Comparing Classification Methods for Asynchronous Brain-Computer Interfaces

Tijl Grootswagers

August 24, 2011

BSc Artificial Intelligence Thesis
Radboud University Nijmegen
Supervisors: J. Farquhar & L. Roijendijk

Abstract

In the field of Brain-Computer Interfacing, the asynchronous approach adds a new ”no movement” class to the classic trial-based synchronous approach. This ”no movement” class is very unbalanced (most of the data will belong to this class) and therefore causes a major classification obstacle. This study investigates the suitability of different classification methods for use in an asynchronous Brain-Computer Interface. We used a Guitar Hero like game to gather EEG data to compare the classification methods using the Readiness Potential in different movement conditions. Based on the results, we recommend to use a hierarchical method that first classifies between movement and no movement, and only if movement is detected, classifies between left and right movement.

1 Introduction

Brain-Computer Interfaces (BCI's) cover the field of systems that let people communicate with a computer by directly measuring brain activity. There are many ways to achieve such communication and a general distinction is made between invasive and non-invasive BCI's. Invasive BCI's measure brain activity directly from inside the brain by implanting electrodes in the brain tissue whereas most non-invasive BCI's use electroencephalographic (EEG) activity measured on the scalp, although magnetoencephalography or functional magnetic resonance imaging have also been used in BCI experiments [14, 26]. The advantage of non-invasive BCI is that the preparation is an easy and relatively fast procedure and causes fewer ethical dilemmas [8]. Non-invasive BCI's using EEG have a fast preparation procedure and are the most used method for BCI research.

Most work in the field of BCI's is targeted at helping paralysed patients communicate or move. Examples are writing with the p300 visual speller [7] or operating a prosthetic device by imagining arm or foot movement [17]. There is also a large area of possible BCI applications aimed at healthy users. Some commercial BCI's already exist, for example to operate games such as the "emotive" [1]. However, these commercial applications tend to focus at global brain states such as different levels of concentration or possibly use artefacts rather than brain signals [18].

1.1 Brains on Fire

Figure 1: A screenshot of the Brains on Fire game used in this study. The user has to hit the notes by moving his left or right hand at the moment the note pass the horizontal line to score points.

An example of a healthy user application is the Brains on Fire game developed to determine if the Lateralised Readiness Potential (see Section 3.1) occurs in imagined movement [25]. Figure 1 shows an in-game screenshot of Brains on Fire. The game is based on the open source Frets on Fire (FoF) game, which in its turn is based on the popular Guitar Hero game [11]. Brains on Fire uses FoF's source code and includes several modifications to make the game suited for BCI. The game is now focused on drumming with both hands instead of playing the guitar and distinguishes between left and right notes only. To provide a usable signal for research, easy to learn songs were created and most visual effects were removed.

1.2 Asynchronous BCI

Most BCI systems are designed using a synchronous approach, recording the users decision in fixed time frames [7, 21, 2, 4]. The synchronous approach facilitates analysis of the EEG, as there exists a fixed time window where the signal should appear. In an asynchronous BCI, no cue stimulus is used and the user is free to decide when to engage in the specific mental activity to operate the BCI. This method requires a continuous processing and classification of the EEG signal. The system has to remove noise and classify the signals as fast as possible (real-time) to provide a usable computer interface [24].

Where a binary synchronous BCI only has to classify two classes every fixed time step (trial), for example left hand movement versus right foot movement, a 'binary' asynchronous BCI automatically has to deal with a third class, for when there is no movement at all. The third or "no signal" class will also have a great impact on the balancedness of the data (most of the data will belong to the no signal class) because for the majority of the time, the user does not intend to give an input.[1] Table 1 shows the difference for two trials.

Input for 2 trials:																			
				L								R							
Synchronous output:																			
Left										Right									
Asynchronous output:																			
0	0	0	0	L	0	0	0	0	0	0	0	R	0	0	0	0	0	0	0

Table 1: Synchronous and asynchronous target output over two trials (trials) showing how the asynchronous method differs in number of classes and number of decisions per trial.

Examples of areas where asynchronous BCI can be used include:

- Driving a car [28]
- Computer gaming as with Brains on Fire
- Operating a wheelchair [15]

1.3 Classification

Many methods, mostly linear [16], have been used for classification in BCI, all of them having their own strengths and all being suited for specific tasks. Examples include:

- Linear logistic regression [23]
- Neural Networks [20]
- Support-Vector Machines [12, 4]

The research to date has tended to focus on synchronous rather than asynchronous BCI. Therefore, much less is known about the suitability of classification methods for asynchronous BCI systems [13, 16].

As mentioned in the previous section, the difficulties with an asynchronous BCI lie in the unbalanced multi-class classification problem. It is clear that when dealing with such unbalance, it is important that the amount of false alarms is minimised.

1.4 Lateralised Readiness Potential

In EEG data of movements in humans, various signals can be identified which are caused by real or imagined movement. One of these signals is the Readiness Potential (RP) which is a decrease in amplitude starting before a movement in the central motor area (around the Cz electrode, see Figure 11 for the location of the electrodes) [9, 22, 19]. In unilateral movement, the RP becomes lateralised on the C3 and C4 electrodes (See Figure 11 for the locations). This is known as the Lateralized Readiness Potential (LRP) and is used in this study. Figure 2 shows a typical LRP in the C3 electrode for finger movement.

[1] When classifying every 100ms and the user is trying to produce a signal once every second, on average 9 out of 10 classifications should yield no signal. Also note that once every second is not realistic for a BCI, the real input rate will be much lower and approximately 99% of the data will end up containing no intentional signal.

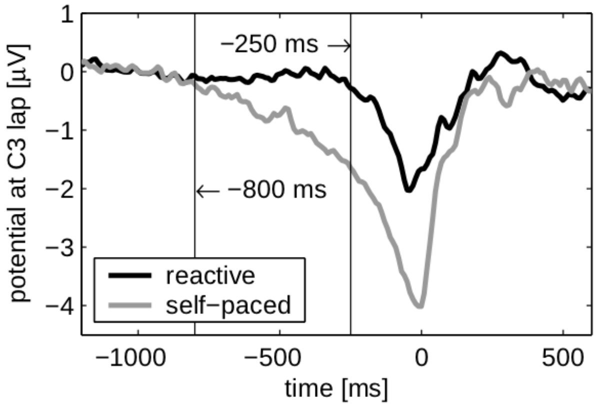


Figure 2: A typical LRP in the C3 electrode for right finger movement showing the increase in negativity before movement onset [10].

