comfort	C10	9.000	.408	Fatigue	C10	2.750	.479

Tables 9, 10: Estimated Marginal Means

From these results it is evident that any mentioned change in condition has a significant effect on comfort, as all the corresponding significance values are smaller than the statistical border of 0.05 (see "Tests of Within-Subject Contrasts"). Combined with the means from the descriptive statistics, it can be stated that any of the experimental conditions is significantly more comfortable than the baseline condition.

Furthermore, all the experimental conditions, except C8, have significant effects (sig.<0.05) on fatigue, compared to the baseline. By also looking at the mean values for these conditions, it is concluded that all experimental conditions, except C8, are significantly less fatiguing compared to the baseline condition. Although C8 does appear to be less fatiguing than the baseline condition (according to the mean value), this effect is not significant.

The effects of visual characteristics in NoiseTag BCI

Part B: Analysis of BCI performance

1. Preprocessed results

The following table is constructed based on the raw results. "BaseCor." is the number of correctly classified characters during the base condition, "BaseTime" the total classification time during the base condition in seconds (both verification runs combined, as described in section 4.d.ii.1), etc.

Subject	BaseCor.	BaseTime	C6Cor.	C6Time	C9Cor.	C9Time
1	40	67.1	35	68.0	39	65.8
2	37	78.8	29	105.2	36	92.7
3	33	87.8	4	159.6	14	148.9
4	34	103.6	22	136	20	127.3
5	27	110.1	18	135	11	149
Mean	34.2	89.48	21.6	120.76	24	116.74

Table 11: BCI classification results

Based on the above, CPM values are given below:

Subject	BaseCPM	C6CPM	C9CPM
1	35.8	30.9	35.6
2	28.2	16.5	23.3
3	22.6	1.5	5.6
4	19.7	9.7	9.4
5	14.7	8	4.4
Mean	24.2	13.3	15.7

Table 12: Characters Per Minute (CPM)

2. Statistical analysis

A 1-Factor Repeated Measures ANOVA is performed, with condition as factor and CPM as the dependent variable.

	Mean	Std. Deviation	N
CPMbase	24.18	8.10	5
CPMc6	13.33	11.18	5
CPMc9	15.67	13.41	5

Table 13: Descriptive Statistics

Measure	Condition	Type III Sum of Squares	d	Mean Square	F	Sig.	Partial Eta Squared
CPM	C6 vs base	588.92	1	588.92	14.91	.018	.788
CPM	C9 vs base	361.85	1	361.85	9.10	.039	.685

Table 14: Tests of Within-Subjects Contrasts

Measure	Condition	Mean	Std. Error
CPM	Base	24.18	3.62
CPM	C6	13.33	4.99
CPM	C9	15.67	5.99

Table 15: Estimated Marginal Means

The effect of changing between the baseline (black/white) condition and C6 (magenta/black) is significant: C6 yields a significantly lower CPM value compared to the baseline. The effect of changing between the baseline condition and C9 (white 70%/white 30%) is similarly significant, if slightly less so: C9 yields a significantly lower CPM value compared to the baseline.

6. Discussion & Conclusion

From all the conditions that are applied in part A of the experiment, the base condition of pure black/white pulses is considered by subjects to be the least comfortable. Any of the other conditions is significantly more comfortable.

When it comes to induced fatigue, it is apparent that the default condition is reported to be significantly more fatiguing compared to any other tested condition, except condition C8 (white 85%/white 15%). Although this condition does appear to be less fatiguing based on figure 8 and the mean data in table 7, this effect is insignificant. An explanation is that this condition is very similar to the default condition, having only a slightly lower contrast, which makes the difference between it and the base condition too small to be significant.

There appears to be a direct relation between contrast and induced fatigue: a higher contrast level yields higher levels of induced fatigue (figure 8). Furthermore, there seems to be an inverse relation between contrast and viewing comfort: a higher contrast level is less comfortable for the user. This relation can be observed clearly in figure 9. Indeed, subjects 1, 3 and 5 have stated their preference for lower contrast conditions (table 5). It should be noted that these statements are made based on figures 7, 8, 9 and tables 6, 7, but the described effects have not been tested for statistical significance.

Altering the colors used for signaling on/off pulses in the NoiseTag BCI speller has a significant effect on the performance of the BCI speller. When switching from the baseline condition to either one of the two experimental conditions (C6 (magenta/black) and C9 (white 70%/white 30%)), the performance of the BCI system, as measured in CPM, degrades significantly. Therefore, the conclusion is drawn that altering the colors will at least in some cases cause significant performance differences.

The significant discrepancy between performance (measured in CPM) of the baseline condition and C9 suggests that purely lowering the contrast of the signals causes a significant performance degradation. Furthermore, the discrepancy between the performance of the baseline, C6 and C9, in relation to the effects of those conditions on comfort levels, suggests that system performance is inversely related to perceived comfort. However, these effects should be studied using a larger variety of contrast levels (at least the complete grayscale set mentioned in this report) and colors, before drawing conclusions, to evaluate what sort of pattern (e.g. linear) they follow.

Some unexpected observations will now be discussed. First of, as mentioned by subject 1 and 3 (table 5), the blue color appeared to be brighter compared to the other colors. It appears that even though colors were adjusted for isoluminance, this may still have yielded an imperfect luma balance. This issue could potentially be resolved in future research by generating a more specific formula than the one discussed in section 3, optimized for the specific setup used. This may require the use of a colorimeter to profile the particular display device used.

