

Radboud University Nijmegen



ARTIFICIAL INTELLIGENCE  
BACHELOR THESIS

---

Comparison of subject-specific,  
cross-subject and adaptive classification for  
auditory oddball BCIs

---

*Author:*  
Eileen Sowa  
0815438  
e.sowa@student.ru.nl

*Supervisor:*  
Makiko Sadakata  
*Institute for Cognitive Artificial Intelligence*  
*Radboud University Nijmegen*  
Loukianos Spyrou  
*Donders Centre for Cognition*  
*Radboud University Nijmegen*

August 9, 2014

### **Abstract**

This article is about the comparison of subject-specific, cross-subject and adaptive classification. The project is based on the research of Alex Brandmeyer et al. who used subject-specific classification for event-related potential (ERPs). The comparison is done with the data from his researched. In the previous research participants had to fulfill an auditory oddball task with visual feedback. The brain responses were measured with an electroencephalography (EEG).

The expectations were that cross-subject and adaptive classification would behave worse than subject-specific. In fact the results showed that cross-subject and adaptive classification behave equal and both had a similar accuracy as subject-specific classification. Summing up, we concluded that adaptive classification has the most potential in use; first the accuracy is high and second it can cover also brain specific responses in individuals.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Classification . . . . .	3
1.2	Background . . . . .	3
1.3	Cross-subject classification . . . . .	4
1.4	Adaptive classification . . . . .	4
<b>2</b>	<b>Method</b>	<b>6</b>
2.1	Procedure . . . . .	6
2.2	The Experiment . . . . .	6
2.3	Classification Analysis . . . . .	6
2.4	Cross-subject classification . . . . .	6
2.5	Adaptive Classification . . . . .	7
<b>3</b>	<b>Results</b>	<b>8</b>
<b>4</b>	<b>Discussion</b>	<b>12</b>
<b>A</b>	<b>Appendix</b>	<b>15</b>

# 1 Introduction

## 1.1 Classification

Classification techniques are used in many areas. A customary example is the categorizing of cars which someone wants to buy. In research classification is used in areas such as classifying brain waves in artificial intelligence or psychology, or to categorize a text to a specific language in linguistic studies. Basically, in classification two or more objects are divided into groups [Tan et al., 2006]. The number of objects and groups depends on the task. Obviously many classifications techniques needed to develop, depending on the different problems. The most common classification methods for brain waves, or signal, detection are personalized; one classifier for one participant. This article will describe the electroencephalography (EEG) which is used to record brain responses, the previous research project, the limitations of the chosen classification, and two possible solutions. The results will give a first impression of possible improvements.

## 1.2 Background

The current study aims at classification to identify signals in the brain through EEG. These signals can be used to determine the humans' brain behavior during cognitive tasks. The EEG records the electrical potentials of the human brain through electrodes placed on the participants scalp. EEG biofeedback techniques are used in training participants by record the responses and give neurofeedback to improve cognitive skills [Angelakis et al., 2007]. Therefore the brain responses are not only recorded but also analyzed. It is needed to distinguish the different responses produced by the brain to get to know which are important for a specific cognitive task. The reasons therefore are: first, as mentioned above, to improve cognitive skills, and second to build up better classification method.

As mentioned above there are different application areas. Brandmeyer et al. used it for healthy population to train their brain for learning. Angelakis et al. tested neurofeedback on participants with brain injuries and healthy participants. The idea was to improve cognitive skills in people with Traumatic brain injuries, learning disabilities, depression, and schizophrenia [Angelakis et al., 2007]. Both injured and health participants, showed an increase in cognitive performance [Angelakis et al., 2007]. To train the human brain can help injured people to overcome limitations depend on their injuries. It can be used for health people to improve cognitive skills. Curran et al. described in their article the possibility of brain activity to control brain computer interfaces (BCI). They used EEG and BCI to generate and control changes in the EEG pattern of brain activity[Curran and Stokes, 2003].