1.5 Our study

The aim of this study was to determine which classification method is most suited for running an asynchronous BCI.

The LRP signature from Section 1.4 was used for classification. A special interest has been taken in the universum method (described in Section 2.3.3), as it has not been used in BCI classification before and we expected this method to fit the asynchronous classification problems introduced in Section 1.2.

The rest of the paper is organised in the following way:
First, we will describe how we acquire the dataset using the BoF platform. Next, we give an overview of the preprocessing pipeline and the classification methods we want to compare. Finally, we list the results and report our conclusions.

2 Methods

The objectives of this research were to evaluate different classification methods for asynchronous BCI. As the Brains on Fire game (introduced in Section 1.1) is well suited for asynchronous BCI, we used this platform to compare the different classification methods. We had to create a dataset for off-line processing and classification. This process is described in the next sections.

2.1 Data acquisition

2.1.1 Equipment

The experiment set up was the same as used by Versteeg [25]. This set up measured the following:

- 64 Ag/AgCl active electrodes (placed in the international 10-20 system) measuring the EEG signal in 2048Hz.
- 4 electrodes, two on the outer side of both eyes and one above and one below the right eye measured vertical Electro-oculographic (EOG) activity (eye movements, eye blinks).
- 4 electrodes on the wrist and below the elbow on both arms measured Electromyographic (EMG) activity (arm movement).
- A midi drum pad recording the physical impact of the drum sticks.

2.1.2 Participants

Three students participated in the experiment, age 21,23 and 26, two male, all right handed, two with EEG experience, none with experience in BCI or imagined movement.

2.1.3 Preparation and instruction

The participants were first informed about BCI in general and asked for permission to use their EEG data in this study. During and after attaching the cap and electrodes, the participants were given instructions about the rest of the experiment. We provided them with a few example trials and let them practice with the drum sticks on the midi pads.

2.1.4 Stimuli

During the experiment, only one song was used. This song consisted of very simple patterns which were played two times with the subject only having to play along in the second pass. The song had 70 notes (trials) evenly distributed over left and right targets. Every block consisted out of three songs, which makes up for 210 trials per block.

2.1.5 Conditions

We measured four different conditions in seperate blocks. Table 2 shows the time-line we used in the experiment.

1. **Real drumming** The first phase where EEG data was recorded consisted of the participant physically hitting the midi pads using the drum sticks. This phase resembles a real drumming situation and was therefore used first to make the subject comfortable with the songs. For this phase the subject was instructed to make short movements with their hand only, and not use the whole arm.

2. **Air drumming** In this condition the subjects were not holding the drum sticks and were instructed to make the same hand movement as they did in the real drumming blocks.

3. **Imagined drumming** We included imagined movement in this study to validate previous findings. We instructed the subjects to keep their hands on their lap and to very lively imagine the hand movements they had made before.

4. **Observing** In this condition, we had the subjects watch one song without doing anything as a control condition. The observing condition could later be used to identify (and isolate) the effects of seeing and hearing the stimuli.

Time (minutes)	Activity
30m	Preparation and instruction
30m	Real drumming (6 songs)
15m	Air drumming (3 songs)
15m	Imagined drumming (3 songs)
5m	Observing (1 song)

Table 2: Experiment time-line. One block consists of three songs, with after every block a short break where instructions were given.

2.2 Preprocessing

To get from raw EEG data to classification trials, the data had to be visually inspected and preprocessed to extract features to classify on. We used a standard ERP preprocessing pipeline [4, 25] consisting of the following steps:

1. Down-sampling the data from 2048Hz to 32Hz.
2. Slicing the data to create windows from -750ms to 0ms (movement onset), and adding no signal trials.
3. Detecting and removing noise/artefacts (using EOG for eye artefacts).
4. Detecting and removing bad channels and bad trials. When a channel or trial differs more than three standard errors, it was rejected.
5. Re-referencing. Using the common average over all channels, we re-referenced with this average to improve the signal to noise ratio.
6. Extracting signal features with a frequency filter in the in the lower frequency range (0 to 12Hz).

When we visually inspected the data, we saw that there was a lot of activity in the frontal and occipital areas after movement onset (see figure 11). This could have been the result of movement artefacts, tension or reactions to the visual stimuli. As the classifiers could benefit from this information, which in some channels was class-dependant as well, we decided to take the time window from 750ms before movement up to the movement onset (0ms) and not provide the classifier with the post-movement information. Although the LRP is located in the motor cortex, for classification we included all channels to provide more information about the noise.

We took the no signal trials from moments between two movements. The time between two movements could be very short and therefore a no signal trial could still contain some parts of the signal from the previous movement. To make sure this signal was not time-locked and could not be used by the classifier, we randomly shifted the no signal windows in time by +/- 100ms.

This procedure gave us 50% no signal trials in the data, resembling the unbalance of the asynchronous BCI. When interpreting the results it should be noted that the real class distribution would consist of over 90% no signal trials (as explained in Section 1.2).

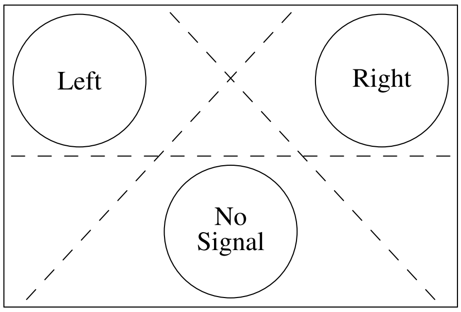
2.3 Classification

As described in Section 1.3, many classifiers can be used for BCI. For this study we used linear logistic regression classifiers as they have been proven to work well in BCI research [6]. A linear logistic regression classifier tries to fit the feature information to a logistic curve to predict the probability of a class [3]. After the preprocessing described in the previous section, we used a kernel function to map the high-dimensional data to a linear space [3]. We trained the classifiers using regularisation which prevent over-fitting on the training data by applying a penalty to large weights [6]. We used ten-fold cross-validation to pick the optimal regularisation parameter and evaluate the generalisation performance of the classification method on the dataset.

