

Variability and nonstationarity in Brain Computer Interfaces

Linsey Roijendijk, 0213942
Master Thesis Artificial Intelligence
Radboud University Nijmegen

Supervisor: Jason Farquhar

October 13, 2009

Abstract

A well-known problem in Brain Computer Interfacing (BCI) research is the large degree of variability in the acquired data. Every user has different brain signals, and also the performance of a user varies widely between sessions, within a session, between online and offline settings, and even from epoch to epoch. This variation can be caused by many factors such as psychological factors (e.g., fatigue, mood), physiological factors, change of task involvement (offline versus online task), learning, or measurement artifacts and noise. For maintaining good BCI performance, a BCI must be able to adapt to these changes.

In our research we attempt to characterize these variability types in a BCI context. Therefore, we have conducted an EEG imagined movement experiment with 8 participants taking part in two sessions. Participants received two types of feedback, namely continuous during movement imagination, or once immediately after movement imagination. The performance of the BCI and some simple bias and gain adaptation methods for improving the feedback were analyzed.

Contents

1	Introduction	5
1.1	Measuring techniques	6
1.2	Brain patterns and mental tasks	7
1.2.1	Imagined movement	8
1.3	Variability and nonstationarity in BCI	10
1.3.1	Adaptive BCIs	11
1.4	Our study	12
2	Methods	13
2.1	Experiment	13
2.1.1	Equipment	13
2.1.2	Participants	13
2.1.3	Task	14
2.1.4	Design	14
2.1.5	Procedure	20
2.2	Data analysis	21
2.2.1	Preprocessing	21
2.2.2	Feature selection and classification	21
3	Results	23
3.1	Plots	23
3.2	Classification performance	24
3.2.1	Online	25
3.2.2	Offline	27
3.3	Bias and Gain Analysis	29
3.3.1	Bias	29
3.3.2	Gain	31
4	Discussion and conclusions	34
4.1	Inter-user variability	34
4.2	Inter-session and user learning variability	34
4.3	Feedback dependent variability	35
4.4	Signal nonstationarity	35
5	Future work	36
	Bibliography	37

A	Experiment	40
A.1	Instruction form	40
A.2	General questionnaire	42
A.3	Imagined movement scale	43

Chapter 1

Introduction

Would it not be wonderful if a paralyzed patient (or even you, the reader), would be able to control the performance of a computer-linked device by mental control only? A Brain Computer Interface (BCI) is a system that allows an individual to do just that: to control a computer by thought alone. It measures the brain activity of an individual and translates it into commands for a computer or some other device.

In the present study, we examined variability and nonstationarity in BCIs by means of an imagined movement task. Before the experiment is described, we will first set the stage by explaining what BCIs are and what their purpose is. Next, the BCI field will be narrowed down to a specific type of BCI that uses EEG signals as input and has certain advantages compared to other systems. This is followed by a consideration of the brain patterns and tasks used in BCIs and in particular, the imagined movement task used in our research. An overview will be given of studies on variability and nonstationarity in BCIs. Finally, we will pose the specific research question of this study.

A major goal of BCIs is to restore motor control in disabled patients, suffering from conditions like amyotrophic lateral sclerosis (ALS), spinal cord injury, stroke and cerebral palsy. A BCI could be useful especially for ALS patients in an advanced stage of the disease, when they suffer from the "locked-in syndrome". These patients are conscious and aware of their environment, but are not able to communicate or move. Studies have shown that BCIs can facilitate the communication of people with severe motor disabilities (J. R. Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002; Birbaumer, 2006). Example applications for patients are navigating a wheelchair around (Millán, Renkens, Mourino, & Gerstner, 2004), writing letters (Farwell & Donchin, 1988), or controlling an internet browser (Bensch et al., 2007).

Besides BCIs for patients, BCIs for healthy users are being developed. Cognitive and affective BCIs use neurophysiological signals for automatical detection of the cognitive or affective state of a user and they feed this information back to the user. For example, an EEG-based BCI was used to detect and mitigate high mental workload in people driving a car (Dornhege et al., 2006). Gaming BCIs provide gamers with an extra dimension for control (Nijholt, 2009). Brain signals can be used for active control, such as moving a cursor on the screen, or for passive control, in which the game adapts itself to the affective state of a user.

A schematic overview of a BCI system is shown in Figure 1.1. A BCI records brain signals when users perform a mental task. Different mental tasks induce different brain signals and the BCI tries to distinguish these signals. Sometimes specific stimuli have to be shown

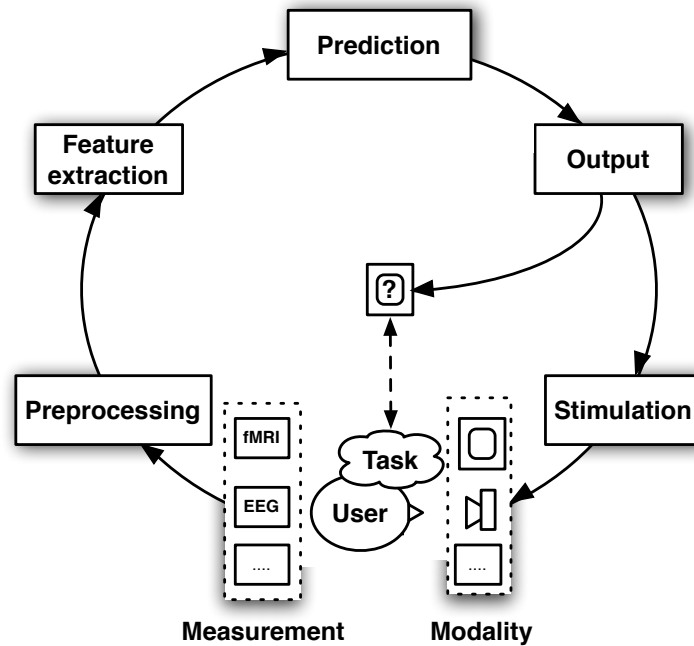


Figure 1.1: The BCI-cycle (Gerven et al., 2009)

to the user to produce the desired signals, such as flashes on a screen. The computer analyzes the recorded brain signals and tries to predict which mental task was executed. Before making a prediction, the data are first preprocessed to remove noise and artifacts, and only the features of interest (e.g., certain spatial positions, frequencies) are extracted for use in the classification process. The predictions can be used to alter stimuli (e.g., moving a ball on the screen) or control an external device like a wheelchair.

State-of-the-art BCI systems use advanced machine learning techniques, that adapt to the user, instead of training a user to adapt to the BCI. The advantage of this approach is that all that is needed is a calibration session of about 20 minutes, rather than a user training that takes about 50-100 hours (Müller, Krauledat, Dornhege, Curio, & Blankertz, 2004).

1.1 Measuring techniques

Several measurement techniques are used for recording brain signals in a BCI. Invasive systems (using electrocorticographic (ECoG) or intracortical activity) as well as non-invasive systems (electroencephalography (EEG), magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), and Near Infrared Spectroscopy (NIRS)) are currently used in BCI research. Invasive systems have a high temporal and spatial resolution compared to non-invasive systems, but an operation is needed to implant the electrodes. Non-invasive systems are less intrusive for users and less expensive, and are therefore used more often in BCI research.

The most used measuring technique in BCI research is the EEG (Mason, Bashashati, Fattourehchi, Navarro, & Birch, 2007). The EEG is the recording of electrical activity on the scalp

produced by neuronal activation. The activity is usually measured from multiple electrodes placed on the scalp. It often contains noise, generated by other sources than the brain such as artifacts due to ocular movement (EOG), cardiac movement (ECG) or muscular activity (EMG).

The temporal resolution of the EEG is higher than for fMRI and NIRS, but its spatial resolution is lower, because the number of electrodes applied to the surface of the skull is limited. Layers of tissue, skull, and hair between the electrodes and the cortex cause spatial smearing, which makes it difficult to localize the source of the signals. For BCI purposes, EEG is easier to use and less expensive than the other non-invasive measuring techniques. Besides, the apparatus can easily be transported out of the scientific environment, for example, to the home of a patient. This is not yet possible for MEG and fMRI. In this thesis we will focus on EEG.

1.2 Brain patterns and mental tasks

A range of brain patterns are used for controlling a BCI, including steady-state evoked potentials (SSEPs), event related potentials (ERPs), slow cortical potentials (SCPs), and sensorimotor rhythms (SMRs).

SSEPs (Regan, 1977) are neurological responses evoked by external oscillations. In BCIs, stimuli that oscillate at different frequencies are presented to the users. The oscillation patterns of these frequencies can be found back in the EEG. When a user selectively attends to one of the stimuli, the strength of that oscillation pattern increases. Different modalities, e.g., visual (Allison et al., 2008), somatosensory (Nangini, Ross, Tam, & Graham, 2006), and auditory (Ross, Herdmann, & Pantev, 2005) evoked potentials have been investigated for BCI purposes.

A more frequently considered brain wave component is the P300 evoked potential (Farwell & Donchin, 1988). The P300 is a positive deflection in the EEG at about 300 ms after stimulus presentation, which can be elicited by a rare event. Attention is required for evoking the P300. A speller application with a letter matrix with flashing rows and columns that evokes a P300 (Farwell & Donchin, 1988) is often used in BCI research. No user training is required in this case.

SCPs are slow voltage changes in the EEG that occur over several seconds, and reflect the amount of activity of the underlying cortical areas. Birbaumer et al. (1999) showed that people can learn to control SCPs and can use the SCPs to control a cursor on a screen. However, SCPs are relatively slow patterns compared to the other brain patterns, and learning to control them is very time consuming. Participants needed more than 280 sessions of 5-10 minutes to reach a stable performance of at least 75 percent correct responses (Birbaumer et al., 1999).

Other widely used signals are the SMRs, which include the μ (8 - 12 Hz) and central β (13 - 28 Hz) oscillations. Most people can modulate the power in these frequency bands by performing real or imagined movements. These signals and the Lateralized Readiness Potential (LRP) that occurs before (imagined) movement will be discussed further in Section 1.2.1.

Furthermore, the error potential (EP) has been investigated in BCIs to improve the reliability of the systems (Schalk, Wolpa, McFarland, & Pfurtscheller, 2000; Buttfeld, Ferrez, & R. Millán, 2006). An error potential is an evoked potential occurring when a person makes

a response error. However, an EP is also generated in response to an error made by an interface as a misinterpretation of a command that a user has given. Therefore, the EP is used in BCIs to detect whether the outcome of an epoch matches with that desired by the user.

1.2.1 Imagined movement

Imagined movement is a widely used mental task in BCI research. During actual movement, specific brain patterns can be determined in the EEG before, during, and after the movements. Before movement onset, an LRP occurs, which is a negative event-related brain potential, that reflects the preparation of motor activity on a certain part of the body (Krauledat et al., 2004; Wang et al., 2004). During a movement, the power in the sensorimotor frequency bands (Event Related Desynchronisation, ERD) decreases. Following the movement, the power increases again in these frequencies, which is called Event Related Synchronization (ERS or beta rebound) (Pfurtscheller & Silva, 1999). The sensorimotor frequency bands include μ (8 - 12 Hz) and central β (13 - 28 Hz) oscillations.

Research has shown that imagined movements produce similar but attenuated signals as actual movements (McFarland, Miner, Vaughan, & Wolpaw, 2000; Kranczioch, Mathews, Dean, & Sterr, 2009). Probably, the signals are less pronounced due to the duality of the task. A participant has to imagine moving, but at the same time avoid performing the actual movement. In particular, hand and feet motor imagery seem to affect sensorimotor EEG rhythms in such a way that these tasks can control a BCI with a high reliability (Pfurtscheller, Brunner, Schlögl, & Silva, 2006). Left hand, right hand, and feet movements are relatively easy to discriminate in recorded brain signals, because the body parts entail a large area of the sensory motor cortex and are located on different locations (see Figure 1.2).

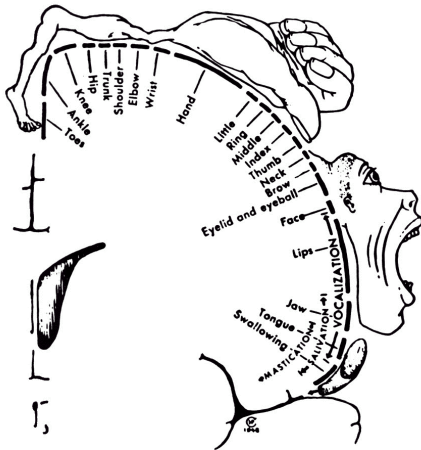


Figure 1.2: Homunculus showing the location of various body parts in one hemisphere of the sensory motor cortex. The larger the body part, the more sensitive the brain is to that body part.

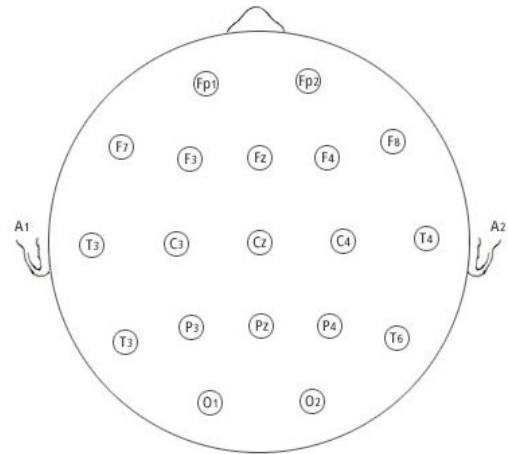


Figure 1.3: International 10-20 system (Klem et al., 1999) illustrating the standard placement of EEG electrodes on the head. The letters and numbers correspond to different areas of the brain.