In the study of Brandmeyer et al. electroencephalography (EEG) signals are used for classification. The research in the current article is based on the study of Brandmeyer et al. who tried to use neurofeedback to train participants' audible perception. The participants watched a film during a passive auditory oddball task with neurofeedback. In an auditory oddball task participants were given a sequence of tones. Within this sequence there were two tones, one used as standard and the other used as stimulus. The stimulus occurs random in the sequence so that participants cannot predict the sequence. It can also occur multiple times. The aim of an auditory oddball task is that the brain of participants should react to the presented stimulus with a peak in the signal measurement. The behavior of the brain signals are measured as event-related potentials (ERPs), in the research of Brandmeyer et al. Mismatch Negativity and P3 were measured. Both are responses in the brain which occur to an oddball stimulus.

The participants were split into two groups for the research of Brandmeyer et al., an experimental and a control group. The film participants watched was used as visual feedback. Both groups were given the feedback by differing of the sharpness of the film. The visual feedback was based on the ERPs for the experimental group. In the case that the brain react to the stimulus and the peak occur in the P3 measurement the film became sharper and less sharp if the brain showed no reaction. For the control

group the sharpness of the film differed, too. But this time it had nothing to do with the behavior of the participants' brain. His results showed differences between the experimental and the control group [Brandmeyer et al., 2014]. From the results Brandmeyer et al. concluded that neurofeedback is usable for pattern classification.

Although in his research is one limitation the time. The classifier has to be made separately for every participant and thus cannot be used for a different person. Every classifier needs to be trained during a training phase before it can be applied in the test phase. Furthermore the classifier is re-trained all four days to prevent inaccuracy. Inaccuracy can be caused by changes in the mental state of the participant e.g., tension or tiredness. An overview about the setup of the experiment is given in the Appendix Figure 8. For that reason it is preferable to have a classification method which needs less time to make and use.

But not only is the time a limitation in the approach of Brandmeyer et al. for classification. The two limitations which will be discussed in detail are personalization and time. Personalization means that every participant needs an own classifier because the brain signals vary between participant and cognitive tasks [Parasuraman and Jiang, 2012]. In addition, the wanted signals are mostly covered by noise signals, so a lot of unwanted signals are presented. To find one specific signal is a great challenge for the classifier and the reason why the most classifier are personalized until now. The time refers here to the time we need to train the classifier for one participant. The participant has to sit down for two phases, the training and the test phase. During the training phase the classifier is trained with the data of the participant. This needs of course more time than only testing. A disadvantage of more time is also the possibility of habituation. In habituation the participant adapt to the cognitive task and the brain signals are changes. To avoid this problem it is preferable to have only the test phase.

Two possible solutions will be introduced: Cross-subject and adaptive classification. Cross-subject is one classifier which can be used for every new participant without training. Adaptive classifier is a classifier based on one classifier for all, but it will adjust for every new participant.

### 1.3 Cross-subject classification

Cross-subject classification is an approach to overcome the limitations of the subject-specific classification. The cross-subject classifier will be trained with all collected data from all participants to train one classifier which can be used for every participant. However, the data to train the cross-subject classifier still has to be collected. This is done by a training and test phase like the subject-specific classifier. But compared to the subject-specific the new method is much faster in applying, because after collecting the data no further training is needed for any participant. The results for cross-subject classification could be as good as subject-specific classification depending on the chosen model itself [Wang et al., 2012]. Wang et al. tested three different models for cross-subject classification. The results showed clear that some approaches of Wang et al. could be as good as subject-specific. This gives a first suggestion for a possible solution for classification.

### 1.4 Adaptive classification

A second approach of improvement is an adaptive classification. An adaptive classification is an in-between step of subject-specific and cross-subject classification. The basic part is built up like cross-subject classification. Therefore an existing cross-subject classifier suitable for the task can be used or data have to be collected and the cross-subject classifier has to be trained. Once the basic part of the adaptive classification is made it will be updated separately for every participant. Updated means in this case that a small amount of data collection of the participant is used. With this data the adaptive classifier is trained again and will be personalized for this specific participant. Of course the total time of training

is less because the basic part of the classifier is already trained and only improved for the participant. Researches of Baldwin et al.(2012) and Yoon et al.(2009) give a first indication of the improvement using an adaptive classifier. In both researches adaptive training was suitable for the given experiments without losing accuracy or increasing time[Baldwin and Penaranda, 2012],[Yoon et al., 2009].