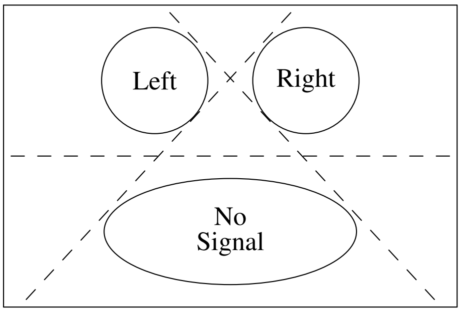
Cross validation works by partitioning the data into ten folds, where each fold uses a different subset of the data as a test set. The remaining data form the training set, used to train the classifier. The classifier's performance is then calculated for the test set and the process repeats for every fold. This way all trials in the data are used for validation exactly once and we can measure a test set performance using all the trials. We applied the same cross-validation folding for every classification method to ensure that differences between methods can not be caused by random differences in the folding.

To use the binary classifiers in the multiclass environment, different methods can be used to divide the multi-class problem into subproblems [5]. The methods we used in this study are described in detail in the next sections.

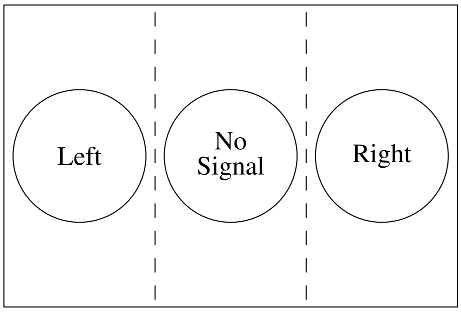
The results of the different methods will also provide information about the distribution of the classes in the high dimensional feature space, as different methods make different assumptions about the class distribution. Figure 3 shows examples of different class distributions in two dimensional space.



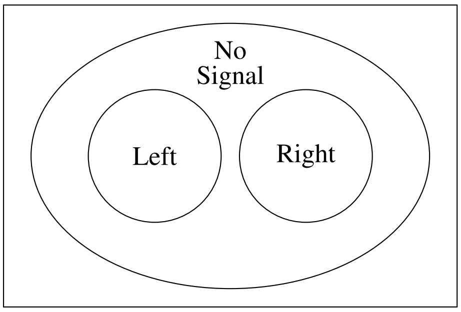
(a) All classes are linearly separable from the others



(b) All classes are linearly separable but the no signal class overlaps the other classes in one dimension



(c) Requires two linear boundaries for the no signal class



(d) Requires non-linear boundaries

Figure 3: Examples of class distribution in two-dimensional space. Dashed lines show the linear decision boundaries. Some methods we used make assumptions about the distribution. If these methods succeed, we get an idea of the class distribution of the high-dimensional EEG data.

2.3.1 Hierarchical

The most straightforward approach of dividing the multiclass problem into subproblems, is to use two sequential classifiers. The first classifier is used for signal detection. Only if the first classifier predicts a signal, the second classifier is used to determine whether the signal was a left or a right hand movement. This hierarchical structure is visualised in Figure 4. The hierarchical method only works under the assumption that the no signal class is linearly separable from the other two classes and the other two classes are linearly separable from each other. In the examples from Figure 3, this assumption is only true for 3a and 3b. Note that these figures show more or less what one would expect the class distribution to be like when imagining that signal strength in the Cz electrode (whether a RP occurs at all) is projected on the y-axis and the x-axis showing the lateralisation of the signal in the C3 and C4 electrodes.

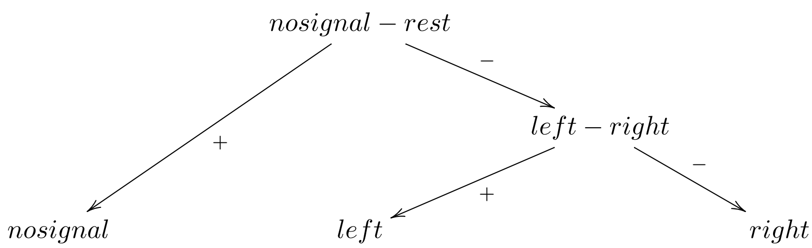


Figure 4: The structure of the hierarchical method. First, a specialised no signal versus rest classifier is used for signal detection and if there is a signal, the left versus right classifier is used to classify between left and right.

2.3.2 1 versus X

Being among the most general classifiers, the "1 versus Rest" (1vR) and "1 versus 1" (1v1) methods make very few assumptions about the class distribution and can deal with most class distributions.

The 1vR classifier tries to find three decision boundaries by using a subproblem of the form 'x versus rest' for every class, e.g. left versus (right or no signal). To implement the 1vR method we trained three classifiers using the subproblems and target outputs from Table 3a.

Subproblem	left	right	no signal
left versus rest	+	-	-
right versus rest	-	+	-
no signal versus rest	-	-	+

(a) 1 versus Rest

Subproblem	left	right	no signal
left versus right	+	-	
left versus no signal	+		-
right versus no signal		+	-

(b) 1 versus 1

Table 3: Subproblem target outputs. One classifier is trained for each subproblem using these target outputs. No target output exclude a class from a subproblem.

After training the classifiers on these subproblems, for every trial, the output of the classifier was multiplied with its target value for every class to make the class with the highest sum over the subproblems the predicted class for that trial.

The 1v1 approach is slightly different from the 1vR as it uses different classifiers for all combinations of classes, as shown by its subproblem decomposition matrix in Table 3b.

The advantage of the 1v1 method is that not all classes have to be linearly separable from all other classes, but must only be pairwise separable. In the 1vR method all classes have to be linearly separable from all other classes. The disadvantage is that the relation between the number of classes (n) and the number of classifiers needed is $\frac{1}{2}(n^2 - n)$ compared to only n in the 1vR approach.

