Bauhaus-Universität Weimar
Fakultät Medien
Studiengang Mediensysteme

Comparison of P300 spelling and input imagery for text selection on a consumer BCI

Task performance, workload and user acceptance

Bachelor thesis

Katja Müller | Matrikelnummer 70194
Geboren am 22. August 1987 in Wittenberg

1. Gutachter: PD Dr.-Ing. habil. Günther Schatter
2. Gutachter: Jun.-Prof. Dr. Sven Bertel

Datum der Abgabe: 20. April 2012

Erklärung

Hiermit versichere ich, daß ich diese Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Weimar, den 20. April 2012

. . . .

Katja Müller

Acknowledgments This thesis was written at the Faculty of Media, Bauhaus-University Weimar at the Chair for "Vernetzte Medien". I would like to thank Dr.-Ing. habil. Günther Schatter for his guidance and for the opportunity to pursue a thesis in the area of Brain-Computer Interaction (BCI); Jun.-Prof. Dr. Sven Bertel for being my co-advisor and for advice on statistical methodology; Alexander Kulik for advice on formal evaluation methodology, and the group for Virtual Reality Systems and Prof. Fröhlich for allowing me to conduct the evaluation in their laboratory. I would also like to thank Dr. Jürgen Mellinger of the BCI2000 team and Christian Walther of the University of Jena for advice in the area of Brain-Computer Interaction and EEG experimentation, and Martin Trenkmann for helping in setting up a local NetSpeak system at the beginning of the thesis work. As much I would like to thank all participants who took part in the evaluation.

Abstract

With the consideration to use it in rehabilitation centres, the text entry capabilities of the Emotiv EPOC as a recent and affordable consumer BCI system were investigated. P300 spelling and switch-scanning (with a single input imagery using the device's software classifiers) were evaluated as possible BCI variants. An additional feedback variant of input imagery was included in the evaluation. The experimental set-up mimicked the text entry process by alternating text formulation and word selection in a 6x6 matrix. A sample of 30 healthy participants evaluated both systems within-subjects in a single session with short training times. Data for task performance, workload (NASA Task Load Index (TLX)), user acceptance (System Usability Scale (SUS)), error types (self-reported) and system preference was collected. During P300 spelling, an average information transfer rate of 1.28 bit per trial and a Task Completion Rate (TCR) of 79.6% could be achieved in our experimental set-up, with the best users reaching 10.25 bit/min and 3.67 bit per trial. TLX was rated 51.92 on average, with an emphasis on physical demand and effort. SUS was rated 23.27 on average. Subjects were more likely to attribute errors occurring during the evaluation to the system. During the no-feedback switch-scanning system, the average bit rate was 0.01 bit per trial (max. 0.39 bit per trial) and the TCR 28.6%, while in the feedback system an average of 0.34 bit per trial and a TCR of 58.6% was reached (max. 1.69 bit per trial). 5 subjects (17%) were unable to create an input imagination sufficiently distinctive for input (BCI illiteracy). TLX of the no-feedback system was rated 70.06 on average, and 49.39 for the feedback system, both with an emphasis on mental demand, effort and frustration. SUS was rated 17.87 for the no-feedback system and 25 for the feedback system. The subjects were likely to attribute errors occurring during the evaluation to themselves. False positive input occurred during the text formulation subtasks (Midas touch problem) and could present a prominent usability issue in a full system. Despite the low task performance of the switch-scanning system, both systems were equally preferred by users for various reasons[1]. However, before software keyboards based on the classifiers in the Cognitive suite of the device can be used in actual writing systems, usability problems such as false positive input need to be addressed. The current P300 input paradigm should therefore be investigated with enhanced writing interfaces on this device, while the input imagery paradigm should be evaluated with prolonged training time and addressed usability and input acquisition problems due to existing user acceptance.

[1]Reasons for system preference are presented in appendix C.1 .

Contents

Chapter 1

Introduction

The locked-in condition, where individuals became paralysed due to a stroke or neuromuscular diseases has left their physicians and caretakers helpless for decades. While patients with remaining muscle functions have been provided with switch-scanning systems for long time, only recently systematic research into communication technologies for the totally locked-in condition began. For the growing number of patients in this condition, the only remaining form of communication and the only means to regain autonomy in the long term will be Brain-Computer Interaction (BCI) technology enabling them to control text entry systems and computers. A working and usable system for communication would dramatically increase life quality for thousands of patients currently cut off from their surroundings.

1.1 Problem definition

The recently released consumer Electroencephalographys (EEGs) systems could prove to be an affordable and easy-to-use solution as BCI communication systems for rehabilitation centres. The Emotiv EPOC that will be used in this thesis was the first consumer EEG with a considerable number of electrodes and sufficient signal quality for specific BCI input tasks. The device is substantially cheaper than comparable specialist devices. It is yet unclear how fast and reliable the possible BCI input variants work on this device, and which of them are more suitable as text entry systems.

Two BCI input variants should be investigated: P300 spelling was chosen because it is one type of widely-researched BCI known to work on the Emotiv EPOC. The device has already been considered as an affordable solution for this text entry system in research (refer to [PK10]). To apply it in a clinical setting, the higher-priced developer's edition would be necessary in addition to the development of an efficient predictive writing system. As a second system, the software classifiers of the EPOC will be investigated. Software keyboard solutions based on switch-access scanning are available for the device, some in ready-to-use solutions with included phrase prediction. However, it has not been investigated how usable this input variant is on this specific device, and which problems interfere with the input.

Until recently, most research into BCI focused on data acquisition of the specific input methods. However, Assistive technology (AT) systems that have been installed in homes of motor-impaired patients were often reported to be abandoned due to low usability and declining user acceptance. The objective of designing AT for communication should be maximizing the information flow and user acceptance while minimizing the mental and physical workload of the user. We therefore assume that taking an Human-Computer Interaction (HCI) viewpoint is crucial to assess the capability of BCI input methods for text entry.

1.2 Objective of the thesis

The evaluation task setting should investigate two possible forms of input on the chosen consumer EEG. The evaluation system limits should allow comparing these two variants, and also mimic subtasks of the process of writing long texts to influence workload and identify usability problems.

The applied metrics should investigate which of the two systems:

- enables higher task performance
- creates less effort or workload for the users
- gains more user acceptance.

The evaluation should also examine usability problems during the input tasks. These should provide a foundation for future research in BCI interfaces. Additionally, the capability of the EPOC device for the widely-studied P300 spelling should be assessed and supported with data on this specific device. The thesis text should as well provide a general overview of the research and technology in BCI for AT at the time of writing.

1.3 Structure of the document

Chapter 2 presents non-invasive BCI based on on-line EEG analysis. After reviewing recent literature on the subject, basics of EEG measurements and the signal components usable for interaction paradigms are described. Free software frameworks usable for BCI are then introduced, as well as the Emotiv EPOC used in this thesis.

Chapter 3 describes BCI designed for AT. After introducing the locked-in condition, literature on BCI spelling systems as well as spelling paradigms is reviewed. Recent research using the EPOC device and available software keyboards at the time of writing is then presented.

Chapter 4 presents the reasoning behind the evaluation methods and the implementation, and discusses the usability measurement methods. There is also a literature review on HCI research in the BCI area.

Chapter 5 describes the evaluation method in detail, including the order of the individual steps, the group assignments, the system calibration as well as special occurrences during the actual evaluation process.

Chapter 6 presents and discusses the collected data and its descriptive statistics.

Chapter 7 summarizes the findings, criticizes the evaluation process and suggests future research directions in the area of the thesis topic.

Chapter 2

Non-invasive EEG Brain-Computer Interaction

This chapter will first provide a general overview about the current literature in the area of Brain-Computer Interaction. This is followed by a short general introduction into Electroencephalography (EEG) technology and to the most prominent Brain-Computer Interaction (BCI) interaction paradigms. This chapter will also list currently available consumer EEGs, with a focus on the Emotiv EPOC used in this thesis and its applicability for research.

2.1 Literature overview

Research into BCI technology first occurred in the early 1970s (refer to [Vid73]). At the time of writing, searching Google Scholar for "Brain-computer interaction" results in more than 18.000 publications in this area. However, only recently the first comprehensive books on this subject were published. Only literature that has been reviewed for this thesis will be listed.

[GAP11] is one recent and comprehensive introduction to BCI technology. The introductory chapters of the manual book of BCI2000 ([SM10]) can also be utilised as a as a general and practically oriented introduction to BCI. [Dor07] can appear slightly dated, but still serve as a general introduction to current BCI research issues. There is another BCI anthology by [BCGM09].

[TN10] is one recent publication in the area of BCI that incorporates Human-Computer Interaction (HCI) knowledge and research. The book discusses both the study of HCI and usability issues in BCI communication and possible applications of BCI measurement technology in HCI research.

For details about the digital processing of the EEG, refer to [SC07]. The book provides a comprehensive mathematical introduction to the automatic (algorithmic) analysis and detection of Event-related potentials (ERPs), medical conditions, source localization and EEG sleep patterns. It also features a physiologically and technically detailed introduction to the EEG.

2.2 Introduction to electroencephalography

The electrical activity on the scalp has first been investigated in-depth by the German psychiatrist and physiologist Hans Berger in Jena in 1924. He published his comprehensive experiences in a 1929 research report (refer to [Ber29]), which became the foundation of the technology. Since then, the recording of the EEG has advanced to a standard in neuroimaging. To name a few, it is being used for research into ERP reactions to stimuli as well as sleep research, diagnosis of neurological disorders such as epilepsy, neurological conditions such as brain death and coma, and as a basis for BCI. In 1968, brain death has replaced cardiac arrest as the scientific-medical criterion for death. Since then, the flat EEG activity of a person has served as one main indication of death.

2.2.1 Origin of the EEG signal

The central nervous system mainly consists of neurons and glial cells. As visible in figure 2.1, nerve cells mainly consist of dendrites, cell bodies and axons. Synapses are the junctions between axons and dendrites, where currents are transmitted through action potentials. Dendrites are connected to axons or other nerve's dendrites and transmit electrical impulses. On average, there are 10000 of these dendrite connections from each nerve cell to other cells. While the cell bodies of a nerve cell contain its main metabolism, the axons are responsible for the transmission of electrical impulses in longer ranges.

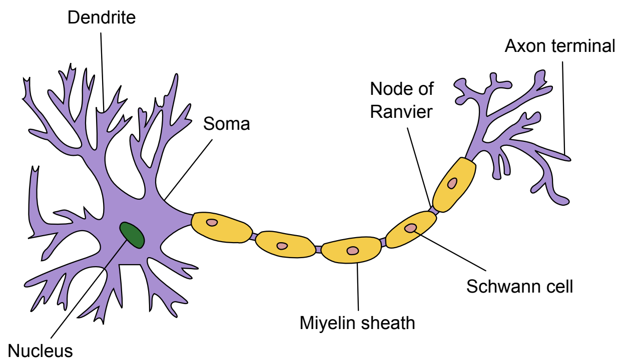


Figure 2.1: Structure of a neuron (by Quasar Jarosz, CC-BY-SA-3.0)

Synaptic excitations of dendrites of pyramidal cells in the cerebral cortex create currents that can then be depicted in the EEG as a measure of the electrical fields generated. Not the action potential reaction to stimuli directly cause the EEG, but the post-synaptic currents.

2.2.2 Acquisition of the EEG signal

Conventionally, 19 to 21 electrodes are placed on the scalp using a conductive gel, however many more can be used for high resolution recording. The raw EEG contains frequency components of up to 300Hz. Within the cell bodies, the potentials reach up to $60 - 70$mV, while the respective signals measured on the surface of the skull only reach voltages from 10 to 100µV, caused by the skull having an approximate resistivity of about 180Ω. They therefore need to be amplified after the acquisition. Additionally, the recording is highly sensitive towards muscle movements at the head, e.g. eye movements and eye blinks, which result in artefacts (Electromyogram (EMG) signals). Many more electrodes can be placed if high resolution should be achieved. The exact positioning of the electrodes is based on the 10-20 International System (see section 2.2.3), which is adjusted to the individual skull shape. It should be noted that in comparison to other imaging techniques such as functional Magnetic Resonance Imaging (fMRI), the recording of EEG has very low spatial resolution, while providing high temporal resolution.