Figure 1.4 shows the signals that typically can be detected when using a left versus right hand imagined movement paradigm. Often, right hand movements produce signals in the

sensorimotor cortex around electrode C3, and left hand movements produce signals around electrode C4 (see Figure 1.3 for the spatial locations of C3 and C4 on the scalp). However, the frequencies and spatial locations of the signals caused by (imagined) movement can differ from participant to participant. Another pattern, however not shown in Figure 1.4, is the 'focal ERD/surround ERS' (Pfurtscheller & Silva, 1999). This is the observation that not only an ERD occurs during the movement in the sensorimotor rhythms, but also an ERS in neighboring cortical areas.

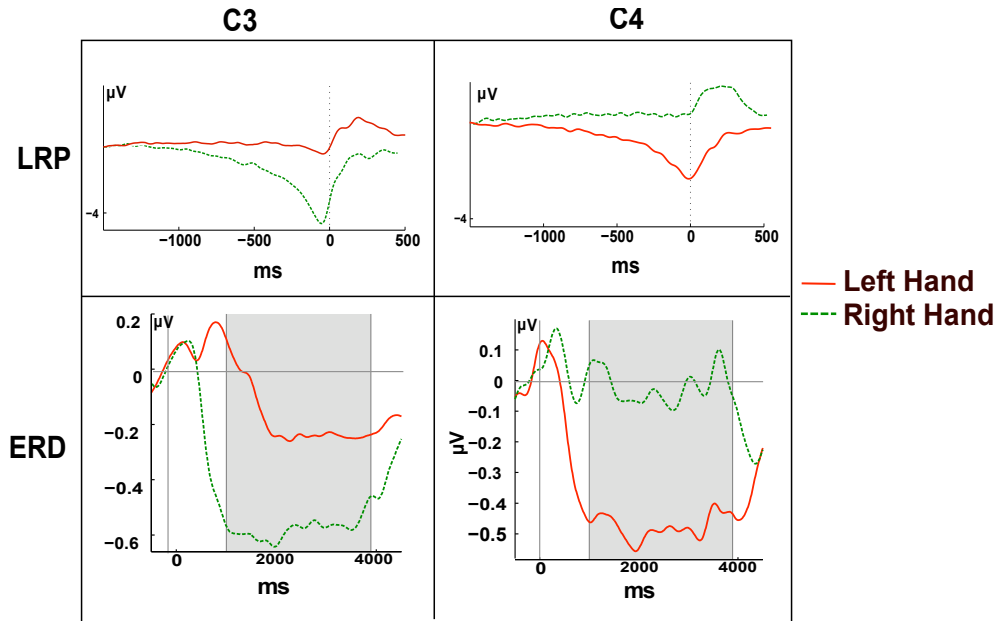


Figure 1.4: Example of EEG signals that can be detected before and during (imagined) movement in the C3 and C4 electrodes. The upper plots show LRPs, occurring contralateral before stimulus onset. The lower plots show the EEG signal in the μ frequency band (8 - 14 Hz), showing a desynchronization short after movement onset (ERD), and again a synchronization afterwards (ERS). The figures are adapted from Blankertz, Dornhege, Krauledat, et al. (2006); Blankertz et al. (2008).

Motor imagery can be performed either in a visual or in a kinesthetic way. When visual motor imagery is involved, users imagine seeing themselves or someone else performing a movement. During kinesthetic imagery, users imagine that they are performing the movement themselves. Neuper, Scherer, Reiner, and Pfurtscheller (2005) showed that the typical sensorimotor rhythms did not occur when visual motor imagery was used. Therefore, it is necessary to use kinesthetic imagery in BCIs.

The motor imagery paradigm is often used in BCIs, because it can be performed spontaneously (without needing an external stimulus) and the signals are relatively straightforward to detect. However, a disadvantage of using this paradigm is that the sensorimotor rhythms cannot be discerned in every person, a phenomenon called BCI illiteracy. Nijholt et al. (2008) estimated that about 20 percent of participants do not show strong enough motor-related mu-rhythm variations for effective motor imagery BCI, although training may improve this aspect (Vidaurre & Blankertz, 2009).

1.3 Variability and nonstationarity in BCI

A significant issue in BCI research is the degree of variability in the signals used in BCIs (Curran & Stokes, 2003; Wolpaw & Jonathan, 2007; Shenoy et al., 2006; Krauledat, 2008). Although most BCIs are used in highly controlled laboratory environments, much variability is present between and within users. Every user is characterized by different brain signals, but the user's brain patterns also vary widely between sessions, within a session and even from epoch to epoch (Blankertz, Dornhege, Lemm, et al., 2006; Blankertz, Dornhege, Krauledat, Müller, & Curio, 2007). This problem occurs with invasive as well as with non-invasive BCI systems (Wolpaw & Jonathan, 2007).

A special case of variability is "nonstationarity", meaning in this context that the characteristics of the signal change over time, such that the performance of a trained classifier changes on data recorded at a different time point. A mathematical definition of nonstationarity in BCI context is given in (Krauledat, 2008).

The variation in the data can be caused by many different variables such as psychological factors (e.g., fatigue, mood, motivation (Bartolic, Basso, Schefft, Glauser, & Titanic-Schefft, 1999; Birbaumer et al., 1999)), physiological factors, change of task involvement, user learning, measurement artifacts and noise.

Changes in the data occurring over time can cause previously functioning BCIs to fail. However, BCIs in real-life situations must be robust in their application. A patient needs a system that functions all the time and is able to adapt to new circumstances. Therefore, it is necessary to investigate this variability and find solutions for improving the BCIs.

Not many researchers have tried to characterize the variability and nonstationarities in BCI systems. To date, only Guger, Edlinger, Harkam, Niedermayer, and Pfurtscheller (2003) have investigated BCI accuracy on a large user population. Ninety-nine people participated in an imagined movement experiment with feedback. Classification performance is shown in Figure 1.5. Also Blankertz et al. (2008) investigated the classification performance accuracy between participants. Therefore, they used the Berlin Brain Computer Interface (BBCI) that adapts to specific users: Best frequency bands, best spatial locations and best task (left hand versus right hand, or left hand versus feet, or right hand versus feet) were selected for each participant, which was not done in the research of Guger et al. (2003). The experiment with 14 BCI-naïve participants showed that 8 participants obtained an accuracy over 84%, and 4 participants an accuracy over 70% (see Figure 1.5). In addition, researchers investigated the user variability with respect to spatial patterns and spatiotemporal characteristics of brain signals (Blankertz et al., 2007; Pfurtscheller et al., 2006).

User learning and signal changes caused by feedback have been investigated before by Neuper, Schlögl, and Pfurtscheller (1999) in an imagined movement paradigm. Four trained participants reached 85% to 95% classification accuracy in the course of multiple experimental sessions. The EEG data revealed a significant ERD over the contralateral central area in all participants, and two participants simultaneously displayed ERS over the ipsilateral side. During feedback presentation the ERD/ERS patterns showed increased hemispheric asymmetry compared to initial control sessions without feedback. Vidaurre, Schlögl, Cabeza, Scherer, and Pfurtscheller (2006) further showed that participants using an BCI with feedback during multiple sessions improved their performance in subsequent sessions.

The first quantitative study about nonstationarities in a feedback BCI was performed by Shenoy et al. (2006). They presented a systematic quantitative study involving the imagined movement paradigm in which multiple (experienced) participants were recorded during

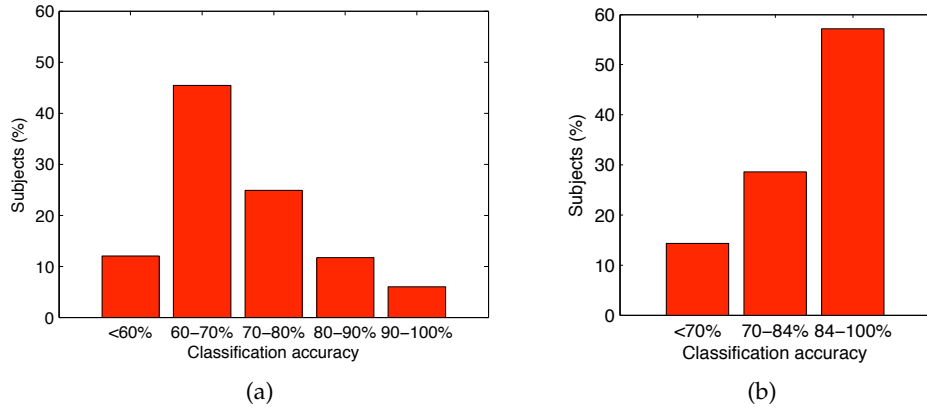


Figure 1.5: Classification accuracy obtained from (a) 99 participants (Guger et al., 2003) and (b) 14 participants (Blankertz et al., 2008).

offline (without feedback) and online (with feedback). Changes were observed in the EEG data between offline and online settings, and within a single online session of 30 minutes. The first change could be interpreted as a shift of the data in the feature space, due to the difference of the online task, which caused different background activity in the brain signals. Especially, the parietal α rhythm (idle rhythm of the visual cortex) increased during the online task, possibly due to a higher demand for visual processing.

1.3.1 Adaptive BCIs

To deal with nonstationarities, adaptive BCIs have been built. Adaptive BCIs try to improve the classification rates by using the classifier from a previous block or session, but adapting it to new data. Typically, the mean of the feature distribution changes between sessions and adapting the classifier can easily compensate for these changes.

Several studies have shown that adapting the classifier improves participant performance (Shenoy et al., 2006; Vidaurre, Schlögel, Blankertz, Kawanabe, & Müller, 2008; Tsui & Gan, 2008).

Shenoy et al. (2006) simulated adaptive classification techniques with common spatial patterns (CSP) based features to improve the classification rates. Shifting the classification results with a bias based on data of an initial window, all labeled data up to the current point, or data from a window over the immediate past improved the classification accuracy compared to the original classification accuracy. Retraining the LDA classifier with the labeled online data using the features from the offline dataset also improved the classification performance.

Vidaurre et al. (2008) simulated three types of unsupervised adaptation procedures based on linear discriminant analysis (LDA) with a pooled covariance matrix. The first method reduced the bias in the classification results by updating the common mean, the second method updated both the mean and the covariance matrix, and the last method applied a rotation in the feature space. All three methods reduced the error rates compared to no adaptation.

Tsui and Gan (2008) also simulated three methods for updating an LDA classifier during

a session. They used a relatively simple method in which a new classifier was created based on updated means and variances. In addition, they assessed the application of two more advanced methods using a Kalman filter and an extended Kalman filter for adapting the LDA's parameters. All three methods improved the classification performance compared to no adaptation and no significant differences were found between the three methods. The adaptation seemed to improve the classification accuracy less for participants who already achieved a good accuracy without adaptation and more for participants with low initial accuracy.

1.4 Our study

As our previous review shows, little quantitative research has so far been done about variability and nonstationarity within BCIs, although it is considered to be an important issue. In several studies, new algorithms were applied, and adapted BCIs were developed. However, few studies have tried to characterize the variability and nonstationarity in EEG-based BCI data directly.

In our study, we therefore investigated the characteristics of the variability of EEG-based BCIs in more detail. We investigated several types of variability:

- inter-user variability: Variability across the population of users, for example, the location and the frequencies of the desired brain signal.
- inter-session variability: When the same user is re-attached to the recording equipment, neither the equipment nor the user will be in exactly the same state. For example, the user can be more fatigued or less motivated in a second session. In addition, the electrode impedances and the measurement noise can be different.
- user learning variability: When a user makes use of the system, the user learns to adapt to the system. A BCI system will give feedback to the users and the brains of the users will try to adapt to improve the user's performance. Performing a certain task more often can also induce different brain signals (e.g., automatizing movements).
- feedback dependent variability: When a user is performing a task offline, the stimuli are often different from the stimuli online, possibly causing changes in the brain signal. In addition, the feedback itself could influence the mental state of the user (e.g., mood, motivation). For instance, a game could engage users.
- signal nonstationarity: During a session the signal can change over time and therefore periodic adjustments are necessary to reduce the impact of transient effects, such as external noise and different user mind states (e.g., tiredness).

To address these variabilities, we conducted a multi-session, multi-feedback, imagined movement experiment using EEG. Multiple sessions were measured to look at the inter-session and user learning variability. Two types of feedback were used for investigating the feedback dependent variability. The duration of a single session was more than 2 hours, to allow an investigation of the intra-session variability, or signal nonstationarities.

Chapter 2

Methods

2.1 Experiment

We conducted a BCI experiment with multiple participants, multiple sessions, and long sessions of more than 2 hours to investigate variability between users, between sessions, between feedback types, and within sessions. We used an EEG based BCI that is controlled by imagined movement. In this section we describe the method used in this experiment.

2.1.1 Equipment

EEG recordings were carried out using 64 AG/AgCl active electrodes disposed according to the extended international 10 - 20 system (Klem et al., 1999). The offsets of the electrodes were kept below 20 mV, when possible. A Biosemi ActiveTwoAD-box amplifier (BioSemi, Amsterdam, <http://www.biosemi.com>) was used for the recordings. The signals were sampled at 2048 Hz.