It is clear that the hierarchical approach fails when the data would be distributed as in Figure 3c whereas the 1v1 method would have no problem dealing with such a situation. Although the 1vR method would suffer from the same problem; the no signal versus rest classifier cat not create a useful decision boundary, a solution could still be found by combining the information from the other two classifiers to single out the no signal class.

2.3.3 Universum

The universum method is a semi-supervised classification method used to increase binary classification performance. When adding universum data that lies around the decision boundary between other classes, as in Figures 3a and 3c, the classifier can use this to improve the recognition rate of true classes. This method has higher classification performance compared to general supervised and semi-supervised learning methods [27].

The universum method has not been used in BCI classification problems and we expected it to fit the problem and class distribution really well. Besides, an advantage of the universum method is that it only needs one classifier, instead of two for the hierarchical methods (Section 2.3.1) or three for the 1vX methods (Section 2.3.2). Hence, it should be less prone to over-fitting.

In our case, we used the no signal class as universum data and trained a binary classifier on the two remaining classes. This approach differs from the general approach in the fact that the universum data is used as a target class as well (the no signal class) instead of being added to aid the binary classification.

Implementing the universum classifier for our data was achieved by assigning all no signal trials to both left and right class to get a class distribution as shown in Figure 5.

With the no signal trials in both the left and right class, the only way for a binary classifier to minimise it's error on a no signal trial is to put it around zero. We trained a classifier on the new set retaining the same folding information for the doubled trials. This yielded a decision value distribution where the decision values for the no signal trials were distributed around zero and the left and right classes were distributed around respectively 1 and -1. See Figure 6 for a plot showing this distribution for one subject in the air drumming condition.

We determined two thresholds by maximizing the correctly classified trials between left and no signal classes and between right and no signal classes. Every trial located between the two thresholds was considered no signal.[2]

[2] This method of determining thresholds is chosen as an example and a proof of concept but will not yield the most optimal thresholds for an asynchronous BCI as the class importance should be taken into account to deal with the unbalance of the no signal class.

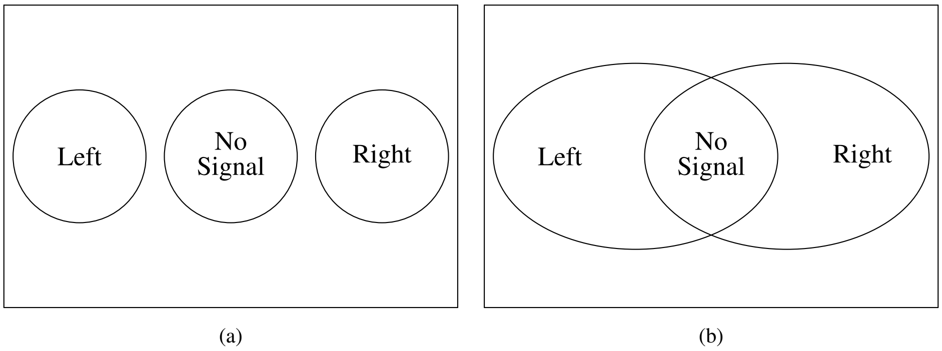


Figure 5: The effect of adding the no signal trials to both left and right classes.

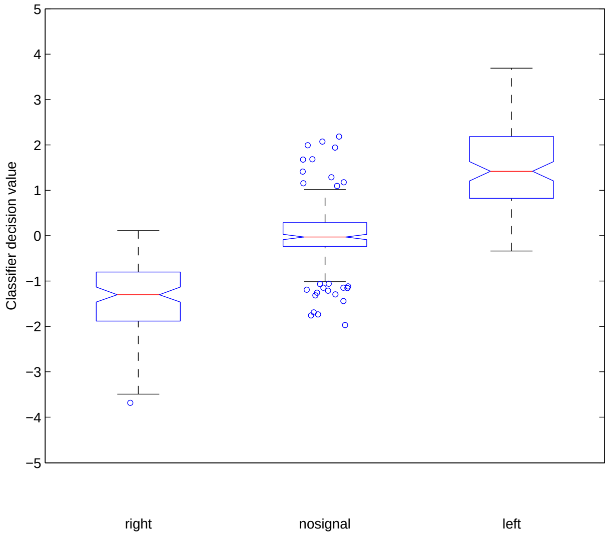


Figure 6: Boxplot of the distribution of the decision values for all test set trials after training the universum classifier. One subject in the air drumming condition is shown (400 trials). The decision values for the no signal class are distributed around zero while the left and right classes can be found around respectively 1 and -1.

2.3.4 Combined Approach

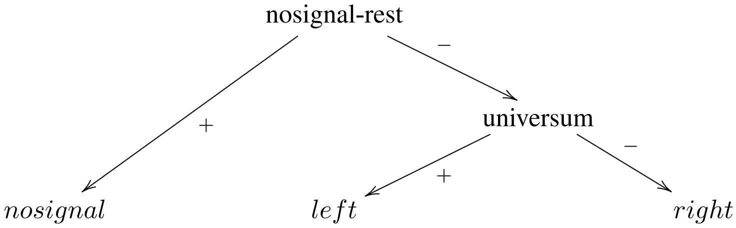


Figure 7: The structure of the combined approach. First, a specialised no signal versus rest classifier is used for signal detection and if there is a signal, the universum classifier is used to classify between left and right.