2.2.3 The 10-20 International System

The spatial resolution of EEG measurements is very low (in the range of several centimetres). To enable a maximal amount of reproducibility, standards for electrode position needed to be stated. The 10-20 International System has originally been defined by Herbert Jasper in 1958 (refer to [Jas58]).

It defines the positioning and naming of electrodes around the head. As visible from figure 2.2, the skull is being subdivided into arcs onto which the electrodes are placed.

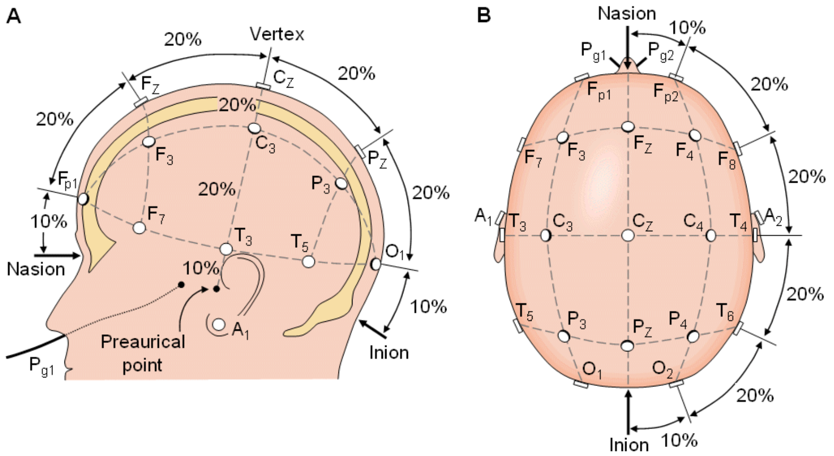


Figure 2.2: Electrode placement in the 10-20 system (image source: [MP95, sec. 13.3])

The electrodes' positions are defined as follows: The left and right preauricular points on the ears are the reference points at the sides of the head. The distance from the *nasion* (between the forehead and the nose) to the *inion* (bump at the lowest point of the skull from the back of the head) is assumed to be 100%. After 10% of this distance, measured from the *nasion* an electrode is being placed. After each further step of 20%, the next electrodes are positioned, until the last one which will have a distance of 10% to the *inion* then. The name 10-20 is derived from this type of subdivision, meaning that adjacent electrode distances are either 10% or 20%.

The letters naming the electrodes are derived from their associated lobes and hemisphere position. F, T, C, P and O stand for frontal, temporal, central, parietal and occipital. C for central is the only letter that is not associated with a physically existing lobe. Z refers to the electrodes on the midline. Even numbers (2, 4, 6, 8) mark the electrode positions on the right hemisphere, and odd numbers (1, 3, 5, 7) those on the left hemisphere.

2.3 EEG interaction paradigms

This section is focusing on non-invasive BCI that uses electrical potentials measured directly on the scalp for interaction systems. To date, almost all BCI research on humans used non-invasive signal recordings. Invasive BCI (i.e. neural implants) will not be described in this chapter. Research into non-invasive BCI mainly focuses on EEG devices. This functional brain imaging technique is especially popular due to its high temporal resolution, and its comparatively cheap technology and local flexibility. Other techniques from neuroimaging (e.g. fMRI) are disadvantageous in these aspects. [TN10, p. 18] provides a complete tabular overview of BCI technology using functional brain imaging considered in research at the time of writing.

2.3.1 Signal components for EEG interaction

The following subsection will shortly describe the four currently most important components of the EEG signal that can be used for EEG based interaction.

P300 The P300 wave is an ERP component that is elicited about 300ms after a stimulus that is surprising or new to the subject. The way of presentation that intentionally should create this reaction is commonly referred to as the oddball paradigm. As in typical ERP experiments, the stimulus has to be presented multiple times to create an average of the signal from which the full ERP waveform can be derived.

Sensory-motor rhythms EEG activity in the motor cortex changes when a person executes or imagines movements. The related signal components are referred to as μ-rhythms, or Sensorimotor rhythms (SMR). Usually, the Event-related desynchronization (ERD) – a gradually diminishing SMR in the planning phase before the movement – is used for detection. EEG patterns can be discriminated by time frequency analysis, i.e. through Fourier transform. The subjects need to be able to actively modulate their brain activity patterns in such a way that they create robust power difference patterns for different mental states. This form of interaction can be learned by advising a subject to use a specific form of movement imagination, e.g. moving one hand or imagining wringing a towel; or by letting a subject explore unspecific movement imaginations until they find the one working best for them (operant conditioning).

In literature, this type of BCI is also often described as motor-imagery BCI or ERD BCI.

Steady state visual evoked potential A subject using a Steady state visual evoked potential (SSVEP) BCI focuses its attention on stimuli oscillating at different frequencies. This procedure attempts to produce similar oscillations at the same frequencies over the visual cortex as well as harmonics of this frequency. These can be detected by analysing the power spectrum created from a Fourier transform. SSVEP paradigms are promising since they need minimal training time and reach high bit rates due to good accuracy. The underlying resulting brain activity can be detected best over the visual cortex (refer to [PAA+03]). It is still debated whether SSVEP interaction relies on the subject's ability to shift gaze direction.

Slow cortical potentials Slow voltage changes deliberately generated in the cerebral cortex can be used for interaction systems. They are referred to as Slow Cortical Potentialss (SCPs) and last 0.5 to 10 seconds. The subjects usually have to be trained over the course of several months to be able to control the negative amplitudes of these potentials. One studied application is moving a cursor vertically over a screen by modifying this amplitude, in order to select one of two presented options (refer to [KKH+99]). This form of communication enables some subjects to enter up to 35 words in one hour.

2.3.2 Interaction paradigms

The signal types defined previously enable the definition of various forms of interaction. Research in the BCI field mainly differentiates between synchronous and asynchronous interaction forms. Refer to [TN10, chapter 2] for further information.

Asynchronous interaction Asynchronous brain-computer interaction usually enables a user to create input events at any time during the time the input system is switched on. Motor-imagery based systems typically use this type of interaction. These systems have to address the problem of unintended input, often referred to as the BCI Midas Touch problem or simply false positives in a broader sense. Depending on the type of signal used, asynchronous interaction may also be suitable for analogue input, e.g. for controlling movements of a cursor or a prosthesis.

Synchronous interaction An asynchronous system will execute input events at clearly stepped time windows. Before these events, data will be collected that is then being classified to detect the intended selection. This type of interaction is primarily suitable for discrete selections. Examples are the P300 spelling paradigm, and SSVEP spelling systems.

Exogenous (evoked) interaction During exogenous interaction, a user focuses his attention on a periodically presented set of stimuli, which create an automatic response in the brain that will then be detected by the BCI, e.g. in EEG patterns. Examples are SSVEP and P300 input.

Endogenous (self-generated) interaction This type of interaction is based on a mental task that the user performs, i.e. imagined movement. Changes in the neuroimage can then be detected by the BCI.

2.4 BCI software frameworks

This section provides a short overview of current BCI software platforms used in research. The evaluation system used the P300 speller of BCI2000, modified in C++. This implementation is frequently used for P300 input experiments.

[BAB+11] provides a comprehensive overview of common and publicly available software platforms for BCI research. The article also discusses strengths and weaknesses of the listed platforms, and an estimation of their individual impact on BCI research. [DKV+10] is a comprehensive and recent review of software frameworks and tools that are freely available to BCI researchers. The applications presented in this paper are based on the MATLAB programming language and computing environment.

BCI2000 BCI2000 is a C++ software framework for stimulus experiments and BCI research. The framework has become a widely used and freely available BCI software. Some of the main BCI input and experiment paradigms are pre-implemented, and signal processing routines are modularized. From the BCI interaction paradigms mentioned earlier (refer to 2.3), it includes P300 spelling, SMR and SCP input. The available experiments can be modified both in the source code and in a graphical user interface. It can also be used for EEG research based on stimulus presentations, e.g. ERP acquisition.

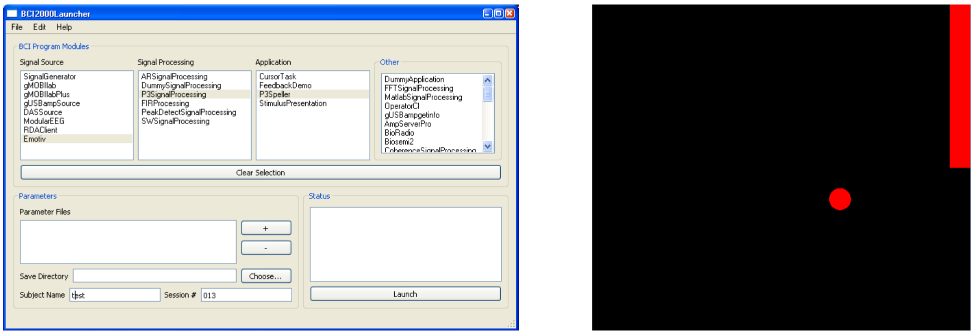


Figure 2.3: The BCI2000 software platform. Left: Module launcher. Right: BCI2000 Cursor feedback demonstration.

Apart from an early overview paper (refer to [SMH+04]), a recent manual book exists (refer to [SM10]). The latest documentation is available from a public wiki system.

OpenViBE This open-source software platform for BCI and stimulus experimentation has been initiated and mainly developed by the French Institut national de recherche en informatique et en automatique (INRIA). The design of stimulus presentation, signal acquisition and classification is based on a detailed set of modules selectable in a graphical user interface, targeted at non-programmers. Alternatively, for designing experiments, the source code can be modified. From the set of BCI interaction paradigms mentioned earlier (see 2.3), it includes P300 spelling, sensory-motor rhythms (motor imagery-scenarios), and SSVEP interaction.

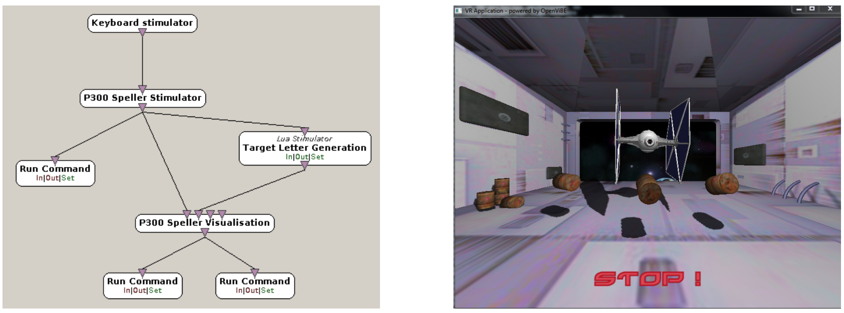


Figure 2.4: The OpenViBE software platform. Left: Module system for experiment scenario design. Right: TIE fighter demonstrator for motor imagery.

Furthermore, there are options to design individual modules. Documentation is provided in the original overview article (see [YFG+10]) and in an online documentation.

Pyff The Pythonic feedback framework (Pyff) by the Berlin BCI group was designed for rapid development and modification of stimulus experiments and feedback paradigms. Since the modification of the C++ experimentation software mentioned earlier can be a bottleneck for both expert and non-expert programmers in BCI laboratories, basing the experiments on the easy-to-learn Python programming language might be one chance to decrease development time and enable more creativity in the design process.

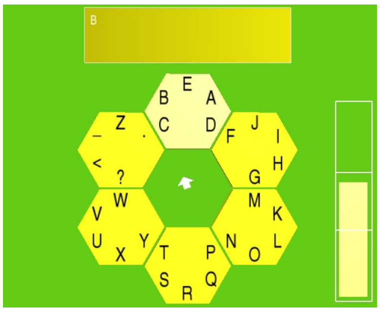


Figure 2.5: The Hex-o-Spell feedback of Pyff.