Electro-oculographic activity (EOG) was measured for detection of eye movements and blinks. Vertical EOG was measured between two electrodes placed at locations directly above and below the left eye. Horizontal EOG was measured by two electrodes at the outer canthi of the left and right eye. Furthermore, hand movements were measured by electromyographic (EMG) activity. These signals were measured for detecting possible movement artifacts during the imagined movement task. Surface EMG was recorded from the main wrist extensors of the right and left arm. Movements of each individual finger were recorded as well, using very sensitive drum pads on which the participant's fingers were placed.

The recordings took place in an electrically shielded room in order to minimize environmental noise in the EEG signals.

The experiments were run with BrainStream, a BCI software platform developed in Nijmegen, in combination with PsychToolbox (<http://psychtoolbox.org>) for stimulus display.

2.1.2 Participants

Eight students (five men and three women, 20-24 years old) participated in this study for two sessions. Seven participants were right handed and one was left handed. All participants were free of neurological disorders and had normal or corrected-to-normal vision.

All participants had never participated in an imagined movement BCI experiment before and gave informed consent after the experimental procedure had been explained to them. Participants were paid for their participation.

2.1.3 Task

A two class time-locked imagined movement paradigm was used in this experiment. Participants had to perform (kinesthetic imagined) left hand movements and (kinesthetic imagined) right hand movements. In each epoch participants had to imagine or execute a movement for 3.6 seconds. During these 3.6 seconds a metronome was played to the participant and the participant had to perform one movement each metronome tick. The metronome is used to make sure that every participant is performing the same task and it could be that participants imagine a movement in a better way when provided with external movement-related stimuli (Heremans et al., 2009).

Pilot experiment

To determine the type of movement and the tempo of the metronome a small pilot experiment was conducted. Two types of movements were investigated in this pilot, namely hand tapping and hand drumming. Hand tapping is moving your hand up and down while the wrist rests on a table. Tapping is widely used in imagined movement research. A new type of movement we tried in this pilot is hand drumming. Hand drumming is moving the fingers of a hand up and down, one by one, starting with the little finger and ending with the index finger. During drumming all the fingers get activated one by one and our hypothesis was that therefore a larger area in the motor region becomes activated, resulting in a larger measured signal. Two metronome tempos, a sound every 300 milliseconds and a sound every 600 milliseconds were tested.

Four participants performed 8 blocks (two times tapping-300, tapping-600, drumming-300, drumming-600) of 7 sequences of movements (3 actual movement, 4 imagined movement). Each sequence contained 12 epochs, counterbalanced over the left and the right hand.

Each epoch started with a cue (an arrow pointing to the left or to the right) on the screen for 1.8 seconds, while the metronome started ticking (see Figure 2.1). Then, a fixation cross was shown on the screen and the participant had to perform or imagine the movement at the tempo of the metronome for 3.6 seconds. Next, the screen turned black for a second. The cues on the screen were coloured red when it was a cue for imagined movement, and white when it was a cue for actual movement.

The data of each condition was analyzed using preprocessing and classification similar to the final experiment as described in Section 2.2. The average classification performance is shown in Table 2.1. These results seemed to show a slight benefit for the "drum 300 ms" condition, especially in the actual movement condition. Therefore, we used drumming with a 300 milliseconds metronome as the task in the final experiment.

2.1.4 Design

In this section, we will first introduce the feedback conditions used in the experiment. Next, we outline the experiment design in a top-down style. First the design of the sessions, then the design of the separate blocks and finally the design of the epochs will be explained.

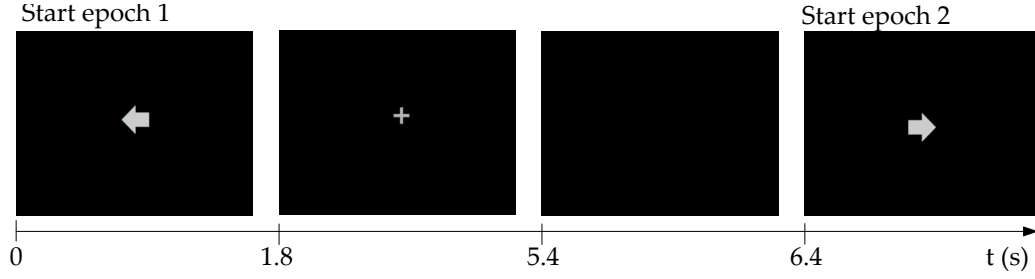


Figure 2.1: Epoch timeline in the pilot experiment.

	Average (Standard Error) (%)	
	Actual movement	Imagined movement
tap 300 ms	72(7)	64(6)
tap 600 ms	77(7)	65(7)
drum 300 ms	87(1)	71(2)
drum 600 ms	78(3)	67(8)

Table 2.1: Average classification performance (in %) in the pilot experiment with 4 participants, which was conducted to determine the type of movement and the tempo of the metronome for the final experiment. The "drum 300 ms" condition seemed to give the highest classification performance across the participants.

Feedback

To investigate the feedback dependent variability, we decided to use two different feedback types in our experiment, namely epoch feedback and game feedback. During epoch feedback, the participants performed the task and immediately after this the participants received feedback whether the computer recognized the correct hand. During game feedback, the participant received continuous feedback during execution of the task. Furthermore, the game feedback gave an indication of how well the computer could predict the hand direction. The feedback is described in more detail in Section 2.1.4. The two feedback types were used, to compare continuous and discontinuous feedback, because each type could influence the participant in a different way. The game feedback could distract the users more from their task, because of the continuous change of the incoming stimuli. Participants could also be more motivated in the game feedback, because points will be given for good performance.

Sessions

Each participant took part in two sessions of about 3 hours in which the second session took place two days after the first session. We measured people on different days, because this resembles the use of a BCI outside of the laboratory, when people will also use it on different days. By measuring on different days we took day-to-day variabilities such as user alertness, cap fitting differences, and effects of sleep into account. Sleep is known to enhance

motor performance, and also imagined movement can enhance motor performance after sleep (Debarnot et al., 2009).

Blocks

A session was divided into 8 blocks of about 12 minutes each. Figure 2.2 shows the timeline of the blocks in our experiment.

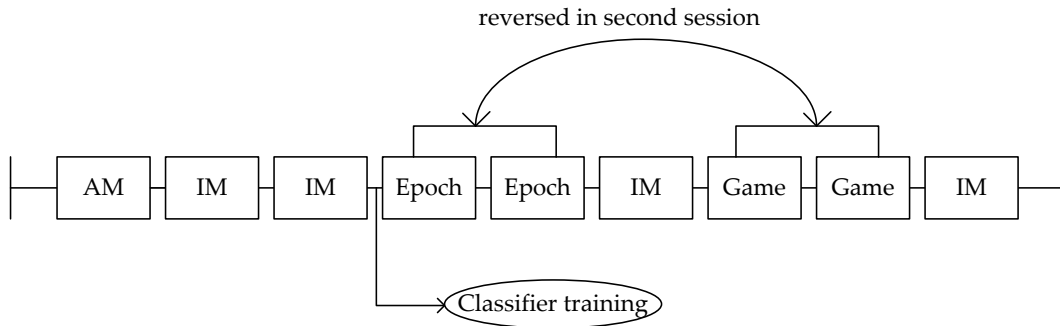


Figure 2.2: Timeline of the experiment including an actual movement block (AM), imagined movement blocks (IM), epoch feedback blocks (EPOCH), and game feedback blocks (GAME).

The recordings began with an actual movement block (AM), so the participant could practice the movement, the experimenter could check if the participant performed the task correctly, and if necessary could correct the participant. Then, two training imagined movement blocks (IM) were performed, in which enough data was accumulated for training the classifier needed in the feedback blocks. After this data was recorded, classifiers were trained as described in Section 2.2. Next, the participants performed two feedback blocks, for example epoch feedback blocks (Epoch). Then, the participants had to perform another imagined movement block without feedback. This "probe" block was used to detect signal changes during the experiment without the influence of the feedback stimuli. Then again, the participants performed two feedback blocks, this time with the other type of feedback, game feedback in the example. Finally, another imagined movement block without feedback was executed as "probe" block. When the first feedback type in session 1 was epoch feedback, the first feedback type in session 2 was game feedback and vice versa. Which feedback type was used first in the first session was counterbalanced across the participants.

Each block consisted of 72 epochs, divided in 6 sequences of 12 epochs in which left and right movements were counterbalanced. In a sequence a similar movement occurred at most three times in a row.

Epoch

Three different epoch types were used in the experiment, namely (imagined) movement without feedback, imagined movement with epoch feedback and imagined movement with game feedback. All three epoch types will be explained separately here.

No feedback First, we outline the epochs that are used during the actual and imagined movement blocks without feedback. Figure 2.3 shows what happens during such an epoch. First, the participants saw an arrow on the screen, indicating whether they should (imagine) drumming with their left or right hand. When the arrow appeared on the screen, the metronome started ticking every 300 ms. After 1.2 seconds a fixation cross appeared and the participants had to start drumming their hand, while the metronome kept ticking. After a further 3.6 seconds the screen turned black and the participants could relax for a second.

Compared to the pilot experiment, we reduced the time in which the cue is on the screen from 1.8 seconds to 1.2 seconds. We chose to do this, because it seemed more logical to use 4 metronome ticks in stead of 6 ticks before starting, like musicians often do. We also needed the epochs to be a bit shorter, to keep the total experiment time within 3 hours.

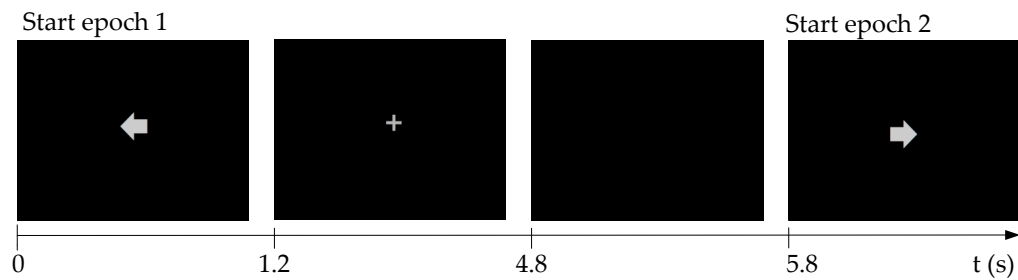


Figure 2.3: Epoch timeline during blocks without feedback.

Epoch feedback The epochs with epoch feedback were as far as possible kept the same as the epochs without feedback. However, an extra screen to give feedback was added after the task performance, see Figure 2.4. Feedback was given by showing either 'correct' in green letters or 'incorrect' in red letters on the screen. The number of correct epochs and the number of epochs performed in the current block were also shown on the screen, so the participant knew his own performance level. The feedback was shown immediately after the epoch data was classified, on average this took about 100 ms.

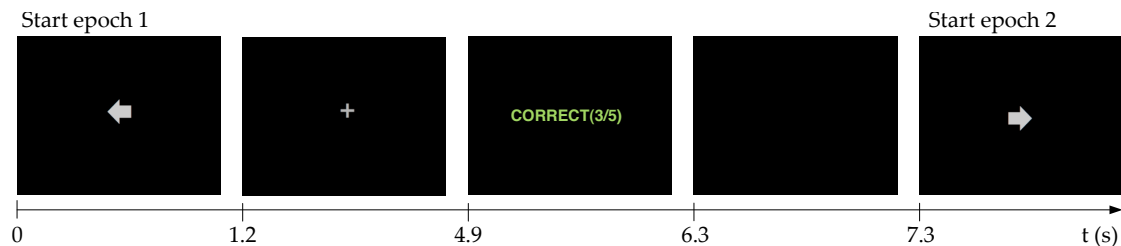


Figure 2.4: Epoch timeline during epoch feedback blocks.

Game feedback For the game feedback, we used a variant of the basket game (Blankertz et al., 2007). This game has been used frequently in previous BCI research. Figure 2.5 shows

how our basket game looks. In the middle of the top of the screen a ball was shown that would fall down into one of the baskets shown at the bottom of the screen. The horizontal position of this ball was controlled by imagined movements. The goal of the game was to move the ball to one of the green highlighted baskets, that indicated whether a left or right hand movement had to be imagined. The further to the side of the screen the ball dropped in a basket, the more points a participant received. The basket scores from the middle to the side of the screen are 1, 5, 10, and 25. After finishing an epoch, the basket score was multiplied by a multiplier value and then added to the total score (both shown in the upper right corner of the screen). Every time the ball was dropped into a correct basket four times in a row, the multiplier increased by one. When a ball was dropped into an incorrect basket the multiplier decreased to 1 again. The maximal value of the multiplier was four.