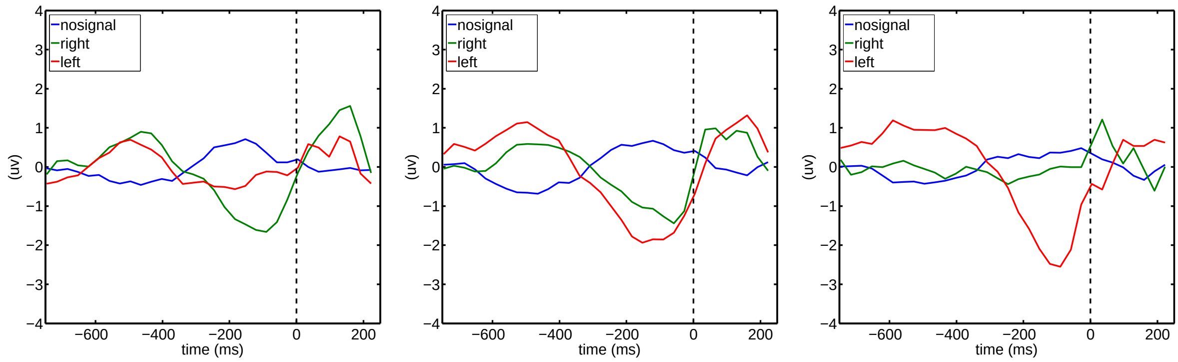
We expected the universum classifier, being a semi-supervised method, to have a higher accuracy on the binary left versus right subproblem as it had more information about the distribution of the noise. However, we also expect a specialised classifier for signal versus no signal detection as used in the hierarchical approach to have a higher signal detection performance if the no signal class is linearly separable. The logical way to get the best of both sides is to use a combined approach.

We implemented this approach by swapping the left versus right classifier in the hierarchical method for the universum classifier. Figure 7 shows this updated structure.

3 Results

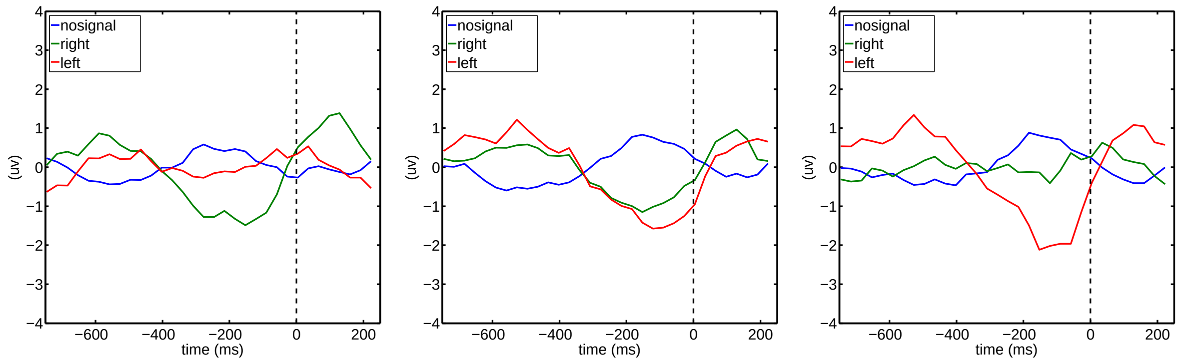
3.1 LRP Signature

We visually inspected the data to see if there was a detectable LRP (See Section 1.4) and to determine the frequency bands we had to filter. Figures 8 and 9 show the LRP signature in the C3, C4 and Cz electrodes for the real drumming and air drumming conditions after applying the preprocessing. Figure 10 shows the same channels for the imagined drumming condition.



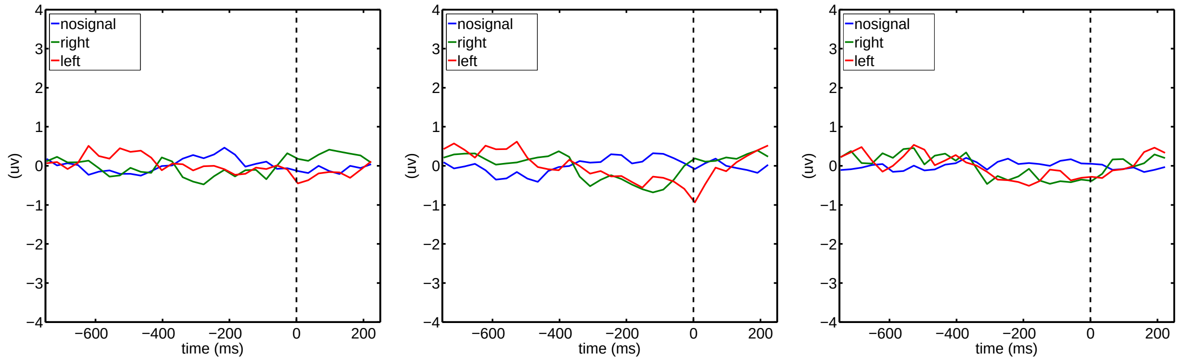
(a) C3 (b) Cz (c) C4

Figure 8: ERP plot averaged over all subjects (800 trials/subject) for the real drumming condition after applying the preprocessing zoomed in on the C3, C4 and Cz electrodes. The LRP is very clear in these electrodes.



(a) C3 (b) Cz (c) C4

Figure 9: ERP plot averaged over all subjects (400 trials/subject) for the air drumming condition after applying the preprocessing zoomed in on the C3, C4 and Cz electrodes showing that the LRP signal is also very clear for the air drumming condition.



(a) C3 (b) Cz (c) C4

Figure 10: ERP plot averaged over all subjects (400 trials/subject) for the imagined drumming condition after applying the preprocessing zoomed in on the C3, C4 and Cz electrodes. These plots show that the LRP signal is not present for imagined movement.