The approach of Brandmeyer et al. so far has a great weak point in execution; a training session has to be done for every participant. Obviously this needs a lot of time and resources to train every participant separately. The cross-subject classification seems to be a faster solution of classifying brain signals to save time and resources. The classifier has to train once and can be used for every participant and does not need a training session again. The adaptive classification seems to be faster than the subject-specific and more accurate than the cross-subject classification. The research from Wang et al. gives a first indication that it could be possible to reach as good results as for the subject-specific classification.

The basic part of the current project is based on the research of Brandmeyer et al.. The effectiveness of two new classification methods will be tested: cross-subject classification and adaptive classification. The cross-subject classification is split into two parts: combining data and combining classifier. Combining data means that the data is first combined and the classifier is then trained on this combined data. Combining classifier means that first classifiers for every data pack are trained and then taking the average of the weights of all classifier to built one classifier. The second way should be faster in time, because less training rounds for one classifier are needed. The adaptive classification is only used with classifier combination because of the less time of combining classifier in cross-subject classification.

## 2 Method

### 2.1 Procedure

We divided the experiment in two main parts. First we began with cross-subject classification. Therefore we had two possible combination methods; first data combining and second classifier combining. For cross-subject classification we took also a look at the best combination of participants. The second part was the adaptive classifier. For adaptive classification we only used the second, classifier combining, method.

### 2.2 The Experiment

The project was done with offline data from the research from Brandmeyer et al.. In his experiment 16 participants were divided into two groups (one experimental and one control group). The participants were students at the Radboud University. Both groups had to fulfill the auditory oddball task with visual feedback. 500Hz and 550Hz tones were used for the experiment. Both tones were used either as standard or stimulus, it changed within the task. The visual feedback was given by a movie film which varied in clarity. An EEG cap measured the brain activity of the participant. If the brain showed response the film became sharper; while it became less sharp when the brain showed no response. Of course for the control group the sharpness changed independently of the occurring brain signals. Both groups were tested over four days in total.

### 2.3 Classification Analysis

The classification itself was the same as described in the article of Brandmeyer et al.. They used a binary regularized linear logistic regression classifier. They trained also two classifiers; one for 500Hz and one for 550Hz. Brandmeyer et al. used a grid search to find the optimal regulation strength. They founded [5 13] as optimal regulation strength and this result was used for the current project for both, cross-subject and adaptive classification as standard strength. All programs of the current project and these which were used from the research of Brandmeyer et al. were written in Matlab.

### 2.4 Cross-subject classification

The first part of the project described in the current article is cross-subject classification. As mentioned above two methods are used for this; first data combination and second classifier combination. For this project could be only used 15 of the 16 data packs because one was damaged. At this point the minimum amount of data pack was worked out too, so how much data packs are needed for cross-subject classification. It started with the combination of three participants up to 14. In addition, also the combinations itself were worked out to see if there are participants which are classify better than other.

In data combination are two main steps; the first step is to combine all possible data and the second step is to make classifier for this new data pack. In this project it was decided to start with the data of day 1 of the experimental group. The combinations here were not randomized, the reason will be explained in the results. For the combination of three data packs the data packs of participant one, two and three are taken. For the combination of four data packs the data of participant one, two, three and four are taken. This was done up to the combination of fourteen data packs.

In classifier combination there are also two main steps; but here the classifiers were made separately for all participants' data packs first and then these classifiers will be combined. For this purpose the weights of all classifiers are taken and the final classifier is made up of the average of these weights. In this method for cross-subject classification it was possible to make random combinations; starting with

all possible combinations for combining three out of 15. This was also done for all other combining steps up to 14 out of 15 data packs.

The best and worst combinations were worked out to see which participants are classifying better than other. The best 10% and the worst 10% were taken from each combination (of the 3 up to 14). After that, the presence of each participant was counted for best and worst.

## 2.5 Adaptive Classification

The second main part of the current project was the use of an adaptive classifier. For the basic part 14 out of the 15 participants were used to build the construction of the classifier. We used half of the data of the one participant, who was left out, for training and the other half for testing. Figure 1 gives an overview which part of data of the left out participant is used for training and which part is used for testing. The first, training, phase was started with a small amount of data and it grew till the half of the data was used for training. Training means here that data was used to adjust the classifier for the particular participant. In the second, test, phase the rest of the data was used to test the performance of the classifier.