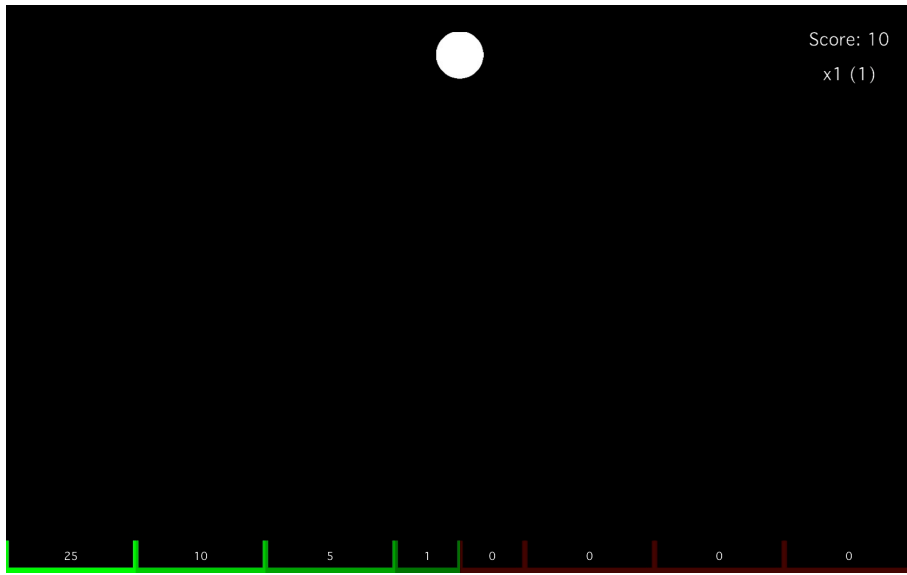


Figure 2.5: Screenshot of the game used in the experiment.

Figure 2.6 shows the timeline of a game epoch. At the beginning of an epoch, the ball appeared in the middle of the screen and the metronome started ticking for 1.2 seconds. Next, the ball started falling down and the participant had to start imagining a left or right hand movement at the tempo of the metronome. After 3.7 seconds the ball fell into a basket and the participant could relax for a short moment before the next epoch started.

The horizontal ball position was updated every 200 ms based upon the BCI output. When the output predicted a left movement, the ball moved to the left side of the screen and when the classifier predicted a right movement, the ball moved to the right side of the screen. The prediction value determined how far the ball moved to the side of the screen. More specifically, every 200 ms a data window of 600 ms was classified and used to determine the next horizontal position of the ball on the screen (in total there were 16 windows each epoch). The horizontal position of the ball on the screen, $x(b)$, was calculated based upon the average of the classifier predictions, using:

$$x(b) = \mathcal{L}(-\sum_i F_i/16) \quad (2.1)$$

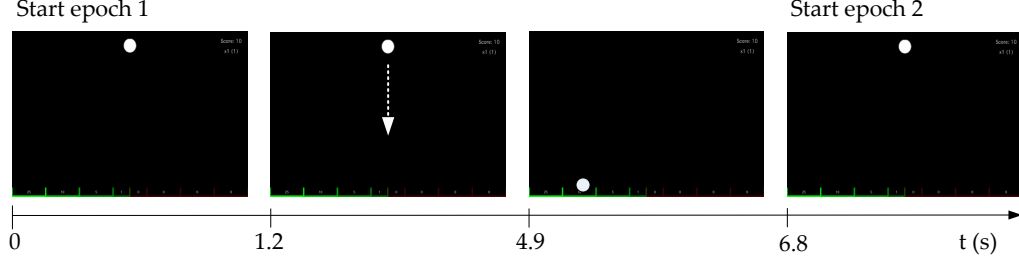


Figure 2.6: Epoch timeline during game feedback blocks.

where, F_i is the classification result of data window i of the current epoch, and $\mathcal{L}(\cdot)$ is the logistic function which maps from the classifier decision values' to class probabilities, as described in Formula 2.2. A moving average was used to give smoother feedback to the user. Furthermore, the use of multiple windows makes the feedback more reliable, because more data is used to determine the class.

Previous research showed that adapting a classifier during an experiment improved the classification performance of the participants (see Section 1.3.1). Therefore, we decided to add bias and gain adaptation to the game feedback. Bias adaptation tries to compensate for shifts occurring in the classification results. Gain adaptation tries to scale the classification results to improve the feedback on the screen, so that the participant is able to reach every possible horizontal position.

We used a very simple kind of bias adaptation in which the previous twenty decision values were averaged and subtracted from the current decision value. In this way the mean of the decision values was kept around zero, exactly in between the classes (all positive results were left, and all negative results were right). We thought this would ensure that the game results were evenly divided between the classes. To adapt the gain, we divided the decision value by the standard deviation of the previous twenty decision values, so that the value of the standard deviation was kept around one and the results could be scaled better onto the screen. For the bias in the first 20 epochs, the average over the already gathered decision values divided by 20 was used. To compute the gain adaptation in the first 20 epochs, we used the standard deviation of the n decision values till now combined with the last $(20 - n)$ decision values of the trainings block.

After we gathered data from the first five participants, we analyzed the results of the bias and the gain adaptation. We had to conclude that the algorithm we were using was too simple. Large outliers appeared in the classification results which skewed the adaptation algorithms. An example of this is shown in Figure 2.7, in which can be seen that the bias adaptation worsens the classification performance due to outliers. Therefore, we added outlier removal. Before computing the bias and gain values on the decision value window, outliers that deviated by more than two standard deviations from the average of the decision value window were removed. We also started using a weighted average for the first 20 epochs of the gain adaptation using a value of 1 (standard deviation of a normal distribution) for the not yet gathered data, instead of using the results from the training block. For information about the bias and gain analyses see Section 3.3.

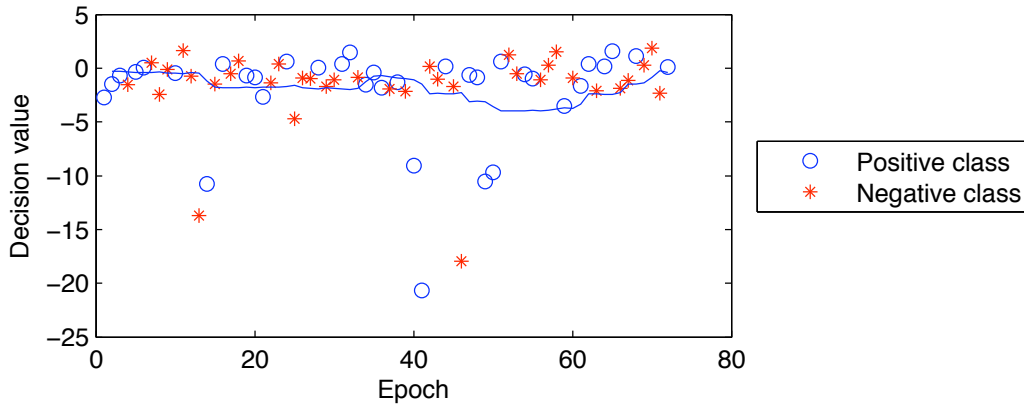


Figure 2.7: Online results of participant 4 during the first epoch feedback block of the first session. The line shows the bias adaptation that was used during the experiment. Everything above the line will be classified as the positive class and everything below as the negative class. The bias is influenced by outliers in such a way that the decision values are not evenly divided between the classes anymore.

2.1.5 Procedure

In the first session of the experiment, the participant started with reading the instructions of the experiment (see Appendix A.1) and signing an informed consent form. In addition to the instructions, the experimenter described the movement while performing it in front of the participant. The participant was also explicitly instructed to use kinesthetic imagery. Next, the researchers connected the EEG, EMG and EOG electrodes to the participant. This preparation process took about 45 minutes in total. The participants were seated in a comfortable chair looking at a 17-inch TFT monitor placed on the table in front of them, on which the stimuli were presented. During the experiment, the participant's hands were hidden from sight in order to encourage the participants to use the kinesthetic strategy of imaginations. After each sequence a participant had to press a foot pedal to continue to the next sequence. After each block the participant had a little break and had to give ratings for three items. First, participants had to give an impression of their alertness on a scale from 1 (absolutely not alert) to 5 (very alert). Furthermore, participants had to scale the effort they put into each block on a scale from 1 (no effort) to 5 (much effort) and after imagined movement blocks a rating of their vividness of imagination had to be given (see Appendix A.3 for the scale). We asked the participants these ratings, to have the possibility to investigate the relation between the experience and the performance of the participants. After the third block there was a longer break of about 10 to 15 minutes, in which the classifier had to be trained. At the end of each session the participant filled in a general questionnaire (see Appendix A.2) to provide the researcher with the participant's background information and an impression of the participant's opinion about the experiment.

A session took about three hours in total. The second session was a bit shorter than the first session, because the participant only got a short reminder from the experimenter about the task, and was already familiar with the procedure.

2.2 Data analysis

This section summarizes all the data analysis methods that have been used. All data analyses have been done using Matlab.

2.2.1 Preprocessing

The first step in handling the recorded EEG data is to preprocess the raw EEG data, because EEG data often contains a lot of environmental noise that masks the real brain signals. Preprocessing was used when analyzing the data offline, but was also needed before applying the classifier in the feedback part of the online experiment. The preprocessing steps for on-line and offline analysis will be explained separately.

Offline

The first step in offline preprocessing was downsampling the raw data from 2048 Hz to 256 Hz. Next, the data recorded during the (imagined) hand movements was sliced into epochs of 3.6 seconds. Linear detrending was applied to the sliced data to remove slow drifts and then the data was re-referenced using an average reference over all channels to improve the signal-to-noise ratio. When the data was preprocessed for creating a classifier, bad epoch and bad channel rejection based on variances was applied, to make the dataset more robust. Specifically, the total power (variance) of each separate epoch, and the standard deviation of all epoch powers were calculated. When the power of an individual epoch deviated more than 3 standard deviations from the mean of all epoch powers, the epoch was removed from the dataset. The bad channel detection worked in a similar way. In this case, channel powers instead of epoch powers were used. When channels were rejected, the data was again re-referenced over channels to remove the noise caused by the now removed bad channels.

Online

During an epoch feedback block, 3.6 seconds of data were extracted every epoch, starting from the appearance of the fixation cross on the screen. During a game feedback block several windows were used: every 200 milliseconds, a window of 600 milliseconds of data was extracted, which resulted in 16 windows of data every epoch. Windows of 600 milliseconds were chosen based on offline analyses on pilot data as they gave the best trade-off between the classification performance and the real time reactivity. Subsequently, roughly the same steps as in the offline preprocessing were executed. First, the data window was downsampled from 2048 Hz to 256 Hz. After that, bad channels, marked during the offline preprocessing of the training data, were removed. Finally, linear detrending and re-referencing was applied.

2.2.2 Feature selection and classification

The next data analysis step was determining the features of interest. First, to select only the sensorimotor rhythms for further analysis, we applied a bandpass filter (Luck, 2005) with a pass band between 7 and 30 Hz. This bandpass filter also removed 50 Hz line noise, slow drifts, cardiac (ECG) and movement artifacts from the data. Following, the data was spatially whitened (Farquhar, 2009). Spatial whitening (or spherising) transforms the data to

a new feature space, in which the new "virtual channels" are uncorrelated with each other and all have unit variance. This prevents the classifier from being biased to high power features. In this way, all features are equally important. Finally, the power of the signals was determined by computing the covariance over space. This feature set contains the information in such a way, that a linear classifier is able to select the most important spatial features (Farquhar, 2009).

The final step was training the classifier. Classifiers were trained using linear logistic regression (Tomioka, Aihara, & Müller, 2007) with L_2 regularization, an algorithm often used in BCI research. Linear logistic regression is a statistical model for binary classification, which creates a hyperplane separating the two different classes from each other. Given feature vector x (in our case the covariance feature set), the probability that this data belongs to the positive class $C+$ is computed as follows:

$$P(C+|x) = \mathcal{L}(x, w) = \frac{1}{1 + e^{-w^T x + b}} \quad (2.2)$$

in which, b is the bias and w is the weight vector, to be found by the classifier training procedure. In logistic regression these values are determined by maximizing the likelihood of the training data given the features. L_2 regularization (Farquhar, 2009) was used to prevent over fitting to the training data. The optimal regularization strength was determined using ten-fold cross validation. In ten-fold cross validation the data is partitioned into ten folds. A classifier is trained on nine subsets and tested on the remaining subset. This process is repeated ten times, in which every subset is used once as test set. Then, the mean of the ten test performances gives an indication of the classifier performance over the entire dataset.

Chapter 3

Results

3.1 Plots

In this section, we show plots of the data to demonstrate that participants have different frequency bands and spatial locations activated during imagined movement.

Figure 3.1 shows topographic displays (topoplots) of the power in specific frequency bands of two participants during the training blocks. A topoplot was made for each class (left and right), to investigate the differences. For participant 1, desynchronization contralateral to the movement (ERD) is visible around the sensorimotor areas and a synchronization lateral. This participant has a strong μ rhythm, however, this is not true for every participant. For example, the difference between the classes of participant 6 is very small and not located around the sensorimotor area.