Another important point to be made is that subject 4 complained about the use of a yellow color

The effects of visual characteristics in NoiseTag BCI

for the on-screen characters during the grayscale conditions, while subject 3 proclaimed his preference for yellow characters. This color was grey by default but was changed to yellow to improve legibility during these specific conditions. This research in principle only focussed on the effects of changing the on/off flashing signal colors, but it could prove interesting to study the effects of other aspects, such as the colors of the characters and the background color on comfort and fatigue as well as performance.

Finally, getting back to the main question at hand: The performance of visual NoiseTag-BCI is highly dependent on the colors used for the on/off signaling. Moreover, the level of comfort and induced fatigue to the user is also highly dependent on the particular set of colors used. By changing these colors, it is possible to optimize the BCI system to either become more comfortable to the user, or to become better performing. However, at least based on the limited set of conditions tested, a compromise between these two qualities has to be made as they appear to be inversely related.

This research project has shown that altering color- and contrast-properties has significant effects on the performance and user experience of the NoiseTag BCI speller. To explore these effects in greater detail, future research should focus on testing more conditions, in particular in relation to BCI performance, as well as other parameters such as the color of the on-screen characters and the background. A larger subset of test subjects and/or longer test sentences could potentially increase the accuracy of the results.

7. Acknowledgements

I would like to thank Jop van Heesch, Peter Desain and Philip van den Broek from the MindAffect company for their incredible support and for supplying me with the tools I needed to do my research. Special thanks also go out to Marzieh Borhanazad, Sara Ahmadi, Karen Dijkstra and Jason Farquhar, who provided me with essential feedback and support.

8. References

- [1] J Farquhar et al. Towards a noise-tagging auditory BCI-paradigm. na, 2008.
- [2] Thielen, J., Van den Broek, P., Farquhar, J., Desain, P. (2015) Broad-Band Visually Evoked Potentials: Re(con)volution in Brain-Computer Interfacing. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0133797>
- [3] Piccione F, Giorgi F, Tonin P, et al. (March 2006). "P300-based brain computer interface: Reliability and performance in healthy and paralysed participants". Clin Neurophysiol.
- [4] Tello, R. J. M. G., Müller, S. M. T., Ferreira, A., & Bastos, T. F. (2015). Comparison of the influence of stimuli color on steady-state visual evoked potentials. Research on Biomedical Engineering, 31(3), 218-231.
- [5] 2005, Luong, et al, "Isoluminant Color Picking for Non-Photorealistic Rendering"
- [6] Rafael C. Gonzalez and Richard Eugene Woods (2008). Digital Image Processing, 3rd ed. Upper Saddle River, NJ: Prentice Hall. ISBN 0-13-168728-X. pp. 407-413
- [7] Poynton (1997). "What weighting of red, green and blue corresponds to brightness?"http://poynton.ca/notes/colour_and_gamma/ColorFAQ.html#RTFToC9
- [8] R. Gold (1967). Optimal binary sequences for spread spectrum multiplexing. IEEE Transactions on Information Theory, 13(4):619-621
- [9] Mischner, I.H.S., Schaefer, R.S., Gielen, C., Desain, P. (2007). Using multimodal frequency tagging for BCI.
- [10] Desain, P., Farquhar, J., Blankespoor, J., Gielen, S. (2014) Detecting spread spectrum random noise tags in EEG/MEG using a structure-based decomposition.
- [11] Vidal, J.J. (1973). "Toward direct brain-computer communication". Annual Review of Biophysics and Bioengineering.

9. Appendix

a. Part A: Questionnaire

Below the questionnaire originally presented to the test subjects is given. Please note that condition 7 was manually removed, and baseline conditions were added before condition 1 and 8, as described in section 4.a.

Questionnaire NoiseTag BCI

Subject No.:.....

Vragen bij het experiment

Je gaat dadelijk proberen te schrijven met de BCI-speller (het werkt niet echt, maar we doen alsof). Telkens wordt je gevraagd te beoordelen hoe comfortabel, en hoe vermoeiend (met name voor de ogen) je het vond. Eventueel kun je toelichten, of een opmerking maken over het experiment.

Er zijn 10 condities, nu begint set 1/3.

Conditie 1: Green/blue

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Conditie 2: Green/magenta

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

The effects of visual characteristics in NoiseTag BCI

Conditie 3: Blue/magenta

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Je neemt nu een korte pauze. Hierna volgt de volgende set (2/3) condities.

Conditie 4: Green/black

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Conditie 5: Blue/black

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Conditie 6: Magenta/black

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Je neemt nu een korte pauze. Hierna volgt de volgende set (3/3) condities.

Conditie 7: White 100%/white 0%

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?

(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Conditie 8: White 85%/white 15%

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?

(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Conditie 9: White 70%/white 30%

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Conditie 10: White 55%/white 45%

Op een schaal van 1 tot 10, hoe comfortabel vond je het om op het scherm te focussen?
(1 = extreem oncomfortabel, 10 = extreem comfortabel)

1 2 3 4 5 6 7 8 9 10

Op een schaal van 1 tot 10, hoe vermoeiend (met name voor je ogen) vond je het om op het scherm te focussen?

(1 = helemaal niet vermoeiend, 10 = extreem vermoeiend)

1 2 3 4 5 6 7 8 9 10

Het volgende blok kun je gebruiken om opmerking of toelichten te schrijven, indien je dat graag wil.

Dit is het einde van het experiment. Bedankt voor je medewerking aan dit onderzoek!