Pyff is no complete BCI system. It is designed to be integrated in an application chain consisting of data acquisition, signal processing and stimulus/feedback presentation, where Pyff only provides the last part. There is a recent overview article about the framework with code examples (see [BSJ+10]).

2.5 Consumer EEGs

Despite the undoubtedly promising possibilities of BCI technology, so far BCI have mostly not arrived within practical Assistive technology (AT) settings. Apart from their still low reliability and bad usability this is also due to their high cost for acquisition, set-up and usage. Consumer EEGs could present a change for extended use of BCI for AT. This section will provide a short overview of available consumer EEGs at the time of writing, and then focus on describing the Emotiv EPOC system used for this thesis.

2.5.1 Consumer brain-computer interfaces

EEG systems intended for personal usage in neurofeedback applications served a small market segment for a while. The first attempts to use EEG input for consumer entertainment started in 2008, when input systems for video game control have been released. For many of these systems it is still unclear to which extend EEG data is being used in contrast to EMG data. Some of the currently available devices are marketed for research.

At the time of writing, the following systems were available:

NeuroSky NeuroSky[1] started to produce an 1-electrode EEG system for use in toys of third-party companies at the end of 2009 (e.g. “Mindflex”, “Force-Trainer”). They also market their own 1-electrode EEG devices (e.g. “Mind Kit”), as well as a future dry-electrode system for use in research. The latter has been demonstrated to be compatible with SSVEP interaction. The NeuroSky chip has gained attention from the Arduino[2] community because it can be easily removed from the low-priced toy devices and integrated with the microcontroller for visual arts projects.

OCZ NIA OCZ released a 3-electrode EEG system for use in games in 2008. The NIA intents to incorporate both EEG and EMG (from muscle and eye movements) signals for their input detections. At the time of writing this product line is no longer being produced, however.

g.tec The manufacturer of medicine technology recently introduced a consumer BCI that integrates SSVEP control in regular applications within their indendiX[3] BCI product line. This system was demonstrated to work within online role-playing games. They also offer a P300 speller in the same product line.

2.5.2 Emotiv EPOC

The Emotiv EPOC EEG system used in this thesis has been released in June 2009. The system is currently the only product of the company Emotiv Systems. It is being sold with a Developer’s Edition SDK license (300$) and a Research Edition SDK license (700$). The main difference between these variants is that the Research Edition enables access to the raw EEG data recordings, making it the only variant that is compatible with third-party standard BCI and EEG analysis software such as OpenViBE, BCI2000 and EEGLAB. Both product variants provide access to the input data via an SDK interface (e.g. through C++).

[1] Refer to `http://www.neurosky.com/` They also provide a list of academic papers: See `http://www.neurosky.com/AcademicPapers.aspx`.

[2] Refer to `http://www.arduino.cc/`.

[3] Refer to `http://www.intendix.com/`.

Figure 2.6: Design of the Emotiv EPOC device

The original software of the device contains pre-build classifiers that should detect events from the EEG data. They are categorized in "suites" as follows:

Expressive suite (facial expressions) The expressive suite detects facial expressions with a latency of up to 2 seconds. These detections have been considered for AT as a replacement for switches-access scanning with remaining face muscle functions ([LWE11], [NDBD11]). However, AT engineers have been reporting that the detections are not accurate enough to replace traditional muscle movement detection methods (refer to [LWE11, p. 74]).

Affective suite (emotions) The software contains a set of emotion classifiers that have been built by incorporating biofeedback and stimulus experiment data. These are denominated as "Excitement", "Engagement", "Meditation", "Frustration" and "Long-term excitement". [COEK11] have been evaluating these classifiers for their accuracy using self-reported validation, and generally attest positive results.

Cognitive suite (conscious imaginations) The Cognitive suite enables the user to train up to 12 classifiers for different interactions. From these, 4 can be used simultaneously via the EmoKey interface within other software applications. The latency of the detections is 2 seconds on average. Since there is no official technical documentation, it remains unclear which features of the EEG signal the software includes for classification.

Electrode positions The device includes 14 electrodes and 2 reference electrodes around the sides of the head. These need to be moistened prior to appliance with an antibacterial saline solution. The electrodes can be loosely associated with the 10-20 system (see figure 2.7). However, due to different head forms the headset can not cover exact electrode positions that would be necessary within EEG based research.

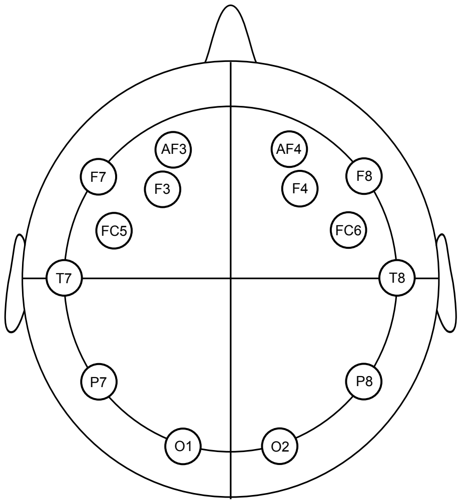


Figure 2.7: Electrode locations of the Emotiv EPOC in the 10-20 system

Emotiv provides more detailed specifications of the system on the product website (see [Emo12]).

2.5.3 Research usage

Emotiv officially advertises the EPOC system as a research device. There has been published research based on the device, with some studies investigating its general capabilities (see [BFC+11]), and some using it for experimental interaction (see the NeuroPhone project: [CCH+10], or the BrainDriver Project: [Bra]) or as a device supporting evaluations in HCI. The company provides an overview of research papers on their website ([Emo11]), which contains 18 publications at the time of writing. It can be stated that the EPOC system has been accepted positively within these types of explorative and proof-of-concept research[4].

2.5.3.1 Research limitations of the EPOC device

Derived from experiences with the device during the evaluation, some disadvantages for use in research can be stated:

[4]There is a forum comment by a neuroscientist reporting about how the neuroscience research community thinks about the device's capabilities and possible improvements. See [Wat11]. According to this, the main criticism is the bad timing accuracy of marker placement within the research software package.

Proprietary technology The exact functionality of the classifiers is not provided by the developers. There are assumptions, e.g. that the classifier features incorporate a mixture of slow cortical potentials and sensory-motor rhythms (refer to [LWE11]). However, since the way the detections work is unknown, it needs to be regarded as a black box by researchers.

Electrode positions Although the electrode positions are derived from the 10-20 system, due to differences in head shapes, the final positions of the electrodes on the head remain mostly inaccurate on the overall static device. In EEG research, due to the already low spatial resolution of this technology the electrodes are usually placed as exactly as possible.

Missing z-electrodes Some signals that can be critical for BCI – such as P300 – are elicited around the electrodes placed around the midline of the 10-20 system (mostly Fz, Cz and Pz). These are missing from the electrode set of the device.

Wet electrodes One of the current engineering research topics in BCI technology is the development of dry electrode systems. The EPOC system needs to be moistened with saline solution regularly to maintain functionality. However, this already represents an advantage to research electrode sets, were the application of gels is necessary.

Also, some advantages of the system are:

Low set-up cost A research EEG and BCI requires expert knowledge, usually through expert training for set-up and during usage supervision. The application of a consumer EEG can be easily teached even to novices. Also, the necessary time for set-up is low (around 5 minutes).

Affordability In comparison to other BCI systems that are suitable for AT communication (e.g. the intendiX system by g.tec, which costs approximately 17.000$ to buy[5]), the EPOC would be affordable for a very low price for rehabilitation centres.

Community support The EPOC has an active online community that does provide help on development and research issues that occur when using the system. The most common software packages for BCI, BCI2000 and OpenViBE integrated the drivers of the EPOC system soon after its release.

[5] Refer to `http://www.intendix.com/`.

Chapter 3

BCI in Assistive Technology

The group of persons who probably will have the most benefit from BCI technology are those who became completely paralysed by illnesses or injuries. Thus, for many years the main focus of research into BCI technology had been to enable locked-in patients to communicate or control prostheses and wheelchairs. The following chapter will both introduce the locked-in condition and possible strategies to enable communication for these patients using EEG interaction.

3.1 The locked-in syndrome

A "locked-in" patient is unable to move any muscles, while cognitive and perceptual functions remain completely intact. The patient is capable of perceiving his environment, but unable to react to it. The condition can be differentiated into the "classic" locked-in syndrome with remaining ability for vertical eye movement and eye blinks, and the "total" locked-in syndrome, where all ability to move or communicate is lost. Other terms like *pseudocoma* or *Monte-Christo syndrome* exist (see [ABJ+01]).

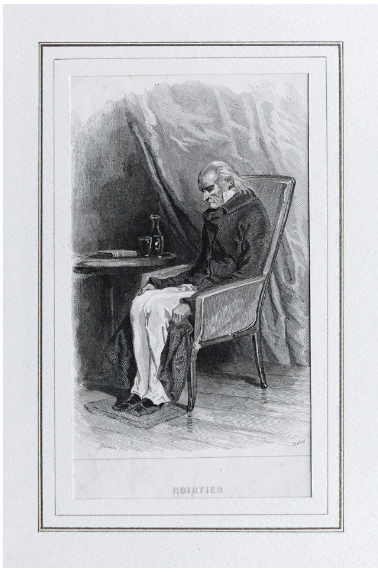


Figure 3.1: Paul Gavarni, Eléonore Sophie Rebel (1845): Noirtier. Copper engraving. Fleurier, Galerie Âme Couleur.

It is necessary to differentiate the condition from the Persistent Vegetative State (PVS), where no cognitive awareness remains. There is concern that many of the patients diagnosed with PVS could be locked-in or have little to full remaining cognitive function[1].

One early description of the locked-in condition can be found in Alexandre Dumas's novel "The Count of Monte Cristo": The character M. Noirtier de Villefort is only able to move his eyes, while being fully conscious. The novel already contains the description of a simple communication system: Noirtier communicates the first two letters of a word by raising his eyes when the desired letter occurs in a recited alphabet, and then chooses the full word by reacting to a finger running over a word list in a dictionary and to the suggestions of his caregivers (see [Dum44, Chapter 58: M. Noirtier de Villefort.]). This communication variant is indeed referred to as partner-assisted scanning in rehabilitation.

Another literary account of the condition is the memoir of the French journalist Jean-Dominique Bauby (refer to [Bau97]), who suffered from a stroke by the end of 1995 and remained in the locked-in syndrome until his death in 1997. With the help of his nurses who recited an alphabet ordered by frequency in French, he was able to communicate the text of the entire book by an estimated 200.000 eye blinks with the remaining muscle functions in his left eyelid. There is a 2007 film based on the memoir[2].

3.1.1 Development of the condition

The locked-in condition can develop from several impact factors. To these, impairment of the spinal cord is common. These will be described in the following section.

Amyotrophic Lateral Sclerosis Amyotrophic Lateral Sclerosis (ALS) is a neuromuscular degenerative illness that typically develops after the age of 30. It is also known as *Lou Gehrig's disease* and *motor neurone disease*. Patients suffering from it gradually and irreversibly loose all muscle functions, up to eye-gaze and vegetative functions such as breathing in its later stages. On average, patients die within three to five years, often due to pneumonia. To date no effective therapeutic strategies exist.

[1] [AMML96] reports that from a sample of 40 patients, 43% were misdiagnosed with PVS. From these, seven have been assumed to be in the PVS state for over one year, including three for over four years. Following the new diagnosis all had sufficient cognitive function to communicate preferences in questions of life quality.

[2] "The Diving Bell and the Butterfly" (USA / France 2007. Directed by Julian Schnabel.)

One well-known patient is the British physicist Stephen Hawking, whose ALS condition is atypical since it progressed very slowly. He uses remaining muscle functions in his cheek to control the switch-scanning interface of a speech generating device. At the time of writing, this last remaining muscle's function is diminishing as well.