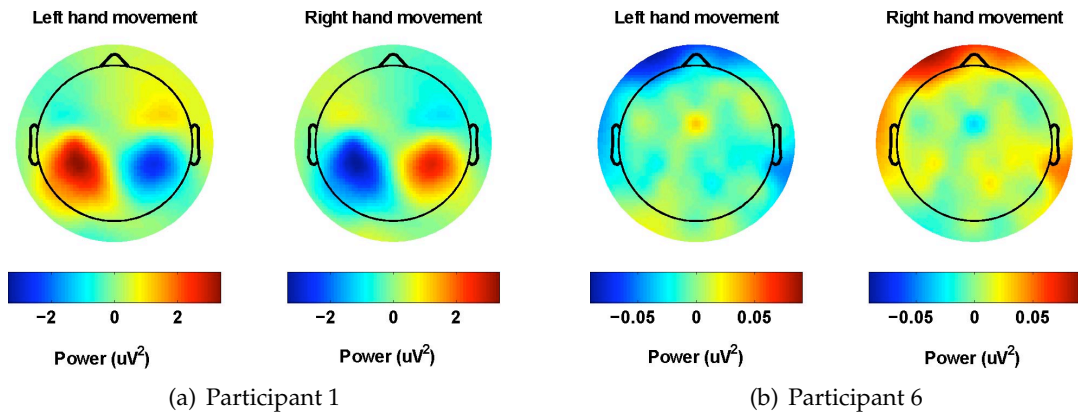


Figure 3.1: Topographical display of the deviation from baseline of the averaged power in the frequency band 8 - 12 Hz (μ rhythm) for both classes in the training blocks of participant 1 and 6 in the first session. As baseline the power of the average of all epochs (left and right) was used. Participant 1 has a clear distinction between the two classes in this frequency band, however participant 6 not.

To demonstrate that participants have different frequency bands and spatial locations activated during imagined movement, we draw area under the receiver operating characteristic curve (AUC) (Bamber, 1975) plots. AUC plots show which frequencies and channels

differentiate the most between the positive and the negative class. Figure 3.2 and 3.3 show AUC plots from the training blocks of two different participants, generated by computing the AUC for each channel and frequency bin over all epochs (in a block) with respect to the true class. The data was transformed from the time domain to the frequency domain using Welch's method (Welch, 1967). This method slices the data into overlapping windows on which the Fast Fourier Transform (FFT) is computed and then the results are averaged. The channels are shown in a topographical view of the brain. Figure 3.2 shows that participant 1 has strong signals around C3 and C4, the area where we would expect that the signals from the hands would appear. The participant shows two different discriminating frequency bands, namely 8-12 (μ rhythm) and 18-24 (β rhythm). Not all persons have those two bands, some only have the lower frequency band. One side of the brain discriminates better for the positive class and the other for the negative class (blue color versus red color). Another interesting aspect shown in the plots is the ERS surrounding the ERD in the sensorimotor area, visible as a change in color the more frontal electrodes (e.g., FC4). Participant 2 (see Figure 3.3) however, shows different frequency bins, signals more to the back of the head, and fewer channels with class dependent value.

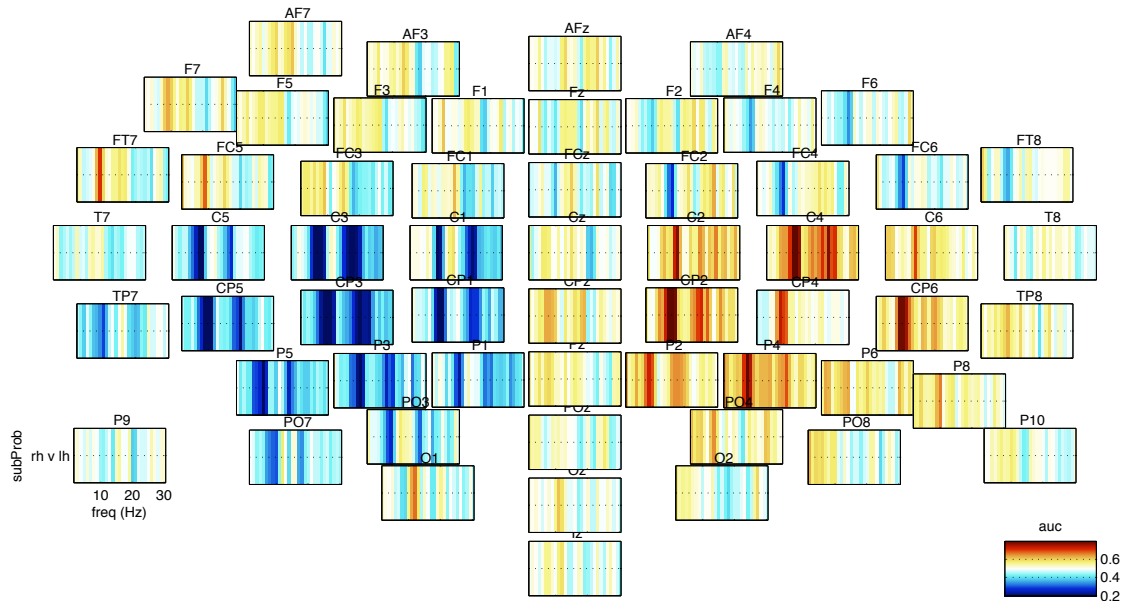


Figure 3.2: AUC plot showing the AUCs for frequencies between 2 and 30 Hz averaged over the epochs of participant 1 in session 1 during the training blocks. The channels are shown in a topographical view. The plot shows high AUC values around C3 and C4, in two frequency bands (8-12 Hz and 18-24 Hz).

3.2 Classification performance

In this section, we compare classification performances to look at the BCI performance variability between participants, between sessions, within sessions and between feedback types. First, the classification results used online (during the experiment) are discussed, and then

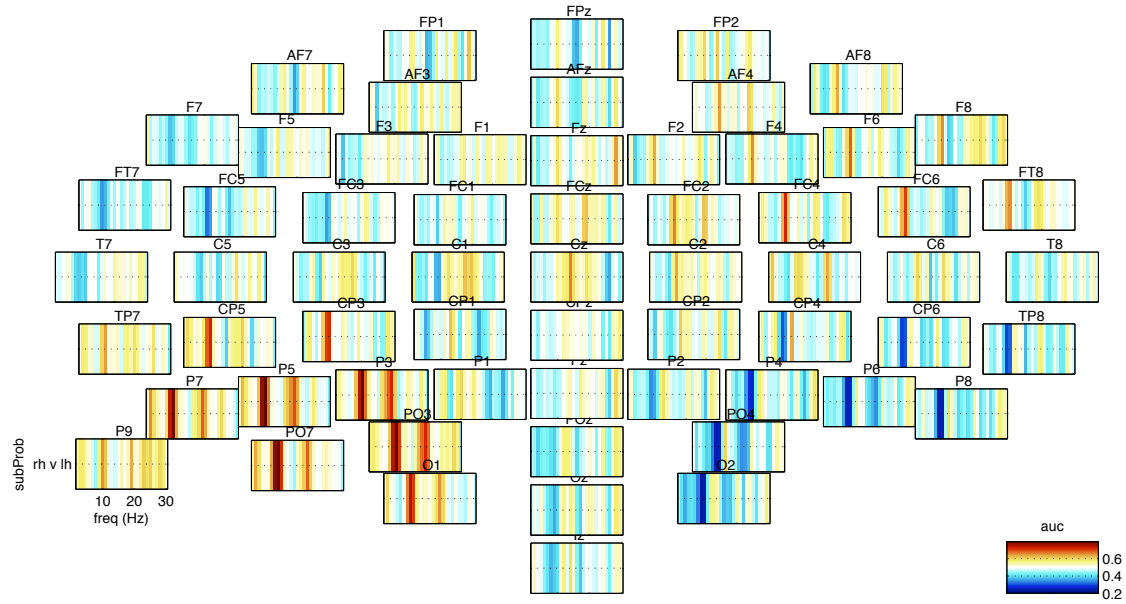


Figure 3.3: AUC plot showing the AUCs for frequencies between 2 and 30 hz averaged over the epochs of participant 2 in session 1 during the epoch feedback blocks. The channels are shown in a topographical view. This plot shows high AUC values around P5 and P6. Furthermore, the frequency bands with discriminating information seem to be smaller compared to Figure 3.2.

offline analyses, in which classifiers are trained on data other than the training blocks, are discussed.

3.2.1 Online

To give an impression of the classification performance of the participants at the start of a session, Table 3.1 shows the cross validated estimated classification performance for the classifiers used during the experiment (online). These classifiers were trained every session on the two imagined movement training blocks (see Figure 2.2) with 144 epochs in total. An epoch classifier was trained for epoch feedback, and a continuous classifier for game feedback. Distinct classifiers were trained, because the game feedback needed to give feedback as rapidly as possible and therefore a smaller data window of 0.6 s was used. The epoch classifier was trained on data windows of 3.6 s and the continuous classifier on data windows of 0.6 s, because training and testing on data of the same length improved the online performance.

The classification accuracy differed from participant to participant, as would be expected. For example, participant 1 performed very well (99%), while participant 6 only performed around chance level (52%). The classification accuracy in the second session was higher than in the first session for both classifiers as can be shown by comparing the classification performance between sessions with a paired t-test ($p < 0.05$ for the epoch classifier, and $p = 0.06$ for the continuous classifier). The epoch classifier performed better in the first session

classifier	session	S1	S2	S3	S4	S5	S6	S7	S8
epoch	1	99(1)*	73(4)*	62(4)*	62(3)*	74(4)*	53(3)	67(4)*	77(3)*
epoch	2	100(0)*	69(4)*	60(5)	58(3)	68(5)*	52(2)	51(4)	68(5)*
continuous	1	93(2)*	68(2)*	56(1)*	55(2)	58(1)*	51(2)	54(3)	57(1)*
continuous	2	90(1)*	58(2)*	54(2)	53(1)	58(2)*	52(1)	51(2)	57(2)*

Table 3.1: Cross validated estimated classification accuracy and standard error rates (in %) for each session, using data from the training blocks (the first two imagined movement blocks of the experiment). Results are shown for both classifier types, i.e. epoch (3.6 s) and continuous (0.6 s). * indicates that the performance was significantly ($p < 0.01$) above chance (50%).

block	session	S1	S2	S3	S4	S5	S6	S7	S8	Avg
epoch	1	94*	84*	53	53	74*	53	55	56	65
epoch	2	96*	61	70*	50	57	49	47	62*	62
game	1	89*	49	50	48	50	53	53	53	56
game	2	87*	55	51	58	62*	62*	50	51	59
offline	1	97*	78*	60	50	65*	55	54	63*	64
offline	2	92*	63*	69*	51	55	55	51	64*	63

Table 3.2: Classification accuracy (in %) of the epoch classifier on the test data sets from the same session, i.e. epoch (2 epoch blocks), game (2 game blocks) and offline (2 IM blocks after feedback) data set. * indicates that the performance was significantly ($p < 0.01$) above chance (50%) calculated with the averaged standard error from the epoch classifier over all training blocks and sessions.

(mean 71%) compared to the second session (mean 66%), and the game classifier performed better in the first session as well (mean 62% versus mean 59%).

The lower classification performance of the continuous classifier in comparison with the epoch classifier can be explained by the shorter data slices that were used during training and testing (3.6 s versus 0.6 s). During the game feedback, multiple classification results were combined and used as feedback in one epoch (see Section 2.1.4) and then the game classifier should give about the same classification performance as the epoch classifier on 3.6 seconds of data.

To demonstrate the performance of the participants during the experiment, Table 3.2 and Figure 3.4 show the classification performance of the epoch classifier, trained at the beginning of the session, tested on the other blocks of the session, where epochs of the same type, i.e. both epoch feedback blocks, both game feedback blocks, and the two imagined movement blocks recorded after each feedback block, have been merged together. The game classifier should give about the same results as the epoch classifier, and therefore we only report the epoch classifier results here.

To compare the performance of the different feedback types (epoch, game and no feedback), the classification performance was analyzed using an analysis of variance (ANOVA) with within-subject factor feedback (Epoch, Game, Offline) with two repeated measures. No significant trends were found. However, by eye, the game feedback seems to perform worse

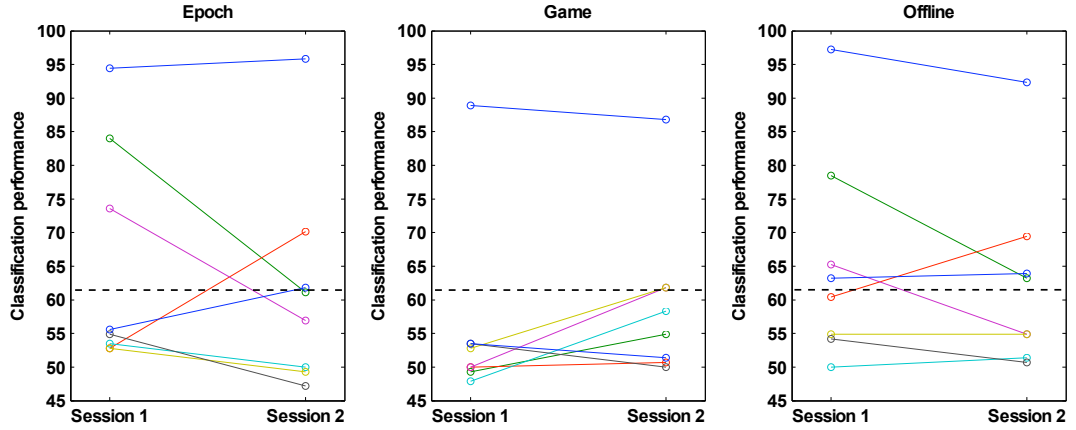


Figure 3.4: Classification accuracy of the epoch classifier tested on the epoch, game, and offline data set from the same session. Each line represents a participant. The horizontal dashed line is the significance threshold determined by the averaged standard error from the epoch classifier over all training blocks and sessions.

than the other two methods (59% compared to 63% for the epoch feedback and 63% for no feedback). The lower classification performance would indicate a change in the features of the game data compared to the training data.