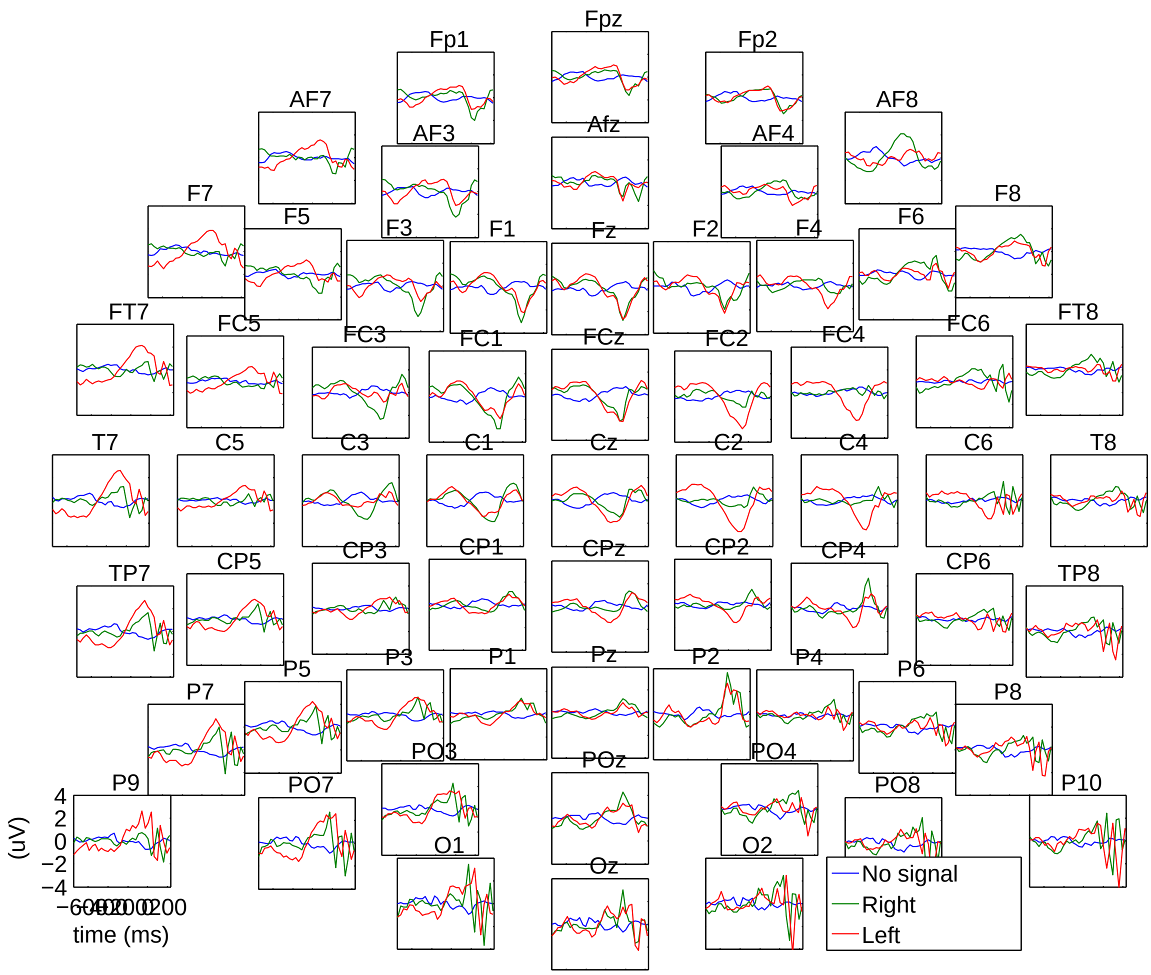


Figure 11: ERP plot of the real drumming condition averaged over all subjects (800 trials/subject). A time window of -750ms to 250ms after movement onset is used. This overview indicates that there is a lot of activity in the frontal (for example the Fpz electrode) and occipital (for example P2 and PO7) areas after movement onset. This could have been the result of movement artefacts, tension or reaction to the visual stimuli.

3.2 Classification

In this section, we show the classification performance of all methods for all conditions per subject, including an average over all subjects. For every condition we show the multiclass accuracy which is the cross-validated percentage of correctly classified trials. The error bars represent the standard error over folds. The results from Figure 12 and 13 show that the overall performance of all methods on the real drumming and air drumming is about 80%. Figure 14 shows that performance on imagined movement is lower than 50%.

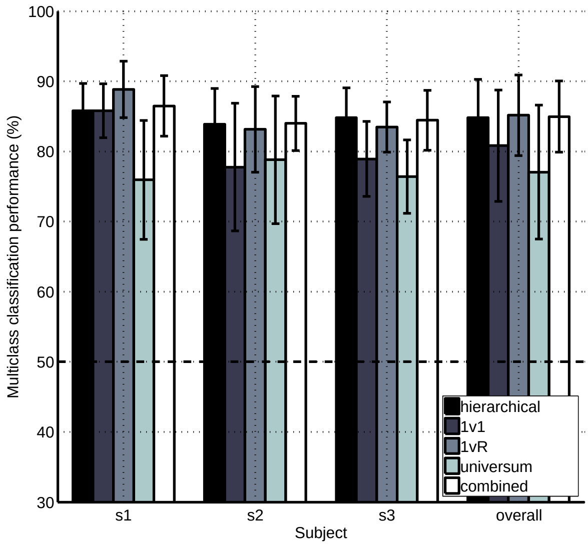


Figure 12: Multiclass classification performance (cross-validated percentage of correctly classified trials) of the different methods in the real drumming condition (800 trials/subject). The error bars represent the standard error over folds and the dashed line shows chance level. All methods perform around 80% and not much difference exist between subjects.

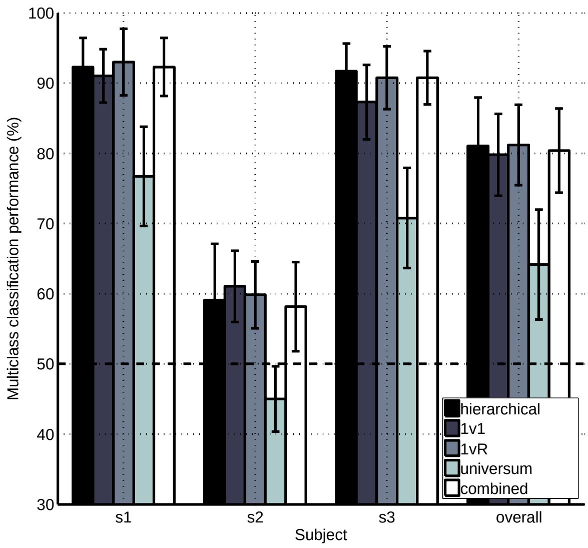


Figure 13: Multiclass classification performance (cross-validated percentage of correctly classified trials) of the different methods in the air drumming condition (400 trials/subject). The error bars represent the standard error over folds and the dashed line shows chance level. The universum method seems to perform less compared to the other methods. The performance for subject 2 is much lower compared to the other subjects.