Brain stem stroke A major causes of the locked-in condition is haemorrhage in the anterior brain stem, or brain stem stroke. Most patients are not in the "total" locked-in condition, but retain the ability to use some of their muscles. For these patients it is sometimes possible to regain muscle functions with training.

Spinal cord injury Damage of the spinal cord can result in a "classic" locked-in condition. With this origin, often some the muscle functions at the head are retained. Recently there has been intensified research into treatment by stem cell therapies, which remains promising for patients.

3.1.2 BCI communication methods

It is obvious that any way that allows locked-in patients to independently communicate with their environment would dramatically increase their life quality. One of the last remaining communication methods for these people are BCI devices. The target group of these are patients in "total" locked-in condition. If just a single functioning muscle remains intact, it is more effective to connect this muscle to a switch and enable control over a simple switch-access interface. If a patient can still move his eyes, eye-trackers are the more efficient choice. It is still debated to which extend "total" locked-in patients without any working muscle are able to use a BCI. Recently, there has been evidence that BCI performance negatively correlates with the degree of paralysis (refer to [AN08]).

[WBM+02] is a frequently cited publication that provides comprehensive descriptions of available input forms, the target group and active research teams at the time of release. The research team of Niels Birbaumer has worked intensively with locked-in patients and thus could include insightful practical experiences. The publication does not incorporate recent promising developments such as SSVEP input. [Cec10b] is a recent and comprehensive overview of EEG non-invasive BCI communication systems. It highlights current issues and directions of research into the main paradigms.

Recently, research into BCI started to additionally experiment with further communication variants that could increase life quality, such as web browsers[3] or systems for creative expression (see [JSS+10]). Research currently focuses on machine learning for on-line input classification, the design of communication interfaces, prosthesis control, and most recently, usability issues.

3.1.2.1 P300 spelling

The P300 signal is a thoroughly investigated elicited signal detectable in the EEG, following a surprising or rare stimulus approximately 300ms after its onset. It can usually be detected around the midline of the head (Cz, Pz or Fz electrodes), and is preceded by two small positive peaks, and therefore also referred to as the P3 signal.

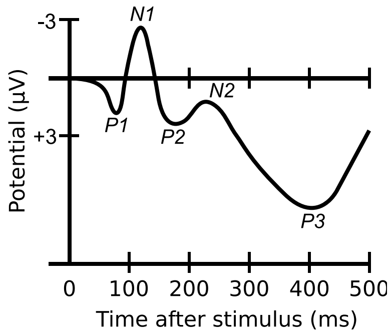


Figure 3.2: Wave showing N100 and P300 components of an ERP.

Due to differences in individual waveforms, it is necessary to calibrate P300 spelling systems. However, in comparison to other paradigms such as motor-imagery the calibration time is low. The oddball paradigm that leads to the elicitation of the P300 wave can be constructed artificially for use in interfaces based on selections. From a set of shortly flashing stimuli that are displayed periodically, the BCI user would focus on the desired one. The P300 signal can then be detected from an average of multiple identical stimulus presentations.

[3] A system that has gained attention is the P300 controlled Nessi browser (see [MBH+08]).

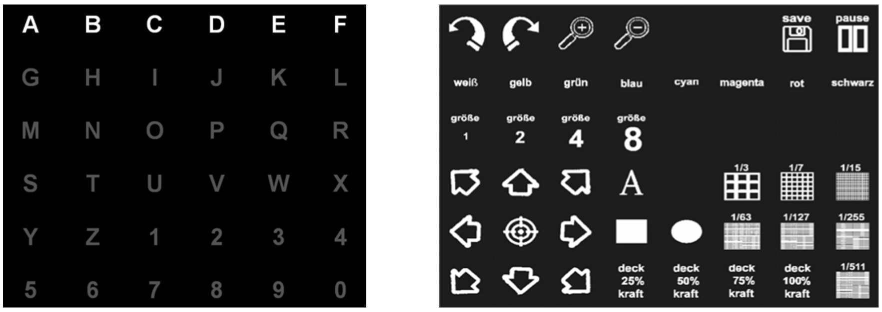


Figure 3.3: Variants of P300 input matrices based on BCI2000. Left: P300 standard matrix. Right: P300 brainpainting matrix by [JSS+10].

This input paradigm is being utilised in the P300-matrix speller system, which was first examined by [FD88]. In its standard implementation, alphanumerical characters are arranged in a 6x6 matrix. Rows and columns of this matrix flash in a random order. If the user keeps focusing on the desired letter, each time the row or column that contains it flashes a P300 signal is elicited. After a predefined number of flashing sequences, the collected data is averaged to detect the P300 signal in the set and the assigned (desired) letter.

Recently (see [BJB+10]), there has been doubt on whether P300 spelling systems rely on the ability to change eye-gaze direction. This outcome would make efficient systems unusable for “totally” locked-in patients. The original paradigm relied on the idea that the P300 signal is elicited whenever the oddball paradigm is fulfilled. Nonetheless, the P300 spelling paradigm generally outperforms motor imagery approaches in terms of training time and BCI illiteracy (see [CSE+09]).

3.1.2.2 SSVEP spelling

SSVEP interaction relies on the reflection of a certain range of frequencies of oscillating visual stimuli in the brain. Usually from a low number of possible selections, the desired one can be detected with good performance and high input accuracy. The Bremen-BCI group recently reported information transfer rates of up to 124bit per minute (refer to [Vol11]). In the same publication, the group reported that they could reduce BCI illiteracy for their system to just 2%. Furthermore, similar to P300 input, these systems need minimal training. However, it is still debated whether patients who are unable to move their eyes can use SSVEP systems.

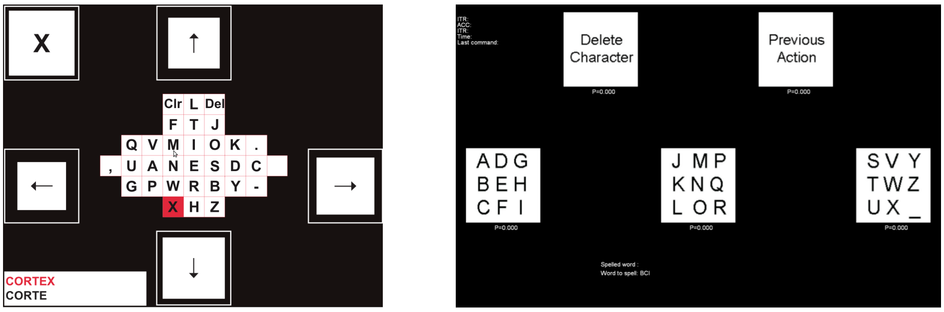


Figure 3.4: Variants of SSVEP interfaces. Left: Bremen-BCI (refer to [Vol11]). Right: The calibration-free CBCI (refer to [Cec10a])

Since the entire number of choices can not be made selectable simultaneously in SSVEP spellers, they usually define a navigation interface (see [Cec10b, p. 3]): For instance, there is a cursor that can be moved over the selectable letters with four types of interaction (Bremen-BCI GUI), or the user navigates through sets of letters until there is only a limited choice left for direct selection (CBCI GUI) .

3.1.2.3 Motor-imagery based spelling

This input paradigm is based on the detection of deliberate changes in the individual EEG by imagining movements. It requires both user and classifier training. Often a limited number of input imaginations are defined which allow control over these systems, e.g. imagining the movement of hands and the movement of feet. Since users must be able to deliberately modulate their brain waves by imagining movement to the extend that it creates detectable changes in their EEG patterns, and since the individual EEG patterns need to be included in a classifier, this system type needs more training time than evoked potential-based systems. However, usually the defined paradigms allow more user control than those relying on evoked potentials.

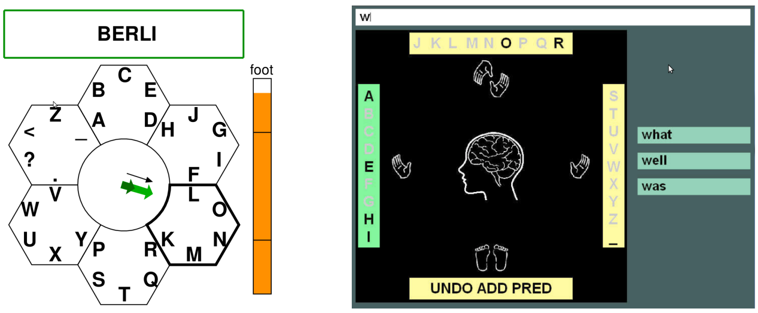


Figure 3.5: Variants of motor imagery interfaces. Left: Hex-o-Spell interface by the Berlin-BCI group ([BDK+06]). Right: AIRLab-BCI predictive spelling interface ([D'A09]).

Since most users can only control only few input imaginations, usually switch-scanning interfaces apply. Examples are Hex-o-spell ([BDK+06]) by the Berlin-BCI group and the AIRLab-BCI GUI ([D'A09], Master's thesis) displayed above.

The Hex-o-Spell system uses two motor imaginations to navigate through sets of letters arranged in a hexagon. Imagining right hand movement controls the movement direction, and feet movement imagination to make a selection. A single letter can thus be selected within two steps.

The AIRLab-BCI is a predictive speller by the Technical University of Milan. Four motor imagination variants can be included to control the system.

3.1.2.4 Slow cortical potential-based spelling

One of the first studies into deliberate EEG pattern modulation was conducted by the group of Niels Birbaumer on SCPs ([WBM+02, p. 773f.]). Users were trained to move a ball on a computer screen within two dimensions to make a selection between four possible choices. Their system is referred to as the “Thought Translation Device”. This SCP system was studied over more than 30 years, including many experiences with patients in late-stage ALS (see [KKH+99]).

The communication system based on it was entitled “Language Support Program”. Here, the user has to select between two halfs of a set of letters until the desired one can be selected among two. This system enables users to write 36 words per hour in the best cases.

3.2 Research into AT using the Emotiv EPOC

Due to the advantages stated in section 2.5.3.1, there has been interest of engineers working in AT to use the EPOC device for paralysed patients. Emotiv systems started to encourage using the device within these environments[4]: They generally recommend no to use the device in security-sensitive environments, were false positives would have dangerous consequences. Usage in communication is stated to be possible.

There has already been research into using the EPOC device for AT applications: One of the first studies exploring this area of application was a wheelchair

[4]There is a thread about this topic by the developers in the Emotiv forums (see [Mac10]).

attached to the device's Expressive suite by the company Cuitech Inc.[5]. At the Microsoft Imagine Cup 2011, the UCEEG team of the University of Canberra placed first with their development of a text entry system controlled by the Expressive suite ([NDBD11]). Also, there has been one recent comprehensive study by AT researchers that explored how the EPOC's Expressive suite compares to traditional muscle-controlled single switch interfaces ([LWE11]). The same publication also examined the training process of the Cognitive suite over several days, however with a small number of participants. There also has been one study about the suitability of the EPOC for P300 spelling tasks ([PK10]), using NASA Task Load Index (TLX) for an estimation of workload. This study generally regarded the device as suitable for P300 detection used for input, however expressed astonishment about the lower-than-expected accuracy and its high variance, assuming that the electrodes were in unsuitable locations for many subjects.

3.2.1 Cognitive suite software keyboards

The cognitive switch paradigm evaluated in this thesis is based on the type of switch-scanning interaction that currently available software keyboards of the Emotiv EPOC device apply. At the time of writing, three software keyboard mainly designed for the classifiers in the Cognitive suite are available:

MindKeyboard This simple application is available for free from the Emotiv website[6]. It enables Cognitive switch control using up to three different thoughts for controlling a cursor moving on a modifiable alphabet. The original alphabet is ordered by frequency. The input thoughts will move the cursor left or right on the alphabet, or trigger a type action. Thus, training of three input variant is necessary for control. Input using the Expressive suite is possible as well.