3.2.2 Offline

To check whether training a new classifier on the later acquired data increases the classification performance compared to testing the classifier trained on the training blocks on this data, separate classifiers were trained on the epoch, game and offline (two IM blocks after the feedback blocks) blocks for each session. Classification performance, again computed by ten fold cross validation, are shown in Table 3.3 and Figure 3.5. Again, an ANOVA analysis comparing the different feedback types showed no significant results, indicating that in general, the participants performed the same for all three blocks, meaning that the brain signals could be discriminated equally well in all different feedback blocks. By eye, the different feedback blocks also seem to perform the same.

Figure 3.6 shows a comparison between the classification performance of the classifier trained on the training blocks (online) and the cross validated performance of the retrained classifier (offline) on the same blocks. To compare the performance, the classification rates were analyzed with an ANOVA with repeated measures with within-subject factors (Epoch, Game, Offline) and classifier type (Online, Offline). A main effect of classifier type was found ($F(7,1) = 13.130, p < 0.01$). The online classification accuracy was lower (mean 61%) than the offline classification accuracy (mean 67 %). The higher classification performance in the offline classifier can be explained by a change in the signals; the online classifier was not trained on the new data and was therefore not able to adapt to possible features changes occurring in the new data set, in contrast to the offline classifier that used the new data as training set. No other significant effects were found.

To check whether the classification performance changed over time within a session,

block	session	S1	S2	S3	S4	S5	S6	S7	S8	Avg
epoch	1	95*	69*	61	62*	66*	46	73*	75*	68
epoch	2	98*	90*	64*	60	70*	51	60	66*	70
game	1	100*	69*	60	58	68*	52	51	68*	66
game	2	97*	68*	69*	55	59	52	58	60	65
offline	1	93*	72*	54	63	62*	52	50	71*	65
offline	2	98*	70*	71*	61	67*	52	58	62*	67

Table 3.3: Cross validated estimated classification accuracy (in %) of classifiers retrained on epoch, game and offline blocks for each session. * indicates that the performance was significantly ($p < 0.01$) above chance (50%) calculated with the averaged standard error from the epoch classifier over all training blocks and sessions.

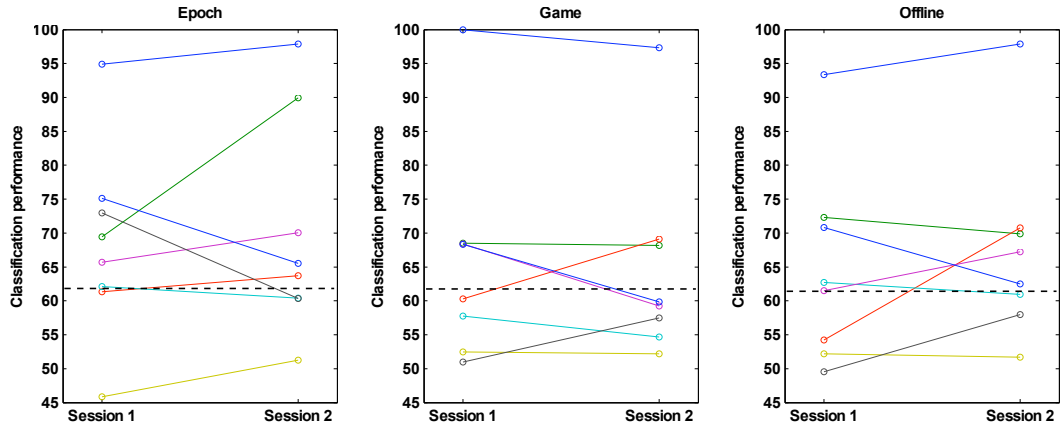


Figure 3.5: Cross validated classification accuracy of classifiers retrained on epoch, game and offline blocks for each session. Each line represents a participant. The horizontal dashed line is the significance threshold determined by the averaged standard error from the epoch classifier over all training blocks and sessions.

the performance of the separate blocks without feedback were assessed. To compare across time, we generated new classifiers, based on 72 epochs each. (Note; these classifiers are less reliable than those used in the previous analyses, because fewer epochs are used for training.) We also took the actual movement block into account to see whether the actual movement resembles the imagined movement, as we would expect if participants perform kinesthetic imagined movement. The results are shown in Figure 3.7. This figure shows large differences between the blocks for participants. Furthermore, the cross-validated estimated classification rates were analyzed with an ANOVA with within-subject factor block (AM,IM1,IM2,IM3,IM4) with two repeated measures to check whether there were any significant changes during a session. No significant effects were found, meaning that the classification accuracy is about the same for each block when a new classifier was trained, which implies that participants' brain signals are equally discriminative through the entire session. This does not mean, that a classifier trained at blocks at the beginning of the session keeps

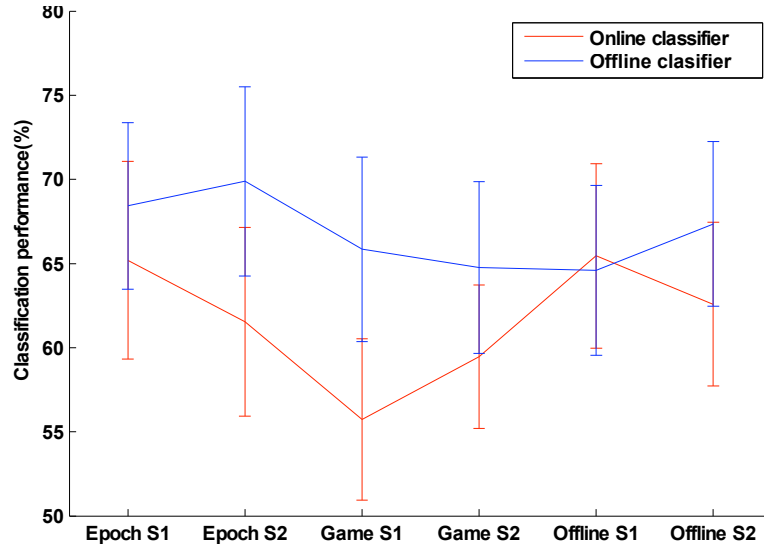


Figure 3.6: Average over all participants of the classification performance of blocks tested with the epoch classifier trained on the training blocks (online) and cross validated estimated classification accuracy of the retrained classifier (offline). Averages and standard errors are shown for each feedback type in both sessions.

functioning during the entire session (the comparison between the online and offline classifiers, already showed there were differences). Finally, the actual movement seems to reflect the imagined movement classification performance.

To conclude, these results show a large degree of user variability in the BCI performance. However, no significant differences have been found between feedback types online and only on the combined training block a significant difference was found between sessions. Furthermore, retraining on new data improved the classification performance compared to using the classifier trained on the training blocks, which indicates that the signals changed during the session, as explained above. Retrained classifiers showed no differences between feedback types, or between the individual imagined movement blocks.

3.3 Bias and Gain Analysis

In this section, we apply bias and gain adaptations to the classification results to test whether the classification performance and the feedback to the user can be improved by these methods, as was shown in previous research (see Section 1.3.1). We tested the adaptations on the epoch and game feedback blocks using the epoch classifier trained on the training blocks.

3.3.1 Bias

Bias adaptation tries to compensate for shifts occurring in the classifier predictions. Figure 3.8 shows examples of the problem we would like to solve with bias adaptation. In the first

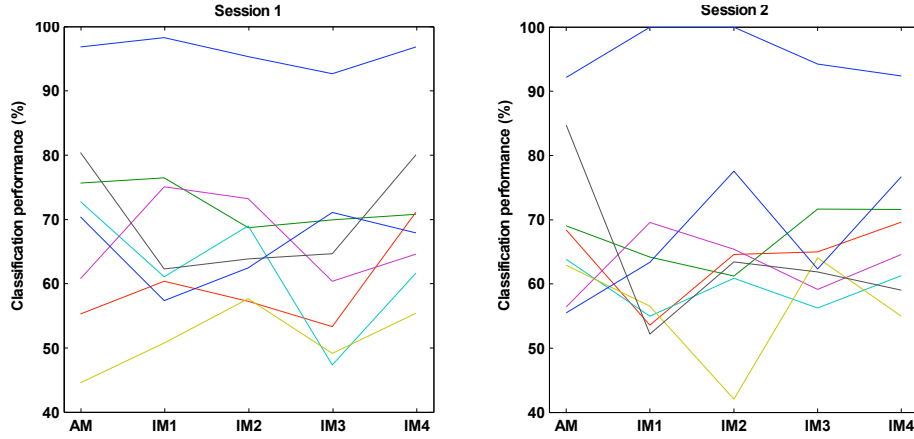


Figure 3.7: Cross validated estimated classification accuracy (in %) of the actual movement block and the individual imagined movement blocks for each session. Each line represents a participant. Error bars are omitted for clarity.

plot more decision values are less than zero, indicating that more epochs were classified as right hand movements than as left hand movements. Visually it appears that this is largely due to a shift in the classifier predictions. We could easily undo this using a bias correction. In the second plot the decision values shift only in the last epochs, also this can be solved with an adaptive bias correction. Therefore, we tried adding some simple bias adaptations on the recorded data to remove shifts, namely:

- Subtracting the average of the last 20 decision values
- Subtracting the average of the last 20 decision values with outlier detection. Where decision values deviating more than 2 standard deviations from the mean of the 20 decision values were removed.

We decided to use unsupervised adaptation (not using class information), because in real use a BCI cannot know what the users true choice was.

Figure 3.9 shows the performance of the two bias methods on the four different feedback blocks. There seems to be a lot of variability between participants and blocks with respect to the adaptation performance. In some case the bias adaptation improves, but in others it degrades the performance. Somewhat surprising however is that the bias adaptation performs most consistently on the second epoch block compared to the other blocks. On average the classification rate before bias adaptation was 60.5%, with bias adaptation (but without outlier detection) 61.0%, and with outlier detection 61.3%. The bias adaptation has improved the performance in our data set a little and the method using outlier rejection appears to perform best.

Although, the bias adaptation did not have a large impact on the classification performance, the adaptation did improve the distribution of the decision values over the negative and positive class. For example, in the game feedback of participant 3 in the first session, 71% of the decision values were negative and 29% positive, but after bias adaptation with outlier detection this was more evenly divided, namely 49% and 51%. This evenly divided

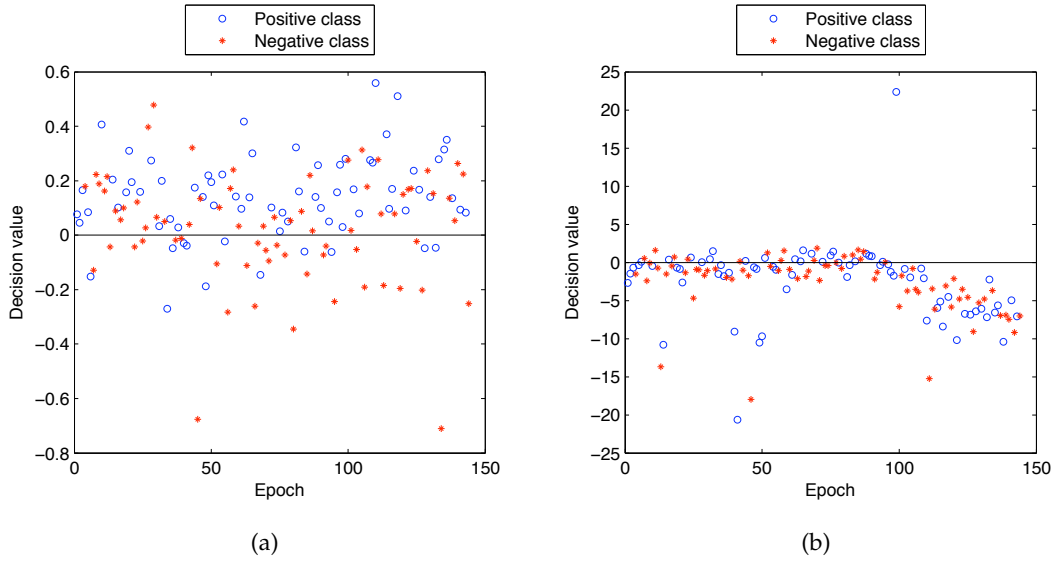


Figure 3.8: Examples of classifier outputs in epoch feedback, showing a shift(a) and a drift(b) in the classifier outputs. Figure a contains the decision values of the epoch block of participant 8 in the second session and figure b contains the decision values of the epoch block of participant 4 in the first session.

distribution makes sure that the ball keeps falling to both sides of the screen, and this may improve the feel of controllability for the user.

3.3.2 Gain

In the game feedback blocks, gain adaptation (see also Section 2.1.4) was used to evenly distribute the position of the ball on the screen. This was important, because often the decision values were lower during the feedback blocks than during the training, which also reduced the standard deviation and resulted in only small, almost imperceptible, ball movements on the screen.

To demonstrate the use of gain adaptation, Figure 3.10 shows gain values for two different participants. The gain values were calculated as one divided by the standard deviation of a window of 20 decision values. Two gain adaption methods, with or without outlier removal, are shown. The outlier procedure used is similar to the procedure applied in the bias adaptation, namely epochs were removed when the decision value deviated more than 2 standard deviations from the mean decision value. The right plot shows very low decision values, and therefore a relatively high gain. There were also changes in the distribution during the session, to which the gain adapted. The highest performing participant (participant one) had almost no distribution changes in the decision values compared to the training session and so almost no gain adaptation was necessary. So, the gain adaptation seems to be more useful for participants with a low classification accuracy than for participants with a high accuracy. The gain adaptation with outlier detection seems to give smoother results than the other methods.