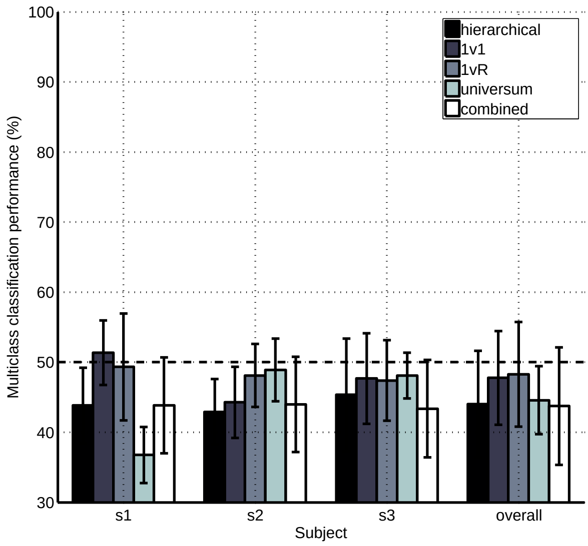
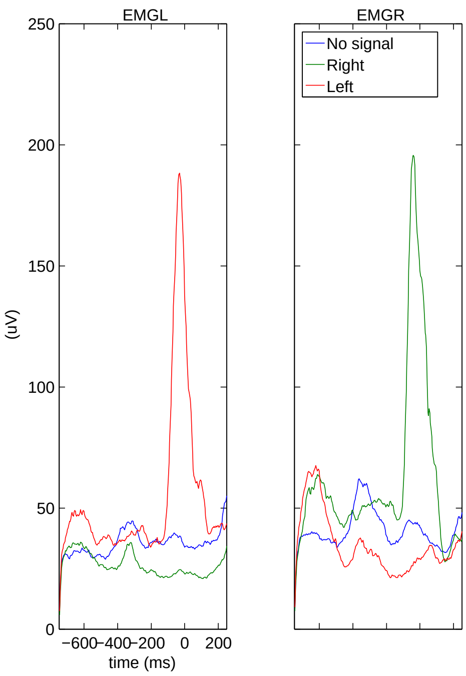
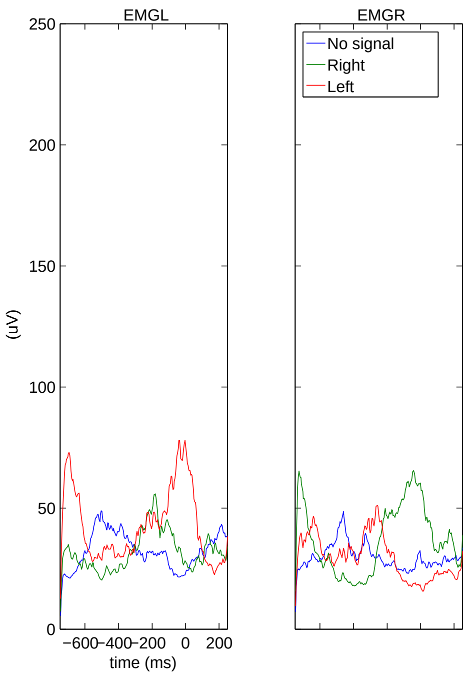


Figure 14: Multiclass classification performance (cross-validated percentage of correctly classified trials) of the different methods in the imagined drumming condition (400 trials/subject). The error bars represent the standard error over folds and the dashed line shows chance level. The performance of all methods is below 50% for this condition.



(a) Real Drumming (800 trials)



(b) Air Drumming (400 trials)

Figure 15: Averaged EMG plots over all trials (real drumming: 800 trials, air drumming: 400 trials) measuring muscle activity on both arms for subject 2. The peaks before the movement onset (0ms) are the result of previous trials as the time between trials was short.

4 Discussion

4.1 Asynchronous BCI

The classification performance for the real movement conditions (real drumming and air drumming) is above 80%. The minimum accuracy to communicate using a BCI is generally considered 70% [21]. However, for asynchronous BCI this number should be higher to compensate for the problems described in Section 1.2.

Considering this, the performance of the classifiers is possibly too low for proper usability. However, classification performance in asynchronous BCI can still be improved by using various post-processing techniques, for example giving the classifier information about the class distribution and penalising two consecutive signal detections.

4.2 Universum method

In the air drumming condition, the classification performance of the universum method was not as good as the other methods. Although less clear, this seemed also true for the real drumming condition. We expected the universum classifier to do better as it seems to fit this task really well, although the results seem to indicate otherwise. Further, a clear benefit of the universum method for the left versus right subproblem could not be identified in this analysis, which implies that adding unlabeled data does not improve the classifiers in this problem.

4.3 Class distribution

Referring to Figure 3, which showed different examples of class distribution, we saw that all methods classify all three classes correctly. Therefore, it seems that all classes are linearly separable as in Figures 3a and 3b. The most likely model, the one that meets the requirements of all methods, would be Figure 3a. In this model, all classes are linearly separable, which is an assumption for the 1vR. Besides, the no signal class lies on the decision boundary for the left versus right subproblem, which is a requirement for the universum classifier to work. Figures 3c and 3d are not linearly separable using one decision boundary for each class and would not have worked for the hierarchical method.

4.4 Air drumming

Particularly interesting is the performance of the classifiers for subject 2 in the air drumming condition (Figure 13), which is much lower compared to the other subjects. This result can be explained by the EMG plots shown in Figure 15. Figure 15a shows that for the real drumming condition, subject 2 produces a clear peak at 0ms (movement onset), indicating a strong signal that has the same timing in every trial. In air drumming (Figure 15b), the peak is much weaker, which means that the EMG signal is not consistently appearing at 0ms, and gets averaged out in the plot. In the real drumming condition the 0ms is located exactly at the time where the marker from the drum pad is inserted, whereas in air drumming, the marker is inserted afterwards by using the information from the song. When the movements are not timed consistently with the notes, the signal in the air drumming condition will not always appear at 0ms, which causes a loss in performance. In other words, subject 2 is a bad drummer.