Figure 1: One data pack is split into training and testing data for adaptive classification



### 3 Results

The baseline results for the subject-specific classification came from the research from Brandmeyer et al. who discover an average classification rate of around 0.62 for the experimental group.

The two different approaches for the cross-subject classification are tested first. It was started with data combination. The calculation of the first half of all possible ordered combinations took two weeks. Due to the time it was decided not to calculate the rest of the data. Also these combinations are only tested for tones of 500Hz frequency. The second way, classifier combination, took 5 days for all of the data. To compare classifier and data combination the same combinations were taken to see if there were differences and which approach is more suitable. The average classification rate of these combinations was 0.6193 for classifier combination and 0.5822 for data combination.

It was decided to go on with classifier combination and to work out all possible combinations and their classification rates for 500Hz and 550Hz. 3 data packs combined give 455 combination possibilities. 4 data packs combined give 1365 and this goes up to 6435 for 7 and 8 data packs. After that it declined till 15 possible combinations for 14 data packs. Table 1 in the appendix gives all results for 500Hz and 550Hz and also the number of total possible combinations. In total were 32646 combination possibilities available. The total average was 0.6013 and the standard deviation was 0.0146 for 500Hz. The results for 550Hz were 0.6067 as average and the standard deviation was 0.0124. Figure 2 and 3 show the maximum (blue line), average (black line) and minimum (red line) performance of the cross-subject classification for each number of combined data packs for 500Hz (Figure 2) and 550Hz (Figure 3). The figures show that the maximum classification rate performance is around 0.75 and the average is around 0.60 which indicates that cross-subject classification can lead to high results depending on the chosen combination. The best data packs for a combination will be discussed afterwards. The graphics show that the around 7 or 8 data packs give the most suitable results. At this point the minimum went up and the maximum and average are still suitable for classification rate.

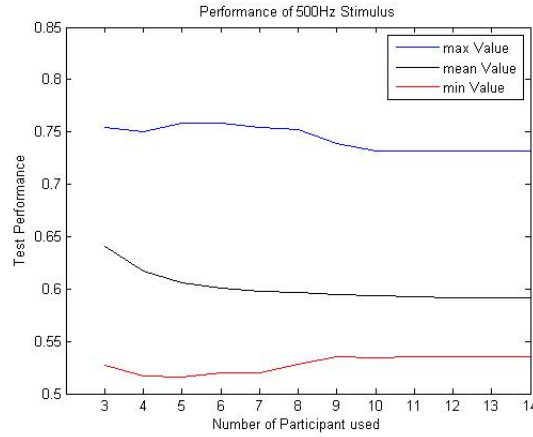


Figure 2: Performance of cross-subject classification of 500Hz

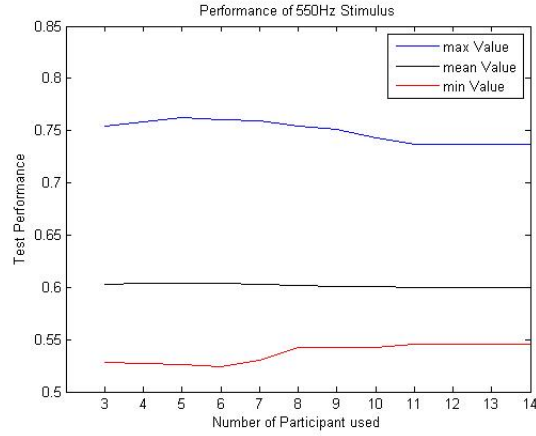


Figure 3: Performance of cross-subject classification of 550Hz

The best 10% and the worst 10% of all combination possibilities were taken to decide which participants are classifying better than other. For best 10% there were 3281 combinations and 3299 combinations in worst 10%. There is a difference between best and worst because combinations above and below the threshold with the same outcome are counted both. This is done to not lose a combination with the wanted outcome. The highest appearance in the best 10% with 2871 times had participant 10 which indicates that he is suitable for a cross-subject classifier. Also participant 14 appears quite often in the best 10% (2267 times) and participant 13 appears less in the worst 10% (191 times), so both can be suitable for cross-subject classifier. On the other hand participant 15 only appeared 5 times in the best 10% and 1900 times in the worst 10% so this one will never be used for a cross-subject classifier. Table 2 in the appendix shows all appearances of each participant in the best 10% and worst 10%. The figures 4 and 5 show the appearance graphically, the blue dots belong to the best 10% and the red dots belong to the worst 10%. Figure 4 shows the appearances by a 500Hz stimulus and figure 5 shows it by 550Hz stimulus.