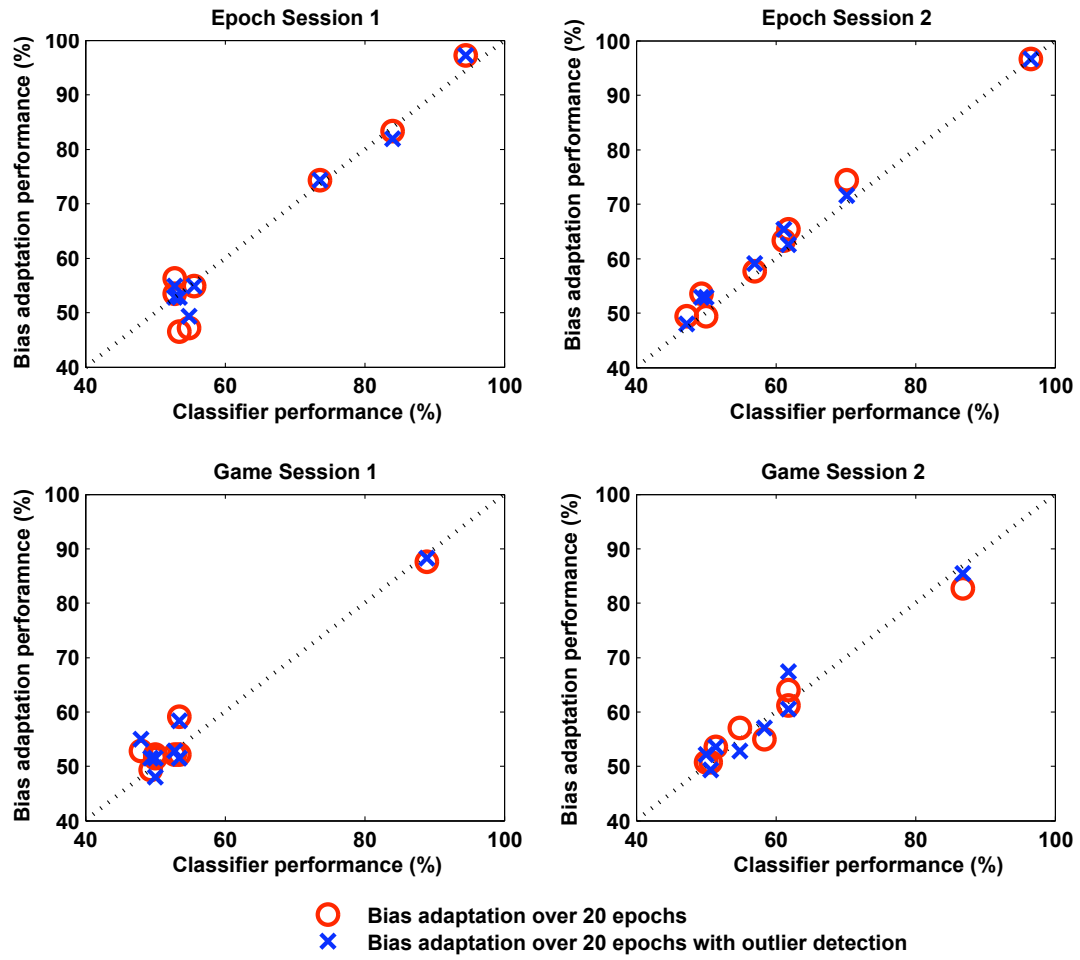


Figure 3.9: Classification accuracy of both bias adaptation methods compared to the classification accuracy without adaptation. Every point represents a participant. The performance of the bias adaptation is better for those data points that lie above the diagonal.

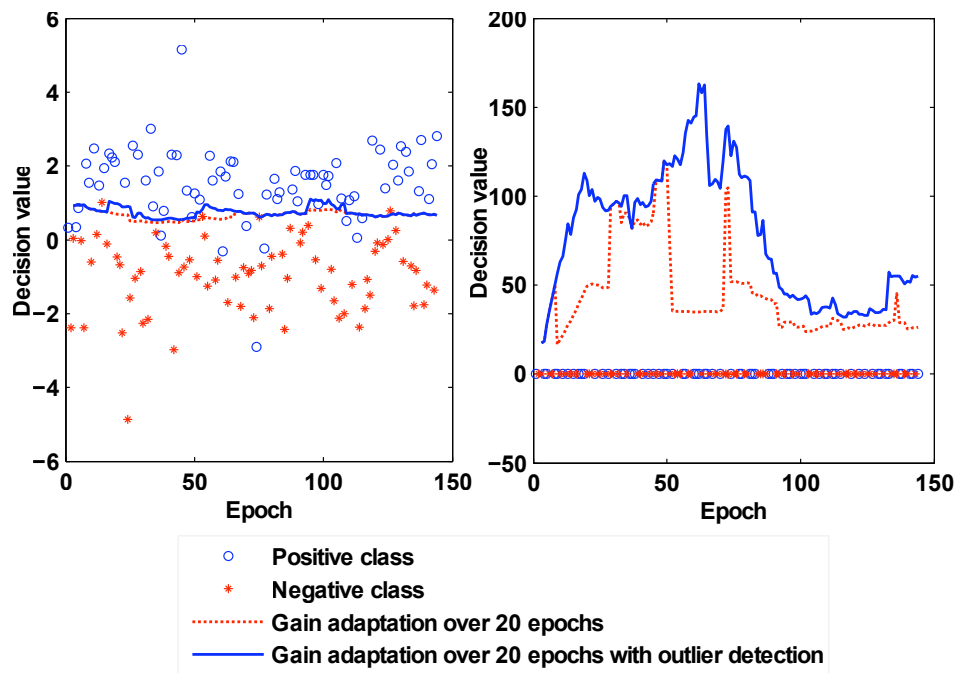


Figure 3.10: Gain adaptation calculated as one divided by the standard deviation using a windows size of 20 epochs, with or without outlier detection. The left plot shows gain adaptation in the game feedback block from the first session of participant 1 and the right plot shows gain adaptation in the game feedback block from the second session of participant 5.

Chapter 4

Discussion and conclusions

In the reported experiment, we explored the variability and nonstationarities in EEG data and their importance for BCI. Different types of variability in the data were indeed observed, namely variability between users, between sessions and within sessions. Let us discuss each type of variability in more detail.

4.1 Inter-user variability

Classification accuracy between participants was very variable, as would be expected from previous research. The classification performance varied between 52 percent and 100 percent during the training blocks. Conspicuous is the relatively low classification accuracy compared to other imagined movement studies. Our classification performance do resemble the classification performance of Guger et al. (2003), but in a more recent study (Blankertz et al., 2008) was found that in a group of 14 fully BCI-naive participants, 8 out of 14 participants could perform at more than 84 percent accuracy in their first BCI session, and a further four participants at more than 70 percent. In our case, only one out of eight participants performed better than 84 percent, and all the others had a performance lower than 70 percent. One of the participants seems to be illiterate for the imagined movement signals. We have no convincing explanation for this low performance yet. Possible explanations could be related to the design of the experiment: The task, drumming in a fast rate, could be too difficult for the participant, although no problems were seen in the pilot and one experimental participant was perfect at it. Alternatively, participants may have used the wrong kind of imagery. Reasons could also be technical, e.g., the EEG equipment, type of feature extraction and classification, classification window onset/length. However, for generalization purposes the data need to be analyzed further and more users have to be measured. Hopefully, this will lead to some clarifications.

4.2 Inter-session and user learning variability

Each participant took part in two sessions, allowing us to examine the inter-session and user learning variability. With respect to the classification accuracy, participants performed differently from day to day. One day, they performed very well, and the second day the performance could be better or worse. The degree of change varied between 1 and 17 percent.

Remarkably, the classifier trained on the training blocks seemed to perform worse in the second session for almost every participant, but there was no performance difference between sessions in the other blocks (epoch, game and offline). The lower classification accuracy in the second session is contrary to our expectations. In general, if participants perform the task repeatedly and receive feedback, their performance should improve (Neuper et al., 1999; Vidaurre et al., 2006). On the other hand, we have only measured during two sessions for each participant; thus, the learning effect may have been masked by the large inter-session variability.

4.3 Feedback dependent variability

In the experiment, we used two types of feedback to investigate the variability between feedback types, epoch and game feedback. We did not find any significant differences in the classification performance between feedback types and "probe" blocks. However, the average classification performance indicated that participants had worse classification accuracy in the game feedback. Most likely, we measured too few participants to obtain a significant effect. The lower classification accuracy in the game feedback resembles the findings of Shenoy et al. (2006) that the feature distribution changed when going from an offline to a continuous feedback BCI.

4.4 Signal nonstationarity

Nonstationarity during a session was found in our experiment. Cross-validated estimated test results, from a classifier trained on the feedback blocks, showed higher classification accuracy than when testing the classifier trained on the training blocks on that same data. The classifiers trained on the feedback blocks took changes in the data into account, whereas the classifier from the training blocks did not. Apparently, between the blocks the signals changed in such a way that the original classifier was not able to reach the same accuracy as retrained classifiers. This resembles the findings of Shenoy et al. (2006) that the signal changed during the session. Furthermore, we did not find any significant changes in performance of the participant throughout the long (more than 2 hours) experiment when we retrained the classifiers, suggesting that participants were performing the task during the entire experiment in such a way that a classifier should be able to differentiate between the classes, although there are inter-block variabilities.

The bias adaptation we used improved the classification results a little: 0.8% on average with outlier detection. However, in some of the participants the bias adaptation worsened the results. Other studies using bias adaptation had larger effects and seemed to improve in almost all cases (Tsui & Gan, 2008; Vidaurre et al., 2008). Tsui and Gan (2008) used 3 different adaptation methods and all naïve participants improved, and on average there were improvements of 2.36%, 4.44% and 3.47%. In our research however, we used different features and a different classification algorithm, which could explain the lower classification improvement.

Chapter 5

Future work

The research presented in this thesis represents some initial steps towards clarifying the variability and nonstationarity present in Brain Computer Interfaces. Although due to time restrictions, only preliminary analyses were performed on the available data, many important issues could be investigated. Besides analyzing the changes in classification performance, the source EEG signals should be investigated. A possible analysis to consider is a space-time frequency analysis, to determine where and in which frequency bands at what times the BCI relevant signals are, and how they are changing between conditions. Furthermore, the lateralized readiness potential could be considered to be used as feature for classification, although the metronome may have made this signal invisible in the data. In addition, varying preprocessing and classification parameters may improve the classification performance. The results of the questionnaire should also be analyzed, to assess the relation of the participant's subjective reports to their BCI performance.

Furthermore, the use of feedback in the experiment's BCI could be improved, for instance, by using more complex adaptation methods, like Kalman filtering. In addition, it needs to be investigated whether changes in electrode impedances with the skull have an impact on the BCI performance during the experiment. Next, the implementation of a more complex bad channel detector might be feasible.

Finally, another interesting issue concerns the effect of user learning. Unfortunately, we were only able to administer two sessions for each participant during this experiment, which did not give us much information about the learning process. Extending the experiment to more sessions and more participants would provide us with more detailed information, also allowing more detailed investigation of the other types of variabilities.

Bibliography

- Allison, B. Z., McFarland, D. J., Schalk, G., Zheng, S. D., Jackson, M. M., & Wolpaw, J. R. (2008). Towards an independent brain-computer interface using steady state visual evoked potentials. *Clinical Neurophysiology*, 119, 339-408.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic. *Journal of Mathematical Psychology*, 12, 387-415.
- Bartolic, E. I., Basso, M. R., Schefft, B. K., Glauser, T., & Titanic-Schefft, M. (1999). Effects of experimentally-induced emotional states on frontal lobe cognitive task performance. *Neuropsychologia*, 37, 677-683.
- Bensch, M., Karim, A. A., Mellinger, J., Hinterberger, T., Tangermann, M., Bogdan, M., et al. (2007). Nessi: An EEG-controlled web browser for severely paralyzed patients. *Computational Intelligence and Neuroscience*, 2007, 5.
- Birbaumer, N. (2006). Breaking the silence: Brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43, 517-532.
- Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., et al. (1999). A spelling device for the paralysed. *Nature*, 398, 297-298.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., & Curio, G. (2007). The non-invasive Berlin Brain-Computer Interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37, 539-550.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Kunzmann, V., Losch, F., et al. (2006). The Berlin Brain-Computer Interface: EEG-based communication without subject training. *IEEE Transactions on Neural Systems and Rehabilitation*, 14, 147-152.
- Blankertz, B., Dornhege, G., Lemm, S., Krauledat, M., Curio, G., & Müller, K.-R. (2006). The Berlin Brain-Computer Interface: machine learning based detection of user specific brain states. *Journal of Universal Computer Science*, 12, 581-607.
- Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., & Müller, K.-R. (2008). The Berlin Brain-Computer Interface: accurate performance from first-session in bci-naïve subjects. *IEEE Transactions on Biomedical Engineering*, 55, 2452-2462.
- Buttfield, A., Ferrez, P. W., & R. Millán, J. del. (2006). Towards a robust BCI: Error potentials and online learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14, 164-168.
- Curran, E. A., & 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.
- Debarnot, U., Creveaux, T., Collet, C., Gemignani, A., Massarelli, R., Doyon, J., et al. (2009). Sleep-related improvements in motor learning following mental practice. *Brain and Cognition*, 69, 398-405.
- Dornhege, G., Braun, M., Kohlmorgen, J., Blankertz, B., Müller, K.-R., Curio, G., et al. (2006).