4.5 Imagined drumming

The results for the imagined movement condition are consistent with those of other studies and suggest that the LRP signal does not appear in imagined movement [25]. Note that while the performance of the classifiers seems to be below chance level, the actual chance level differs per classifier and is at most 37,5%.[3]

[3] Given the distribution of 50% no signal, 25% left and 25% right trials, classifying everything as no signal will yield a 50% performance. However, assigning class labels randomly will correctly label only 33% of the trials. Assigning class labels randomly while taking the class distribution into account will get 37,5% right.

5 Conclusion

The purpose of the current study was to determine which classification method is most suited for an asynchronous BCI. Taken together, these results suggest that no differences seem to exist between different methods, as all methods are within one standard error from each other. The universum method may be an exception, as it seems to do worse than the other methods, especially in air drumming.

Based on these findings, we would recommend the hierarchical method as it is the fastest and most straightforward method to implement in the asynchronous BCI. However, considering the small number of subjects, caution must be applied when interpreting these findings.

5.1 Future Work

As the results of this study are based on three subjects, in order to get statistically significant results more subjects can be tested in a follow-up study.

To be able to train classifiers on air drumming, the marker timing issue (Section 4.4) has to be fixed.[4] This can be achieved by using the EMG data to determine the movement onset.

A very obvious next step is to implement the asynchronous BCI using the hierarchical classification method and post-processing (For which an example was given in Section 4.1) to play a game of Brains on Fire!

References

[1] Emotive Systems: www.emotiv.com, 2010.

[2] F. Babiloni, F. Cincotti, L. Lazzarini, J. Millan, J. Mourino, M. Varsta, J. Heikkonen, L. Bianchi, and M. Marciani. Linear classification of low-resolution EEG patterns produced by imagined hand movements. *IEEE Transactions on Rehabilitation Engineering*, 8(2):186–188, 2000.

[3] C. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer, 2006.

[4] B. Blankertz, G. Curio, and K. Müller. Classifying single trial EEG: Towards brain computer interfacing. In *Advances in neural information processing systems 14: Proceedings of the 2002 conference*, page 157. MIT Press, 2002.

[5] T. Dietterich and G. Bakiri. Solving multiclass learning problems via error-correcting output codes. *Arxiv preprint cs/9501101*, 1995.

[6] J. Farquhar. A linear feature space for simultaneous learning of spatio-spectral filters in BCI. *Neural Networks*, 22(9):1278–1285, 2009.

[7] L. Farwell and E. Donchin. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.

[8] P. Haselager, R. Vlek, J. Hill, and F. Nijboer. A note on ethical aspects of BCI. *Neural Networks*, 22(9):1352–1357, 2009.

[9] H. Kornhuber and L. Deecke. Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv European Journal of Physiology*, 284(1):1–17, 1965.

[10] M. Krauledat, G. Dornhege, B. Blankertz, G. Curio, and K. Müller. The Berlin brain-computer interface for rapid response. *Biomedical Tech*, 49(1):61–62, 2004.

[11] S. Kyostia. Frets on Fire: http://fretsonfire.sourceforge.net, 2006.

[12] H. Lee and S. Choi. Pca+ hmm+ svm for EEG pattern classification. In *Proceedings of the Seventh International Symposium on Signal Processing and Its Applications*, volume 1, pages 541–544. IEEE, 2003.

[4] If the marker timing is the only reason of the performance hit in air drumming, then real drumming blocks could be used to train classifiers as the marker timing is not a problem when classifying in real time.

[13] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4:R1, 2007.

[14] J. Mellinger et al. An MEG-based brain-computer interface (BCI). *Neuroimage*, 36(3):581–593, 2007.

[15] J. Millan and J. Mouriño. Asynchronous BCI and local neural classifiers: An overview of the adaptive brain interface project. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):159–161, 2003.

[16] K. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz. Machine learning techniques for brain-computer interfaces. *Biomedical Technologies*, 49(1):11–22, 2004.

[17] G. Muller-Putz and G. Pfurtscheller. Control of an electrical prosthesis with an SSVEP-based BCI. *IEEE Transactions on Biomedical Engineering*, 55(1):361–364, 2008.

[18] A. Nijholt, J. van Erp, and D. Heylen. *BrainGain: BCI for HCI and Games*. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2008.

[19] G. Pfurtscheller, C. Brunner, A. Schlogl, and F. Lopes da Silva. Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006.

[20] G. Pfurtscheller, J. Kalcher, C. Neuper, D. Flotzinger, and M. Pregenzer. On-line EEG classification during externally-paced hand movements using a neural network-based classifier. *Electroencephalography and Clinical Neurophysiology*, 99(5):416–425, 1996.

[21] G. Pfurtscheller, C. Neuper, and N. Birbaumer. *Human brain-computer interface: Motor Cortex in Voluntary Movements*. New York: CRC Press, 2005.

[22] H. Shibasaki and M. Hallett. What is the Bereitschaftspotential? *Clinical Neurophysiology*, 117(11):2341–2356, 2006.

[23] R. Tomioka, K. Aihara, and K. Muller. Logistic regression for single trial EEG classification. *Advances in Neural Information Processing Systems*, 19:1377, 2007.

[24] G. Townsend, B. Graimann, and G. Pfurtscheller. Continuous EEG classification during motor imagery-simulation of an asynchronous BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(2):258–265, 2004.

[25] R. Versteeg. Brains on Fire. August 2010.

[26] N. Weiskopf, K. Mathiak, S. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, and N. Birbaumer. Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Transactions on Biomedical Engineering*, 51(6):966–970, 2004.

[27] D. Zhang, J. Wang, F. Wang, and C. Zhang. Semi-supervised classification with universum. In *SIAM International Conference on Data Mining (SDM)*, pages 323–333. SIAM, 2008.

[28] Q. Zhao, L. Zhang, and A. Cichocki. EEG-based asynchronous BCI control of a car in 3D virtual reality environments. *Chinese Science Bulletin*, 54(1):78–87, 2009.