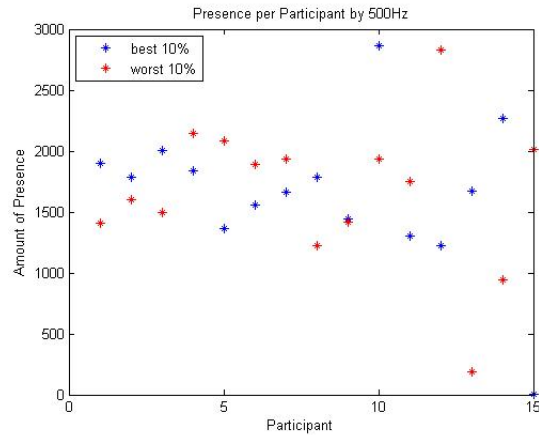


Figure 4: Presence of each participant within best and worst outcomes

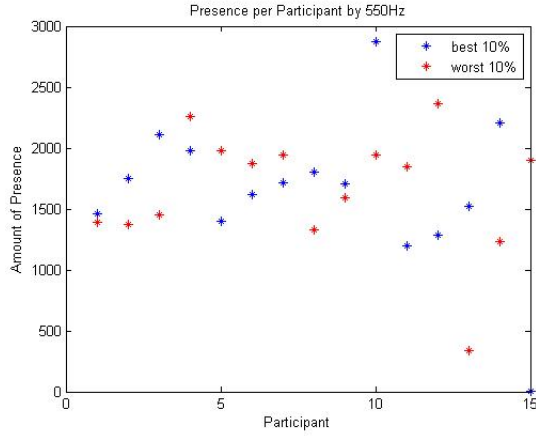


Figure 5: Presence of each participant within best and worst outcomes

The adaptive classification was measured in different steps. The training phase was split in 27 steps. It started by a few data points up to the half of the data. This was done separately for every participant and also the subject-specific and cross-subject were here calculated and compared again. For the cross-subject changed nothing during the steps, because if the classifier is built once the additional training data is not taken into account. Figure 6 shows the overall average for a 500Hz stimulus for subject-specific (green line), cross-subject (blue line) and adaptive (red line) classification for the different amount of training data used. The average classification rate performance across the participants for 500Hz was 0.604 and 0.0441 as standard deviation for the full amount of training data. The classification rate performance for 550Hz were slightly lower; 0.5733 as average and 0.0531 as standard deviation. The performance by a 550Hz stimulus for the different amount of training data is shown in figure 7.