[5] At the time of writing, only a video of this project remained available. Refer to "EPOCWheelchair.mp4".

[6] Refer to `emotiv.com/store/apps/applications/130/9246`.

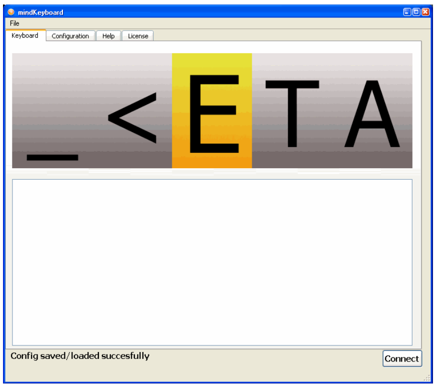


Figure 3.6: MindKeyboard typing application

Neurokey This solution should provide a full keyboard with extended functionality for e.g. browsers based on a scanning and single-switch activation. The user can intentionally change the scanning direction using all available input variants (Cognitive and Expressive suite) of the device. At the time of writing, there is only a demo version available on the Emotiv website[7].

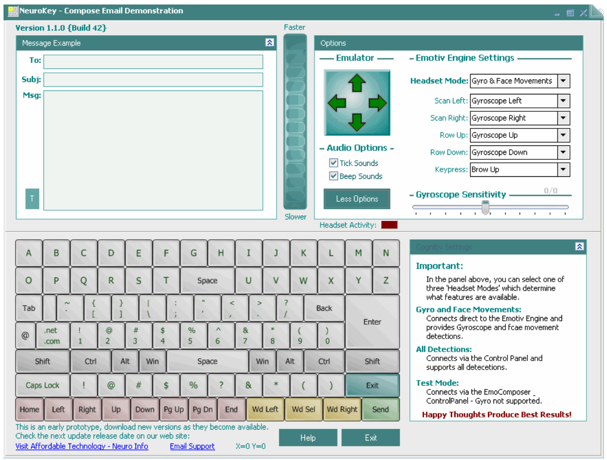


Figure 3.7: Neurokey typing application

Tilvus Assistive Interface The Tilvus solution provides single-switch access to Windows operating systems as well as a comprehensive writing system with phrase prediction and speech synthesis. It has been demonstrated[8] to be compatible with the Emotiv EPOC Cognitive suite. From

[7]Refer to `http://emotiv.com/store/apps/applications/130/727`
[8]Refer to "EPOCTilvus.mp4".

the three presented communication applications this one may be the most flexible and efficient one. However, a full version remains unreleased at the time of writing (see [Til12]).

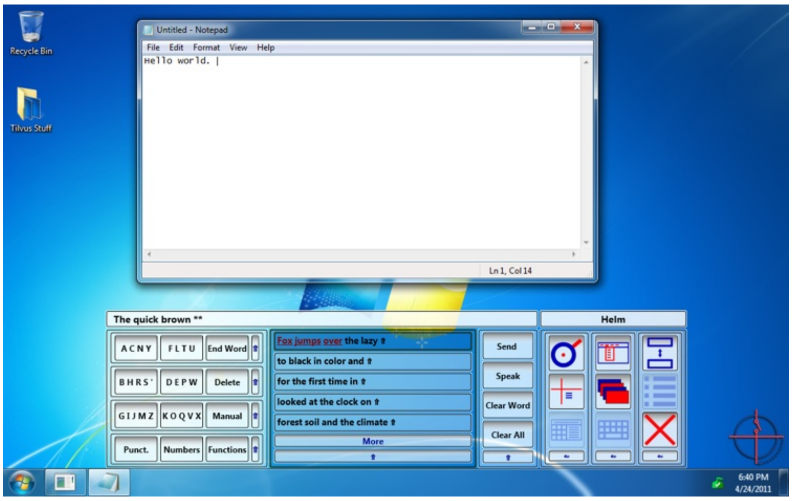


Figure 3.8: Tilvus Assistive Interface

Chapter 4

Evaluation background and implementation

4.1 HCI research into BCI

At the time of writing, all available BCIs can be stated to be difficult to interact with in general: All interaction methods investigated so far suffer from high error rates, and most of the current BCIs also require long training time (up to several weeks for SCP input). These usability problems need to be regarded from the perspective of paralysed patients, for whom even a minimum of elaborate communication increases life quality.

Only recently research into usability and HCI issues of BCI technology began. The released publications are still explorative. There have been initial publications by the University of Twente, focusing on thoughts about usability testing for BCI (see [GPv$^+$11], [vNG$^+$11] and [PGv$^+$11]) and on the lack of usability assessments, and also on the explorative application of self-reported indices (see [PSG$^+$11]). [Sut11] provides a literature review of usability research into BCI. The introductory book [TN10] that was mentioned before focuses on the connection between BCI and HCI within problems of the existing interface technology and the development of metrics derived from neuroimaging technology for HCI studies. Also, there has been a recent doctoral thesis on the subject of usability of BCI for AT (see [Pas11]), which incorporated experiments with paralysed patients for their evaluation. This dissertation does also include lists of usability indices applicable to BCI. Due to the degree of novelty of this area, only ideas of possible interaction problems (e.g. the Midas Touch problem) and a basic methodology set for BCI could be derived from literature review for this work.

It has been realized that the study of HCI and usability of BCI communication systems is necessary in addition to research into input paradigms, classifiers, and physiological possibilities. Patients could otherwise abandon the technology due to demotivation as a response to initially bad results, which is not uncommon for AT.

4.2 Evaluation system

Initially, the question whether a long text can be written on the EPOC as one current consumer EEG should be investigated. Two BCI input variants functioning with this device should be evaluated. One of the approaches (P300) is well-researched, however requires the higher-priced research edition of the device and additional development of a more efficient input system. The second approach (cognitive switch) is based on a single trained imagination in the Cognitive suite and would be working immediately with the developer's edition and already existing predictive software keyboards. The evaluation approach should identify difficulties of the BCI input variants in a sufficiently limited system. With the data it should be possible to provide a suggestion on which communication system to use on the EPOC, or which system to explore further with paralysed patients.

We refrained from implementing a complete system (e.g. with full phrase prediction), since the evaluation should focus on the differences of the BCIs paradigms and their individual problems. It should not focus on other necessary features of a full system, such as prediction quality for the phrase prediction, or clustering for the switch-scanning input. Also, such text input systems had been developed with effective technologies for AT before, and are commercially available (refer to [Til12]).

The system limits should nevertheless leave room for modelling elements of the writing process. The frequently cited Flower & Hayes model of writing (refer to [FH81]) served as an initial theoretical foundation, however the system model that was used in the end only loosely bases on it.

To model the writing process using a phrase prediction system, the system was reduced to the following elements:

Language task The revision step of the Flower & Hayes model was reduced to language processing and language formulation in this evaluation. The main aim of the language task was the activation of the Broca's region, which is relevant for syntactic processing. One way of doing this is presenting subjects with artificial syntactic violations. We expected that a change of focus would result in false positive input (Midas touch problem) during the cognitive switch evaluation, and also distract the subject from their trained activation pattern and thus affect memorability of the system.

Shuffling of word matrix Phrase prediction systems for text entry suffer from re-orientation processes within the constantly updating input matrices. The evaluation system attempted to model this process by shuffling the matrix after solving the language task in 50% of the cases, and the BCI input starting at the same time. We assumed that this would result in reduced performance during the fast-paced P300 input, since this system requires instantly focusing on the desired input option. This is contrary to the rather slow-paced cognitive switch.

Selection The selection of the right answers is performed individually with each BCI input system by each subject. The performance within the selection task is the foundation for task performance measurements and self-reported evaluating.

Developing the system based on these assumptions should allow the additional investigation of a small set of the problems of the BCI input. However, the main background of this three-step system design was to define system limits where a comparisons of task performance metrics were possible, and which could give the subjects an impression of the difficulties of the BCI input variants.

The evaluation compares two BCI input paradigms for text entry. Both paradigms have been demonstrated to work with BCI communication systems on the EPOC.

4.2.1 P300 input

This input paradigm has been introduced in section 3.1.2.1. The evaluation should investigate to which extent the EPOC is suitable for P300 spelling. It should be noted that during text revision it would be necessary to switch off the input system. In the BCI2000 implementation this is solved with a sleep key that needs to be activated twice to continue text input. Expected user interaction problems are:

Physical strain EMG signals from tiny muscle movements are the main confounding variables of this input system. The subjects are told not to move or close their eyes during the input, which can result in dry eyes. The resulting fatigue can interfere with data acquisition.

Constant focus The oddball paradigm must be fulfilled repeatedly during the input, which can only be achieved with constant focus and motivation.

Missing feedback In the BCI2000 implementation, the subject will get no feedback about the letter that is currently about to be selected. Feedback will be given at the end of the input process. Low focus periods during the input process can therefore remain unnoticed.

The implementation is based on BCI2000 v3.0.4, stable code release r3798, with the Contrib package that includes the EPOC drivers. The C++ code of the P3SpellerTask was enhanced and modified for the implementation of the task described above, based on the online documentation of BCI2000 describing the design of a StimulusPresentationTasks.

4.2.2 Cognitive switch

This input method is used for the currently available software keyboards for the EPOC system (previously described in section 3.2.1). A single switch activation pattern will be trained in the Cognitive suite of the EPOC device for a switch-access-interface. It resembles other BCI input paradigms based on voluntary and learned modulation of the EEG, but must be differentiated since the developer of the EPOC system does not provide information on the internal classifiers. The expected user interaction problems are:

BCI illiteracy There will be more subjects who can not create an activation pattern distinctive enough for input in the 30 minutes of available learning time.

Memorability The exact mental state necessary for pattern activation can be forgotten in the course of the evaluation, or during distraction with other tasks (such as the language tasks).

False positives Many of the errors occurring when using the system will be caused by false positive activation. They will also occur when concentrating on other tasks than input.

Midas touch problem This interaction problem is well-known from eye-trackers, with one of the first descriptions in [Jac91]. A user fixates on a selectable region of the screen to read or obtain information, and the program interprets this process as an input command. It can also occur in imagery-based BCI, where normal thinking processes or abstraction can result in a detection of the activating imagination, since the BCI processes all brain activity as possible input. The name is derived from the King Midas legend.

The application for the cognitive switch was implemented using Processing[1], which is a sufficient script language for prototypes of user interaction systems. Signals are sent from the device to the processing application via UDP in the OSC format using Mind your OSCs, which is freely available for the EPOC.

Items in the input matrix will be selected using a switch-scanning interface with row-column-scanning (see figure 4.1). The time necessary for recalling the activation pattern (switch activation time) will be measured before the evaluation (refer to 4.2.2). The switch activation time also determines the time the cursor will stay on one option. It is necessary to determine the switch release time as the time span between a selection and continued scanning. If a subject selects a wrong row, he can escape that row by waiting until the column scanning reaches its end. At the end of a row scanning the whole input field is highlighted, and the application switches back to the row scanning if the pattern is activated in this moment.

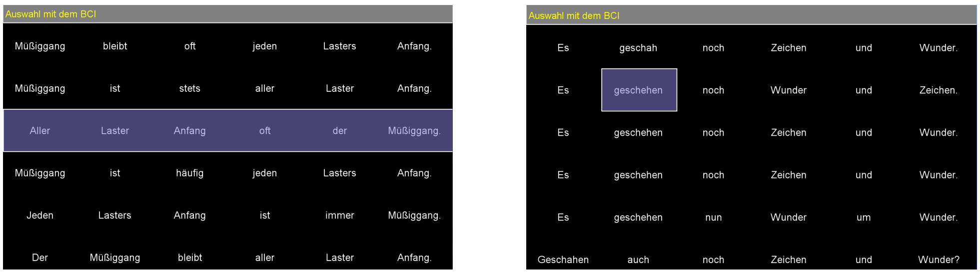


Figure 4.1: Screenshots of the evaluation system for the cognitive switch showing row selection and column selection.