- Improving human performance in a real operating environment through real-time mental workload detection. In *Towards brain-computer interfacing*. MIT Press.
- Farquhar, J. (2009). *A linear feature space for simultaneous learning of spatio-spectral filters in BCI*. (Neural Networks (2009))
- Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70, 510-523.
- Gerven, M. van, Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., et al. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4), 1-9.
- Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., & Pfurtscheller, G. (2003). How many people are able to operate an EEG-based brain-computer interface (BCI)? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11, 145-147.
- Heremans, E., Helsen, W., Poel, H. D., Alaerts, K., Meyns, P., & Feys, P. (2009). *Facilitation of motor imagery through movement-related cueing*. (doi:10.1016/j.brainres.2009.04.041)
- Klem, G. H., Luders, H. O., Jasper, H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement*, 52, 3-6.
- Kranczioch, C., Mathews, S., Dean, P. J. A., & Sterr, A. (2009). On the equivalence of executed and imagined movements: evidence from lateralized motor and nonmotor potentials. *Human Brain Mapping*, 30, 3275-3286.
- Krauledat, M. (2008). *Analysis of nonstationarities in EEG signals for improving brain-computer interface performance*. Unpublished doctoral dissertation, Technische Universität Berlin, Fakultät IV – Elektrotechnik und Informatik.
- Krauledat, M., Dornhege, G., B. Blankertz, Losch, F., Curio, G., & Müller, K.-R. (2004). Improving speed and accuracy of brain-computer interfaces using readiness potential features. In *Conference proceedings IEEE engineering in medicine and biology society*.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.
- Mason, S. G., Bashashati, A., Fatourechi, M., Navarro, K. F., & Birch, G. E. (2007). A comprehensive survey of brain interface technology designs. *Annals of Biomedical Engineering*, 35, 137-169.
- McFarland, D. J., Miner, L. A., Vaughan, T. M., & Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain topography*, 12(3), 177-186.
- Millán, J. R., Renkens, F., Mourino, J., & Gerstner, W. (2004). Non-invasive brain-actuated control of a mobile robot by human EEG. *IEEE Transactions on Biomed Engineering*, 51, 1026-1033.
- Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G., & Blankertz, B. (2004). Machine learning techniques for brain-computer interfaces. *Biomedical Technology*, 49, 11-22.
- Nangini, C., Ross, B., Tam, F., & Graham, S. (2006). Magnetoencephalographic study of vibrotactile evoked transient and steady-state responses in human. *NeuroImage*, 33, 252-262.
- Neuper, C., Scherer, R., Reiner, M., & Pfurtscheller, G. (2005). Imagery of motor actions: Differential effects of kinesthetic and visual - motor mode of imagery in single-trial EEG-based. *Cognitive Brain Research*, 25, 668-677.
- Neuper, C., Schlögl, A., & Pfurtscheller, G. (1999). Enhancement of left-right sensorimotor

- EEG differences during feedback-regulated motor imagery. *Journal of Clinical Neurophysiology*, 16, Journal of Clinical Neurophysiology.
- Nijholt, A. (2009). BCI for games: A 'state of the art' survey. In *Entertainment Computing - ICEC 2008* (p. 225-228). Springer Berlin / Heidelberg.
- Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C., R. Millán, J. del, Allison, B., et al. (2008). Brain-computer interfacing for intelligent systems. *IEEE Intelligent Systems*, 23(3), 72–79.
- Pfurtscheller, G., Brunner, C., Schlögl, A., & Silva, F. L. da. (2006). Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery task. *NeuroImage*, 31, 153-159.
- Pfurtscheller, G., & Silva, F. da. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, 110(11), 1842-1857.
- Regan, D. (1977). Steady-state evoked potentials. *Journal of Optical Society America*, 11, 1475-1489.
- Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement imagery ability: development and assessment of a revised version of the vividness of movement imagery questionnaire. *Journal of Sport & Exercise Psychology*, 30, 200-221.
- Ross, B., Herdmann, A., & Pantev, C. (2005). The effect of attention on the auditory steady-state response. *Neurology and Clinical Neurophysiology*, 22, 1-4.
- Schalk, G., Wolpa, J. R., McFarland, D. J., & Pfurtscheller, G. (2000). EEG-based communication: presence of an error potential. *Clinical Neurophysiology*, 111, 2138-2144.
- Shenoy, P., Krauledat, M., Blankertz, B., Rajesh, Rao, P. N., & Müller, K.-R. (2006). Towards adaptive classification for BCI. *Journal of Neural Engineering*, 3, R13-R23.
- Tomioka, R., Aihara, K., & Müller, K.-R. (2007). Logistic regression for single trial EEG classification. In *Advances in neural information processing systems 19* (p. 1377-1384). MIT Press.
- Tsui, C. S. L., & Gan, J. Q. (2008). Comparison of three methods for adapting LDA classifiers with BCI applications. In *Proceedings of the 4th international brain-computer interface workshop and training course 2008* (p. 116-121).
- Vidaurre, C., & Blankertz, B. (2009). Towards a cure for BCI illiteracy. In *7th NFSI & ICBEM 2009*.
- Vidaurre, C., Schlögl, A., Blankertz, B., Kawanabe, M., & Müller, K.-R. (2008). Unsupervised adaption of the LDA classifier for brain-computer interfaces. In *Proceedings of the 4th international brain-computer interface workshop and training course 2008* (p. 122-127).
- Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2006). A fully on-line adaptive BCI. *IEEE Transactions on Biomedical Engineering*, 53, 1214-1219.
- Wang, Y., Zhang, Z., Li, Y., Gao, X., Gao, S., & Yang, F. (2004). BCI competition 2003-data set iv: an algorithm based on CSSD and FDA for classifying single-trial EEG. *IEEE Transactions on Biomedical Engineering*, 51(6), 1081-1086.
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15, 70-73.
- Wolpaw, & Jonathan, R. (2007). Brain-computer interfaces as new brain output pathways. *The Journal of Physiology*, 579(3), 613–619.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767-791.

Appendix A

Experiment

A.1 Instruction form

Thank you for taking part in this experiment. Please read these instructions carefully.

In this experiment you have to drum with your hand or imagine drumming at a tempo set by a metronome. When you imagine a movement, please try to imagine it as vividly as possible. Try to imagine the feeling of the movement as if you were actually performing the movement, but do not move your hands (or any other part of your body) while imagining.

This experiment consists of 9 blocks of about 8-10 minutes each. Between the blocks you can rest for several minutes if you want to. There are two types of blocks and the experimenter will tell you before each block which type it will be.

The first type consists of the following sequence, for either actual movement or imagery:

1. A black screen: you do not have to do anything.
2. An arrow: focus your eyes on the screen; the direction of the arrow indicates whether you have to move/imagine moving your left or right hand. The metronome will start ticking. Do not start (imagining) moving yet.
3. A fixation cross: focus your eyes on this cross, try not to blink your eyes, try to sit still, and start (imagining) moving your left/right hand in the tempo of the metronome. You should perform the (imagined) movement each tick. Your last drum-movement has to happen after the last tick.

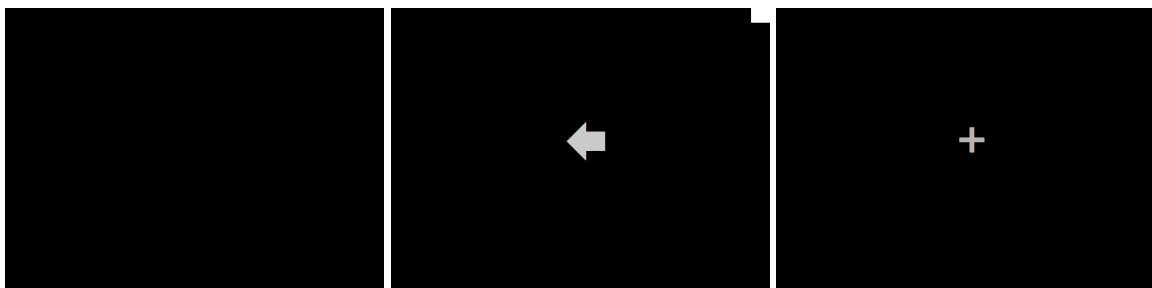


Figure A.1: Left hand movement sequence

In the second type of block you are going to play a game by imaging left/right hand movements. In the middle of the screen you see a ball that will fall down. You can control the horizontal position of the ball by imagining movements.

In the beginning, the ball appears in the middle of the screen and the metronome starts ticking. When the ball starts falling down you have to start imagining a left or right hand movement at the tempo of the metronome. If you imagine a left hand movement, the ball will move to the left, and if you imagine a right hand movement the ball will move to the right. Your goal is to let the ball drop in the basket with the highest number of points. The baskets that will give you points have a green color. The more the ball is to the side of screen, the more points you will receive.

When you drop the ball in one of the correct baskets multiple times in a row, you will receive a score multiplier. For example, if you have correctly moved the ball to a green basket 4 times in a row, the next ball in a green basket will give you twice as much points. When you put the ball in a basket without points, the multiplier will reset.

Do not worry if the ball does not move to the direction which you are imagining movement. In the first blocks the computer collects information from your brain signals to predict which signals indicate a left or a right hand movement. The computer may not be able to tell every time which hand movement you are imagining.

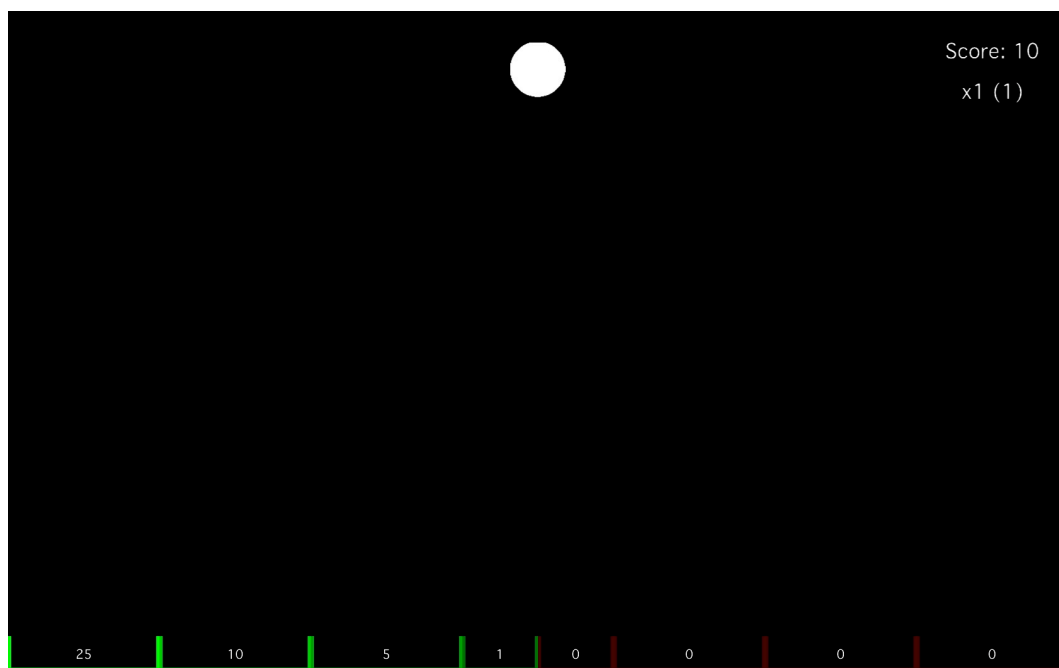


Figure A.2: Game

During the experiment you have to put your hands on the little black blocks that are on the table. During actual movement you have to perform the movement on the blocks, and while imagining you rest your hands on the blocks.

If you have any questions, please ask them. Good luck!

A.2 General questionnaire

Subject ID: S

General

Age:

Gender: male / female

Course/profession:

Are you left- or right handed? left handed / right handed

Visual aids: contact lenses / glasses / none

Medication:

Attention/neurological/psychiatric disorders:

Have you ever participated in an imagined movement BCI experiment before? yes / no

Are you a morning or an evening person? morning o o o o o evening

Session 1

How many hours of sleep did you get last night?

What time did you wake up today?

What do you think of the game? no fun at all o o o o o much fun

Other remarks:

Session 2

How many hours of sleep did you get last night?

What time did you wake up today?

What do you think of the game? no fun at all o o o o o much fun

Do you have the feeling that your task performance has improved?
worsened performance o o o o o improved performance

Other remarks:

A.3 Imagined movement scale

After imagined movement blocks, participants had to scale the vividness of imagination according to the kinesthetic imagery scale from the revised Vividness of Movement Imagery Questionnaire (Roberts, Callow, Hardy, Markland, & Bringer, 2008). The following scale was used:

1. No image at all, you "know" that you are thinking of the skill.
2. Vague and dim
3. Moderately clear and vivid
4. Clear and reasonably vivid
5. Perfectly clear and as vivid (as normal feeling of movement)