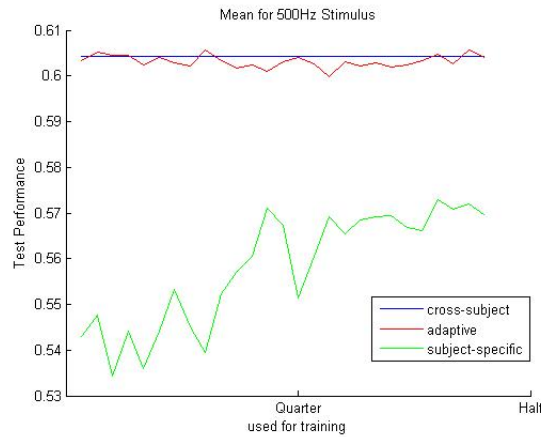


Figure 6: Average of subject-specific, cross-subject and adaptive classification

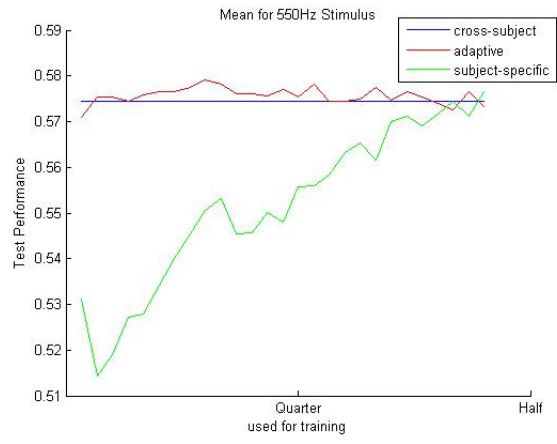


Figure 7: Average of subject-specific, cross-subject and adaptive classification

## 4 Discussion

The aim of this project was to compare three classification methods. The results of subject-specific came from the research of Brandmeyer et al.. These results were compared to cross-subject and adaptive classification. The cross-subject and adaptive classification have similar results in comparison to each other. Compared to subject-specific classification adaptive classification needs less amount of offline training data. The presence of specific individuals can improve the performance.

The first part of this project was the cross-subject classification with two different methods to train the classifier. It was started with the first method the combination of all data for cross-subject classification. This data combination gives us around 58% for half of the data. But the problem with this method was that it took two weeks and the computation had to be stopped. This first method of cross-subject classification was done with ordered data packs, so it started with participant one, two and three and so on for the next combination. This was done in the beginning and it emerged that it needed a lot of time for combining and classifying. In total there were 32646 combinations, but we already needed 2 weeks for ordered data packs, so it would take too much time to classify all random combinations. But randomizing is here an important issue, too. In the results of classifier combining were shown that the range between maximum and minimum values that was reached during the testing was around 20%. This means that there could be combinations that are bad and other that are really good in classifying.

At this point the conclusion was that the data combining method would take too much time to classify all data. The second method was the classifier combining. For all the data it only needed five days and its performance was around 61%. This led us to the conclusion that classifier combining has two advantages: first it needs less time and second it is more accurate. It was decided that classifier combining would be used during the rest of the project. At this point all possible combinations could be made and tested because the classifier combining went faster. The classifier combining needs less time because it has less data to classify than the data combining. The classifier combining only needs to classify the data of a single data pack. Afterwards it takes the average of all data, but this is still faster than classify one big data pack. To find the best combinations the best 10% and the worst 10% of all combinations were taken. It seemed that participant number 10 and 14 had high results and they quite often appeared in the best 10% of all combinations. On the other hand, participant number 12 and 15 are considerable often in the worst 10%. It is more suitable to train a classifier with participants who often appeared in the best 10%. But summing up, no one appears only in the best or worst 10%, besides participant number 15 who appears only in the worst 10%. The conclusion at this point is that we can choose the best participants, but we need also to check whether these are in combination still the best one. But on the whole we can say that some participants are have better classification rates, so that we can use these as baseline for classification of other brain responses and have still a good classification quality.

Another interesting issue within cross-subject classification is the question; is there a number of data packs to combine to reach the best result? In this project we started with combining three data packs up to 14. But there were only 15 data packs available as a result that 7 and 8 combined data packs have the highest number of combination possibilities. This means that the number of total combinations increased till 7 and decreased after 8. This showed that combining three data packs has the highest results in maximum value reached and average value. But after combined four data packs the average is still around 62% and it falls slightly after combining five data packs. So that also four and five gives high results in maximum and average. And another issue is that the minimum went up after combined 8, this has a result that the overall accuracy improved.

In brief, this project couldn't give an advice how much data packs will needed for the best cross-subject classification. It was seen in the results that the average notable fall after three data packs, but really

slightly after that. Of course was the conclusion here that the best number should be around 7 or 8 data packs. On the other hand the number of combinations also decreased after 8. Simultaneously increased the performance of the minimum slightly and the maximum was still high. The increasing minimum can be an indication for the stable results. But these results will depend on the chosen data packs itself. Further research can help to see if more data packs improve the performance of cross-subject classification.