Interest into the technology and the simple applicability of the device allowed evaluating a high number of participants. The last third of the participants could therefore be evaluated with a switch-scanning system including simple feedback on the current classifier results for the trained pattern. The height of the color in the selectable item visualizes the continuous classification values from the device, as well as the input threshold. See figure 4.2.

[1] Refer to http://www.processing.org/.

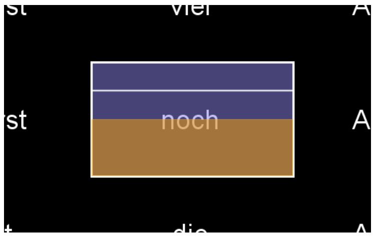


Figure 4.2: Cognitive switch system with feedback

Since the cognitive switch is an asynchronous system and the switch-activation user-controlled, there is the chance that it does not need to be switched off during text revision, which would make writing more efficient. However, input conflicts such as the Midas touch problem known from eye-tracking systems have been reported. The evaluation system attempts to investigate this for writing processes: If false positives occur frequently during solving the syntactic task (as a model for text revision) in the evaluation, it would also conflict with user interaction in a fully-functional application.

Switch activation time For switch-scanning interfaces, often a switch-activation time and a switch-release time are measured before appliance. This determines how long the cursor will stay on each option and give the user enough time to press and release the switch. This is necessary for motor-impaired users who often have only few muscles left, and difficulties using them for controlling computers (e.g. when using head switches). For the cognitive switch, this time includes the classifier latency of approx. 1-2 seconds and the time necessary for the subjects to modulate their brain waves.

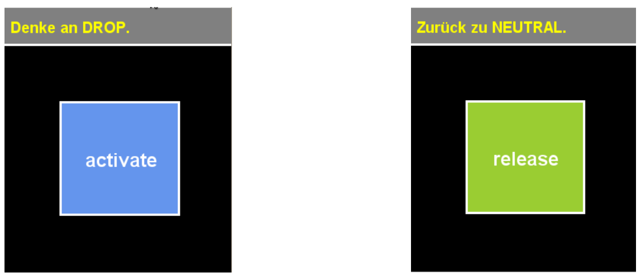


Figure 4.3: Screenshots of the Switch activation measurement application.

The small Processing application measuring these times in this evaluation is depicted in figure 4.3. This time should not exceed 5 seconds, otherwise the subject was asked to train the classifiers again. The maximum times were estimated and rounded to a 500ms step (i.e. 2500ms, 3000ms . . .). The cursor then stayed on each option for the switch scanning time. After selecting a row, the cursor would disappear for the time that was set in the release time variable.

4.3 Usability scales

During the evaluation self-reported metrics were used to quantify general impressions of the input variants. The standard questionnaires TLX and System Usability Scale (SUS) were selected since they serve as measures for workload and user acceptance during system evaluations. There is a lack of this kind of self-reported scales specifically for BCI in research methodology.

This section shortly describes these scales, and other self-reported scales that were considered for BCI usability research before. For a more comprehensive overview, [Pas11, p. 75-82] can provide a good reference. [vGP+11] discussed the design of a user experience questionnaire specifically for BCI.

4.3.1 NASA Task Load Index

The TLX scale was developed by the Human Performance Group at NASA's Ames Research Center, and released in 1986 (refer to [Har88]) in order to measure workload as a representation of the cost for a user for accomplishing something. It had a notable influence on HCI research ([Har06]), with more than 5.500 results on Google Scholar at the time of writing.

TLX originally consisted of two parts: The total "workload" is divided into six subscales that are represented on a single page, serving as one part of the questionnaire. Each of these represents a cluster of subvariables in understanding. They are rated for each task within a 100-points range with 5-points steps. These ratings are then combined to the task load index.
The second part of TLX creates an individual weighting of these subscales by letting the subjects compare them pairwise. The weights are then applied to create the overall task load index.

Many researchers eliminate these pairwise comparisons alltogether, referred to as "Raw TLX" (as described in [Har06, p. 3]). There has also been statistical evidence (see [Bus08]) supporting using the shortened version.

TLX had been applied exploratively to BCI usability research before (refer to [PSG+11]). During the evaluation, a single-page paper and pencil version of "Raw TLX" translated to German was used (refer to appendix B.2).

4.3.2 System usability scale

John Brooke developed this scale at the Digital Equipment Corporation in 1996 as a measure for the satisfaction of a user of an evaluated system. It has become one frequently used questionnaire, probably because it is very quick to apply and to evaluate, and also applicable for a wide range of different types of systems.

The subjects evaluate ten usability items on five points Likert scales, using a dimension from “strongly disagree” to “strongly agree”. As in TLX, the final scale is formed with these individual ratings. The original SUS was carefully designed using a pool of 50 potential items tested with 20 participants for two inherently different systems. For the scale he selected those items that led to the most extreme responses. Brooke did not evaluate the items for reliability or validity, and referred to it as a “quick and dirty” scale in the title of the original publication.

Since half of the items were not applicable to the interaction forms in BCI, the original SUS was modified for this evaluation (refer to appendix B.3). Since these modified items could not be derived from an evaluation as the original SUS, the changes can be viewed as problematic. The results can not be compared to other HCI experiments using the SUS. The original SUS has been applied to BCI usability research by [PSG+11].

4.3.3 SUXES

The SUXES method (refer to [THM+09]) has been applied to BCI exploratively by [GHP11]. Usability is assessed based on the differences between the two-dimensional scales in two questionnaires completed before and after a product is used. The subjects state which expectations they have towards a system, which values are desirable and acceptable, and in the end how the subjects experienced the actual system during the evaluation.

Chapter 5

Evaluation

5.1 Objective

The evaluation will collect data on performance and user experience of the P300 and the cognitive switch input paradigms on the Emotiv EPOC with a mimicked writing process. The measurements will focus on task performance, information transfer, error types and user experience (TLX, SUS) metrics.

For each BCI input system, the participants will calibrate the system and subsequently complete ten input tasks. Each tasks consists of up to three cognitive steps:

1. Finding a syntactic violation in set of sentences. Press [STRG] button to start the input process.
2. During 50% of the tasks, the rows of the text matrix will be shuffled.
3. Input of the subject's solution using the BCI.

Every subject will test both systems ("within-subjects design"). Two groups were defined:

- Group A|PE (with 1. P300 input, 2. Cognitive switch) (A|PE)
- Group B|EP (with 1. Cognitive switch, 2. P300 input) (B|EP)

Due to general interest in BCI, the intended number of participants could be found quickly. Therefore it was possible to include a second variant of the cognitive switch paradigm that included a feedback mechanism. The evaluation thus needs to be regarded as a mixture of a within-subjects and a between-subjects design.

5.1.1 Limitations

The following limitations of the evaluation situation will be in place:

Emotiv EPOC signal quality The signal quality of the device is sufficient for ERP (P300) detection. The signal processing and classification algorithms are optimized for functionality in interactive entertainment applications.

Emotiv EPOC electrode positions The electrode positions are optimized for easy application of the device. They do not include the electrodes around the midline (e.g. Pz, Cz, Fz), which are usually included in P300 experiments ([KSM+08]). → The data for the P300 signal classification is not collected the optimal locations, which may influence the measurements. The results can be vulnerable to head shapes.

Artificial task environment The option to evaluate a fully functional or semi-functional prototype was rejected due to concern about the measurability and comparability of the individual performances. Instead, the evaluation attempts to mimic the writing process. → The systems are not being evaluated in an ongoing writing process.

Singular evaluation Since each participant will only be evaluated once, the experiments will focus on instant usability. In clinical studies, BCIs are often tested over the course of several days or event months with regular calibration and learning steps, resulting in improved task performance over time. → This presumably lowers Task Completion Rates (TCRs). The evaluation data can only estimate the performance of subjects that are new to the system. [LWE11] demonstrated specifically on the EPOC that the input accuracy improved when training over the course of several days.

Healthy participants Due to practical feasibility and ethical concerns, no paralysed patient will evaluate the system. This is general practice for explorative evaluations in BCI research. → The applicability of the Emotiv EPOC neuroheadset for clinical AT situations is not being evaluated.

5.1.2 Hypotheses

The evaluation system and its objective presupposes underlying assumptions. These hypotheses will shortly be reviewed and discussed during data presentation. H1 and H2 state that communication is possible at all using the investigated BCI variants and the EPOC device. H3 and H4 expect different outcomes for the total workload and user acceptance scales. H5 and H6 assume the influence of two usability problems that should be investigated with the evaluation system.

H1 Non-impaired subjects can attain Information Transfer Rates (ITRs) greater than zero during P300 input.

H2 Non-impaired subjects can attain ITRs greater than zero during cognitive switch input.

H3 The combined TLX ratings will be different for P300 input and cognitive switch input.

H4 The combined SUS ratings will be different for P300 input and cognitive switch input.

H5 False positives will occur during the syntactic tasks for non-impaired subjects in negative correlation with their TCR (Midas touch problem).

H6 On average more errors per task will occur if the input matrix changes at the start of the BCI input during P300 input in comparison to cognitive switch input.

H7 The subjects will be more likely to accredit errors to themselves during cognitive switch input in comparison to P300 input[1].

5.2 Experimental method

30 able-bodied volunteers aged from 19 to 59 years (average age: 27.5; median: 26) took part in the evaluation. 24 of the participants were male and 6 were female. No participant had previous experience with BCI. All of them were students or employees of the Bauhaus University, with 26 of them studying or employed at the Computer Science division of the Faculty of Media. Due to that, most of them were familiar with Interface evaluations or Machine Learning terminology (e.g., concepts like "classification").

The participants were searched through the university electronic bulletin board and by personal contact. After making an appointment, they received the evaluation instructions (see appendix A) and were asked to skim them before the evaluation to familiarize themselves with the evaluation situation. The evaluation situation was expected to create stress for the participants and impair

[1]This reflects the problems of deliberately recalling and withholding an exact imagination for input in the right moment.

creativity. Therefore the participants were asked to think about an activation pattern for the cognitive switch before the evaluation. Additionally, they should not consume caffeine[2] for four hours before the test if they were not regular coffee or tea drinkers who need caffeine for concentration.

Each participant tested both input methods in a single session that lasted 164 minutes on average. At the beginning and end of the evaluation, and after each input system the subjects were asked to fill in questionnaires and usability scales. The subjects were asked to "work quickly" through the tasks. Refer to table 5.1 for an overview over the whole evaluation process.

Table 5.1: Order of the complete evaluation

A|PE	**B|EP**
demographic questionnaire	
P300 input	Cognitive switch
NASA TLX, Modified SUS, Final questionnaire	
break	
Cognitive switch	P300 input
NASA TLX, Modified SUS, Final questionnaire	

The mean cognitive switch evaluation time was 24.5 minutes. P300 evaluation lasted 22.8 minutes. Both system evaluations were preceded by approximately 20 minutes of calibration and followed up by 5 minutes of questionnaires. Before evaluating the second input system, the Emotiv EPOC was shortly removed for sensor moistening. See table 5.2 for an overview of the individual procedures.

Table 5.2: Order of input system evaluations

P300 input	time (min)	Cognitive switch input	time (min)
P300 calibration	20	pattern training	up to 30
build classifier	5	switch activation time	5
evaluation (10 tasks)	22.8	evaluation (10 tasks)	24.5
TLX, SUS	5	TLX, SUS	5

[2]Caffeine is proven to increase P300 amplitudes [ANO06].

Confounding variables In test trials of the evaluation, light created noise in the data and impaired the P300 classification results. Therefore, both evaluations were conducted in a darkened, large-area technical laboratory at the Bauhaus University, were light conditions could remain identical for all subjects. The subjects were seated in a distance of 50cm in front of a single LCD screen. Due to construction noise at daytime during the evaluation, all participants conducted their tests in the evening between 6pm and 12pm.

During most of the evaluations people were working in the large-area laboratory. The background noise for most participants consisted of silent keyboard sounds, a closing door and silent conversation. We deliberately included the sound of a closing door once at a certain step during the evaluation to allow the subject estimating whether environmental sounds were disturbing for him. If noise became too loud (e.g. construction noise), the evaluation was postponed.

5.2.1 Calibration procedures

The following sections describe the calibration procedure for both systems in detail. During the calibration and data collection processes, the experimenter withdrew from the participant, stating that he would not be watching the scene. Apart from the obvious fact that it was not necessary for someone to sit next to the participant, especially in the cognitive switch evaluation the experimenter would have been influencing the input process since somebody watching would place additional pressure on the subject.

5.2.1.1 P300 speller calibration

The calibration process of the P300 spelling was in line with the instructions in the BCI2000 manual book (see [SM10, p. 73-79]). The process took 20 minutes for each subject. An unmodified variant of BCI2000 was used with standard settings (see Table 5.3). Keep in mind that variations of many of these parameters have been demonstrated to influence the P300 input process.

Table 5.3: BCI2000 settings during P300 calibration. Refer to [SM10, p. 185ff.] for parameter descriptions.

epoch length	Stimulus duration	ISIMinDuration	ISIMaxDuration
800ms	31.25ms	125ms	125ms
HighPassfilter	LowPassFilter	TargetWidth	TargetTextHeight
at 0.1Hz	Disabled	16	4
TargetHeight	Number of sequences	SampleBlockSize	SamplingRate
14	25	4	128Hz

To allow the subjects to shortly close and rest their eyes between entering the letters, the PostSequenceDuration was set to 4 seconds (or more if the subjects requested it).

Before starting the evaluation, we presented on-line EEG recordings to the subjects to sensitize them for the noise created in the signal by tiny muscle movements. The subjects got a general introduction to input using their P300 signal (see A.3). The input instructions were presented with emphasis on critical issues, i.e. asking the subjects to "rejoice in the same way each time the desired letter flashed", and to "stay calm and happy". The data collection for the calibration was done using the word set:

THE • QUICK • BROWN • FOX • JUMPS

A pre-classification was conducted on the current data set after calibrating the word FOX. This performance feedback presumably increased the subject's motivation before starting the last data collection sequence. If a subject appeared to be tired, a pre-classification was conducted after BROWN or QUICK to increase their motivation. In addition in these cases, to maintain their interest, topographic maps or waveforms of their P300 signals were presented to the subjects after the data collection. These were created using BCI2000's Offline Analysis tool.

After this data collection, BCI2000's P300Classifier application was used to build a classifier. At least the three best word data sets were used if there was a bad data set created during the calibration (i.e., one that clearly reduced the classifier accuracy). However, bad data occurred rarely, so all five data sets were used for most of the subjects. The resulting parameter file was then loaded into the modified P300 spelling task.

The necessary number of epochs for 100% classification accuracy was determined using the classifier results. The BCI2000 manual book suggests to use the number of flashes to reach 100% once (see [SM10, p. 78-80]). However, to increase stability of the input classifier during the system evaluation, the corresponding parameters (EpochsToAverage and NumberOfSequences) were set to the epoch number with the *third* occurence of 100% accuracy.

BCI illiteracy: P300 spelling There was no rule for interruption due to BCI illiteracy during the P300 input evaluation, since this case did almost not occur. The evaluation was interrupted for one participant (27t), who failed to enter the solution for the first 5 input tasks after 5 attempts for each task. He is the only one who appears as P300 BCI illiterate in the data.

5.2.1.2 Cognitive switch calibration

Due to a lack of experience with the calibration process in the Cognitive suite, no strictly organized calibration process was determined for the evaluation. Advice was derived from personal and user experience in the Emotiv forums. This assessment is in line with previous research on the Emotiv EPOC (see [LWE11, p. 75]). Also, the input ability of the subjects could not be as clearly defined as in the P300 input. The skill rating provided in the Cognitive suite proved to be no reliable metric for the accuracy of the classification and the input ability of the subjects. During the evaluation, one of the subjects with a TCR of 100% only had a rating of about 20%.

Therefore, we decided that the aim of the calibration process was to enable the subjects to use the cognitive switch as good as possible during the evaluation design's minimal training time. During the calibration a set of criteria for input ability was determined, which was assessed by watching the subjects and relying on their personal reports:

- Almost no false negatives occur (e.g., when the subject was asked to demonstrate his cognitive switch).
- Almost no false positives occur (e.g., during talking to the subject).
- The subject reports that he feels in control and is mostly secure about the switch activation.
- The switch activation time does not exceed 5 seconds.

Using these criteria, we decided that it would be best if the users had full control over the calibration. This free training process was also expected to reduce the influence of one of the prominent confounding variables – the mental pressure. The subjects were allowed to familiarize themselves with the input system, try different patterns and find the one that worked best for them. They could also find out whether there are special triggers in their patterns that served as switch activations.

The subjects were free to select their own input pattern. They got advice on possibly well-working activation patterns prior to the evaluation (see appendix A.4). Vivid imaginations, e.g. derived from clear memories were stated to be well-working activation patterns. From a BCI point of view, this approach can be described as operant conditioning (see [TN10, p. 10]).

The subjects first got a short introduction to the Cognitive suite. The main functions were explained, and demonstrated for some. After that, the subjects were given control over the system (without an experimenter watching them), and were asked to train classifiers for two mental states using the self-selected activation pattern:

DROP The subjects should clearly think of their complete activation pattern. They should prefer to train the classifier few times in the same way, and then try to "reach" the imagination voluntarily. It was presumed that training too often would blur the data, since it was expected to be difficult and tiring to recreate the very same imagination over and over again.

NEUTRAL The subjects were advised to behave "normally", without mistakenly falling into an unnatural "Zen" state of mind. As suggested in the Emotiv EPOC manual and in the user experiences from the Emotiv forums, they should look around, read, talk to people, or think of everyday things. They should additionally record data for this classifier if many false positives occurred.

The subjects should inform the experimenter as soon as they felt confident using the DROP imagination. Afterwards the switch activation time was measured using the system described in section 4.2.2. In case they succeeded in this application, we continued with the evaluation system. For subjects failing the Switch activation time measurement, additional or new training data was collected.

Please also see chapter on future work and for a critical view on this approach.

BCI illiteracy: Cognitive switch BCI illiteracy was attested if a subject would not be able to create a stable input pattern after two 30 minutes training sessions with the criteria mentioned earlier. If a subject was not able to create an activation pattern after 30 minutes of training in the cognitive suite, he was re-invited. If the subject failed to do so in a second training session, BCI illiteracy was attested, meaning that the subject was illiterate for this present experiment. Refer to [TN10, p. 35-54] for a comprehensive discussion of BCI illiteracy.

5.2.2 Group assignment

The subjects were randomly assigned to A|PE and B|EP. Refer to table 5.4 for an overview. The number in the parentheses is the planned number of subjects, with the number in front being the actually evaluated number of subjects for this group. Missing subjects arose due to BCI illiteracy, which is listed in the last row (with E. . . Cognitive switch illiteracy and P. . . P300 illiteracy).

Table 5.4: Group assignment of the subjects

Group assignment			
	task set	1—2	2—1
A	no feedback	3(4)	4 (4)
	feedback	2(3)	4 (4)
B	no feedback	4 (4)	4 (4)
	feedback	3 (3)	2 (3)
	BCI illiteracy	EEE	EP

As evident from table 5.4, additionally to the system order the task sets were permuted to equilibrate difficulty differences.

BCI illiteracy occurred for 5 of 30 subjects (17%) for one or both system(s) using the criteria stated in 5.2.1.2. The available literature reports between 15% and 30% of BCI illiterate subjects for sensorimotor rhythm modulation (see [DSM+09]), which is possibly relevant for the cognitive switch paradigm on the Emotiv EPOC. The possibility that subjects with very low task performances could also be added to this group was not investigated systematically.

5.2.3 Stimulus material

The stimulus material was not created with a stable algorithmic approach due to unavailable documentation on this approach. Instead, the stimulus material for the syntactic tasks was derived manually from German proverbs. Each task consisted of six similar variants of one German proverb. The meanings of the proverbs were modified to the extent of inexistent sense.

Two types of syntactic violations typically used during N400 experiments were created:

Missing word Select the last word before the gap:

Millionen Fliegen können sich nicht irren.
vs.
Millionen **Fliegen** sich nicht erneut irren.

Wrong inflection Select the wrong inflection:

Eine Schwalbe macht noch keinen Sommer.
vs.
Eine Schwalbe **machten** noch keinen Sommer.

An example stimulus task is presented below:

Der nicht gewinnt, der nicht wagt.
Wer nicht wagt, der nicht gewinnt.
Wer auch wagt, der gewinnt mit.
Auch wer wagt, der gewann nicht.
Wer wagt nicht, der noch **gewannen**.
Wer nie wagte, der gewann nicht.

5.2.4 Error types

After each input error, the subject will self-report on an error type. The subjects got following descriptions (written (see A.2, and orally) of the error types:

"Konzentration" (concentration) The subject noticed that he lacked attention, that his mind was wandering, that he became tired or could not remember how to recall the activation pattern in time.

"System" (system) The subject could not state a concentration problem and therefore assumes that the response of the system was incorrect.

"falsche Antwort" (wrong answer) The subject selected a wrong answer.

In addition to these instantly determined error types, after each input system evaluation the subjects will select two items in an additional short questionnaire on detailed error types.

5.2.5 Special occurrences

Three special occurrences should be mentioned: One of the subjects began to feel nauseous during the P300 spelling task (14t), informing the experimenter about this situation at the end of the evaluation. When one female subject (10t) evaluated the systems, the EPOC slipped from the head during both evaluation situations, which remained unnoticed halfway through the evaluation. Both events surely impaired the results, leading to low task completion rates for both subjects. The data set of 20t was not included because he aborted the evaluation due to feeling uncomfortable wearing the device (headaches).

Chapter 6

Results and discussion

This section lists, depicts and describes the findings of the evaluation. A short interpretation and discussion of the results is attempted for each resulting item. The findings are placed in literature where it is useful. Please note that the cognitive switch paradigm is often abbreviated to "EPOC" in the plots. For an overall conclusion, please refer to the last chapter.

6.1 Task performance

The task performance measurements consisted of task completion rates, the frequency of errors and ITRs. These were the main non-self-reported measurements during the evaluation. As mentioned before, the error types contain a self-reported component, where subjects had to categorize the error right after it occurred. It should be noted that task completion rates and error frequencies do not directly reflect each others distributions. To give an example, task completion can be 100% if a subject needed 4 attempts for each evaluation task, but did not fail one. This resulted both in a high TCR and a high total error rate.

6.1.1 Task Completion Rate

The TCR is the ratio of successfully completed evaluation tasks to the total number of tasks presented:

$$\text{TCR} = \frac{\text{successful tasks}}{\text{total tasks}}$$

During the evaluation, for each task the subjects were given 5 attempts. Also, during the cognitive switch task, a task was aborted and labelled as unsuccessful after a 10th loop of the scanning cursor occurred (both rows and columns), in which case the subjects could not reproduce their activation pattern during the task.