The second and last part of this project was the adaptive classification. In this last step all three classification methods were compared. Due to the results founded in the cross-subject part of the project, classifier combining was chosen as method for the cross-subject classification. In this second part the amount of training data of the new participant is not important for the cross-subject classification. The cross-subject classifier is trained before with 14 participants and used for the 15th, afterwards the new data from the participant is not used for improving. But the cross-subject classification was used as a baseline to compare the classification rate performance to adaptive and subject-specific classification. Still, for subject-specific classification the performance improved with more training data. The adaptive classification is the most interesting part. It had an equal result as the cross-subject classification and it doesn't need much of training data. It had a high result from the beginning of the training phase. This means that adaptive classification is usable with less amount of training data. But an advantage against cross-subject classification is that an adaptive classifier also covers subject-specific brain responses that could be lost in a cross-subject classification.

This project showed that the results of cross-subject and adaptive classification are not significantly different. But the advantage of covering subject-specific brain responses are more satisfied in the adaptive classification. It was also shown that it cannot give an advice for the amount of data packs which are needed for the best cross-subject or as basic for adaptive classification. But in any case it depends on the chosen data packs itself. The project showed that some participant responses are better than other for the use in cross-subject classification. On the whole the project showed clearly that there are advantages in the use of cross-subject and adaptive classification. This has the consequence that both methods can be used in classification without large losses of performance. Summing up, all three classification methods have different advantages and disadvantages and the best one is be determined by the problem itself. There is no guideline for the best classification method and it has to be chosen separately for every project.

## References

- [Angelakis et al., 2007] Angelakis, E., Stathopoulou, S., Frymiare, J., Green, D., Lubar, J., and Kounios, J. (2007). EEG Neurofeedback: A Brief Overview and an Example of Peak Alpha Frequency Training for Cognitive Enhancement in the Elderly. *The Clinical Neuropsychologist*, 21(1):110–129.
- [Baldwin and Penaranda, 2012] Baldwin, C. and Penaranda, B. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *NeuroImage*, 59:48–56.
- [Brandmeyer et al., 2014] Brandmeyer, A., Sadakata, M., Spyrou, L., McQueen, J., and Desain, P. (2014). Modulation of auditory evoked responses using online decoded-EEG neurofeedback: Towards enhanced perceptual learning. In press.
- [Curran and Stokes, 2003] Curran, E. A. and Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and Cognition*, 51:326–336.
- [Parasuraman and Jiang, 2012] Parasuraman, R. and Jiang, Y. (2012). Individual differences in cognition, affect, and performance: Behavioral, neuroimaging, and molecular genetic approaches. *NeuroImage*.
- [Tan et al., 2006] Tan, P.-N., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining*. Addison-Wesley: Pearson International Edition.
- [Wang et al., 2012] Wang, Z., Hope, R., Wang, Z., Qiang, J., and Wayne, D. (2012). Cross-subject workload classification with a hierarchical Bayes model. *NeuroImage*, 59.
- [Yoon et al., 2009] Yoon, J., Roberts, S., Dyson, M., and Gan, J. (2009). Adaptive classification for Brain Computer Interface system using Sequential Monte Carlo sampling. *Neural Networks*, 22:1286–1294.

## A Appendix

Number of data packs combined	Average Accuracy 500Hz	Average Accuracy 550Hz	Total possible combinations
3	0,6409	0,6409	455
4	0,6175	0,6202	1365
5	0,6061	0,61	3003
6	0,6007	0,6052	5005
7	0,5977	0,6026	6435
8	0,5962	0,6018	6435
9	0,5954	0,6018	5005
10	0,5945	0,6019	3003
11	0,5931	0,6016	1365
12	0,5908	0,6008	455
13	0,5878	0,599	105
14	0,5843	0,596	15

Table 1: Average Accuracy for cross-subject classification

Number of participant	Appearance in Best 10%	Appearance in Worst 10%
1	1900	1407
2	1787	1601
3	2005	1497
4	1841	2142
5	1361	2082
6	1557	1886
7	1660	1930
8	1783	1222
9	1447	1415
10	2867	1936
11	1300	1752
12	1226	2826
13	1668	191
14	2267	941
15	1	2011

Table 2: Appearance in best 10% (3281 per participant possible) and in worst 10% (3299 per participant possible)





Figure 8: Procedure of the experiment of Brandmeyer et al. (Source:[Brandmeyer et al., 2014])