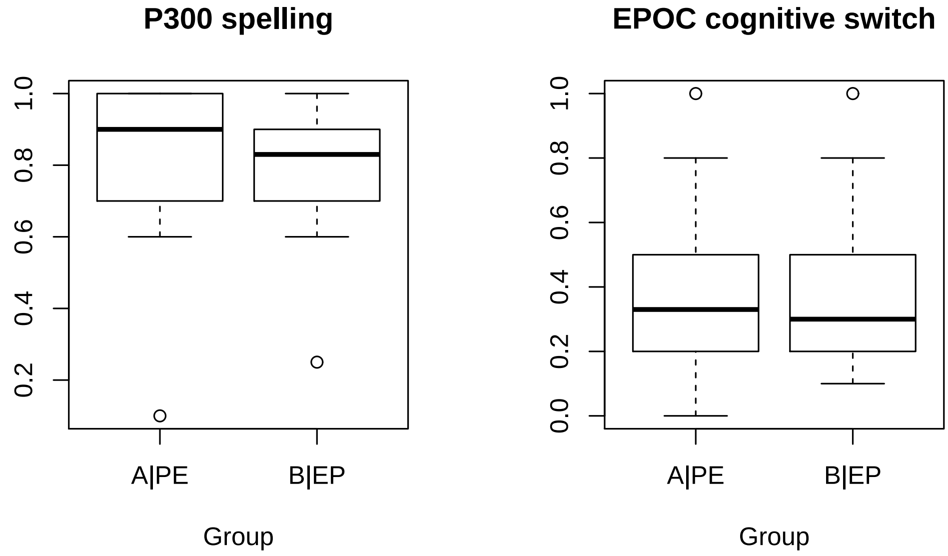


Figure 6.1: Comparison of TCR by input method and group

Data presentation Most subjects achieved their highest TCR during P300 input (mean: 79.6%, median: 87.5%, confidence interval $[0.714, 0.878]$). In this task, 8 subjects achieved a TCR of 100%, while only 2 subjects had a TCR of less than 50%. These subjects can be classified outliers: For subject 10t (TCR 30%), the EPOC device slightly slipped from the head and thus changed electrode positions (and thus classifier accuracy) during the evaluation. This was noticed at the end of the tasks. Subject 14t (TCR 10%) mentioned at the end of the system evaluation that he felt nauseous when focusing on the flashing matrix.

During the cognitive switch task, the subjects achieved a much lower TCR (mean: 41.3%, median: 33%, confidence interval $[0.301, 0.526]$, significant mean difference to P300 input with $p \approx 0$). 17 subjects did not achieve a TCR higher than 50%. Only two subjects, one in each group achieved a TCR of 100%.

The order of the input system presentations had a stronger influence on the task performance rate in the P300 spelling task (group A|PE mean: 90%, group A|PE median: 80.5% vs. group B|EP mean: 83%, group B|EP median: 78.6%)

than during the cognitive switch (group B|EP mean: 42.5%, group B|EP median: 30% vs. group A|PE mean: 40.1%, group A|PE median: 33%). diff. sign

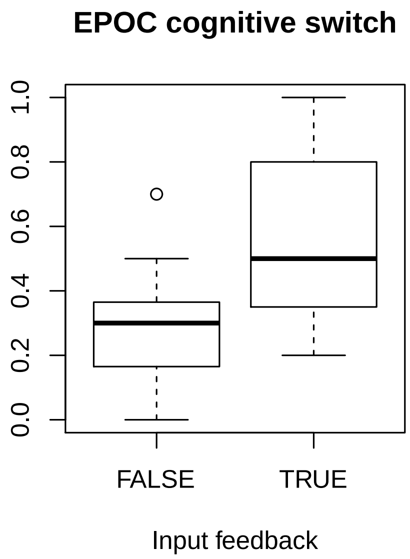


Figure 6.2: Influence of input feedback on TCR

Adding feedback to the cognitive switch input clearly improved the performance in this evaluation. Without feedback, the mean TCR was 30%, with the median of the distribution at 28.6% and a confidence interval at $[0.184, 0.388]$. With feedback, the mean TCR raised to 58.6%, with a median at 50%. The mean difference is significant ($p = 0.0097$).

Discussion It can be assumed that the higher P300 task performance is due to it being based on a physical reaction. Focusing on stimuli requires far less mental contribution of the subject in comparison to the deliberate cognitive switch activation. The subjects were not required to learn how to modulate their brain waves, and did not have to remember an activation pattern between the language tasks. Also, the cognitive switch paradigm counted each pattern detection above the threshold as an activation, which resulted in frequent false positives and thus many failed tasks. In comparison, due to the nature of ERP detection – averaging the shape of the signal over multiple trials –, erroneous P300 classification and input is prevented by verifying the recognition with a high number of stimuli presentations and response recordings.

The influence of the group assignment (i.e. the order of the input systems) is more pronounced during the P300 input task. The TCR of P300 input weakens when the cognitive switch is evaluated first. Possible carry-over effects from the cognitive switch task are a lack of focus, increased tiredness and low motivation towards BCI. P300 does not require as much mental effort as the cognitive switch input, however constant and strenuous focus. Also, there could be frustration and disillusionment due to low performance in the asynchronous task. The aim and design of [THM+09] suggests that there is experience with carry-over effects of the low reliability of certain BCI paradigms. It could also be assumed that the subjects expected the EPOC input to be similarly simple as the P300 input. This would result in a lower TCR of group A|PE. However, this is not reflected in the data.

6.1.2 Information transfer rate

The ITR for BCIs was defined in [WBM+00] as a general measure for comparing the performance of different BCI methods. Since then it has been an accepted measure in BCI research and can be found in many studies. It is calculated as follows:

$$B_{trial} = log_2 N + P log_2 P + (1 - P) log_2 \frac{1 - P}{N - 1}$$

B_{trial} are the bits transferred per input trial. P represents the probability that the desired item is selected ($\frac{\text{desired selections}}{\text{total selections}}$). N is the number of available selections per trial. An estimation of transferred bits per minute can then be obtained by:

$$B_{min} = \frac{60sec}{T_{sec}} \cdot B_{trial}$$

Here, B_{min} are the bits transferable per minute, while T_{sec} is the length of a trial in seconds. Before the recent research into SSVEP methods, ITRs for BCIs were reported to be in maximum ranges of 5-25 bit/min (refer to [WBM+00]).

P300 input ITR For P300 input, ITR was determined with $N = 36$ and

$$T_{sec} = (\text{epochs} \cdot 2 \cdot 6) \cdot (31.25ms + 125ms) - 125ms$$

(with Stimulus duration = $31.25ms$ and Inter stimulus duration = $125ms$, refer to table 5.3). The number of *desired* selections consisted of the correctly selected items and the answer errors. For the total number of selections, the system and concentration errors were added to this number.

It is notable that the original P300 speller publication ([FD88]) reported an ITR of 12 bit/min. However, the it did not take the accuracy P into account, resulting in the calculation of a value maximally possible on the system used.

Cognitive switch ITR With 6 rows and 6 columns, N was set to 6. The number of *desired* selections thus consisted of the numbers of row selections, row escapes (leaving a wrongly selected row), answer errors and the number correctly selected items. For the *total* number of selections, the system and concentration errors and the wrong line selections were added to this number.

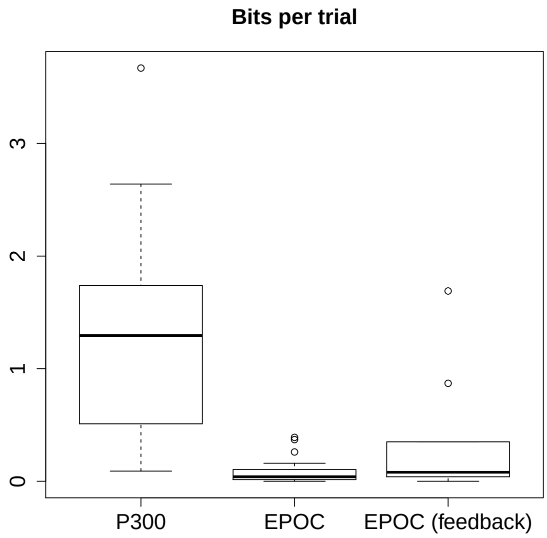


Figure 6.3: Comparison of B_{trial} by input method

Data presentation With a bit rate mean of 0.1, the cognitive switch without feedback can safely be denoted as unusable with this experimental set-up. The feedback variant achieves a higher bit rate mean of 0.343 bit per trial, however with $p = 0.1883$ this difference can not be regarded as significant. The difference of the cognitive switch feedback variant bit rates to the P300 bit rates (mean: 1.284 bit per trial, median: 1.295 bit per trial, confidence interval: $[0.98, 1.59]$) is significant ($p = 0.0004$). Since the ITR of the cognitive switch is still greater than zero, we reject the null hypothesis for H2. Using the Cognitive suite, communication is generally possible with a low bit rate. We can also reject the null hypothesis for H1 and state that communication using the P300 spelling paradigm is possible for non-impaired subjects using this device.

Discussion In contrast to the cognitive switch results, the P300 bit rates are in the usable range. They are similar to the lower average speeds of those reported in other studies. It should be pointed out that the experimental task design was probably influencing the resulting bit rates positively or negatively.

The outliers (maximum P300: 3.67 bit per trial (24t), maximum cognitive switch: 1.69 bit per trial (4t) predict the possibility of much better average information transfer rates on this device. It can be assumed that they could be achieved with extended user training or calibration, and in the copy spelling task that is commonly used for determining ITRs. For P300 input, probably the electrode placement on the subject's individual head shapes influenced the detectability of the P300 signal.

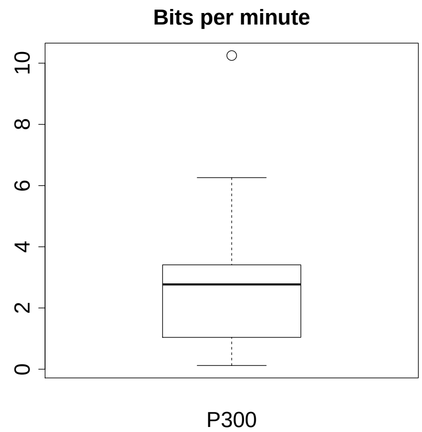


Figure 6.4: B_{min} for P300 spelling

This ITR of the P300 input is comparable to the lower average of results from studies with research devices (mean: 2.85 bit/min, median: 2.77 bit/min, confidence interval: $[2.0, 3.7]$). The fastest subject (26t) could achieve an ITR of 10.25 bit/min, which is the average performance of values attained with recent research P300 devices (refer to [EM09]). When calculating based on the original BCI2000 calibration suggestions (classification after first epoch with 100% accuracy, epochs = $epochs_{evaluation} - 3$), theoretically 3.83 bit / min on average (max. 16.5 bit / min). could have been possible.

Refer to table 6.1 for an overview of ITRs determined using the [WBM+00] definition. Keep in mind that most studies used the common copy spelling task, and that in this work an enhanced system was evaluated, so these values are not directly comparable.

Table 6.1: Exemplary information transfer rates (maximum values in parentheses)

publication	bit/trial	bit/min	BCI class
[NYY+09]	n/a	1.76	P300 spelling
[SKM+06]	n/a	5.25 (7.39)	P300 spelling
[GUM11]	n/a	14.9 (16.2)	P300 spelling
[TLB+10]	n/a	17	P300 spelling
[PK10]	1.4 (4.26)	3.0 (9.14)	P300 spelling (EPOC)
[LWE11]	0.173 (1.01)	0.65 (3.78)	input imagery (EPOC)
this work	1.28 (3.67)	2.85 (10.25)	P300 input
this work	0.1 (0.39)	n/a	cognitive switch
this work	0.343 (1.69)	n/a	cognitive switch (feedback)

The rates per minute for the cognitive switch variants were not determined since T_{sec} of a switch-scanning interface is varying too much. Using the half of the full scanning time was considered a too risky assumption. [LWE11] evaluated within the Cognitive suite whether the input imagination or neutral state on the EPOC could be activated or not. They set a time limit of 8 seconds and could therefore determine values for bit / min. Their maximum bit rate was achieved after 4 days of training in the Cognitive suite. The table represents the values from their standard method for evaluating two mental conditions. [PK10] used a fixed epochs number of 15 for a sample of 5 subjects.

6.2 Error frequencies

6.2.1 Total number of errors

Whenever an error occurred, the subjects self-assessed whether it was caused by themselves ('Konzentration' error), was a wrong solution ('Antwort' errors) or is unexplainable to them ('System' errors), as described earlier in section 5.2.4.

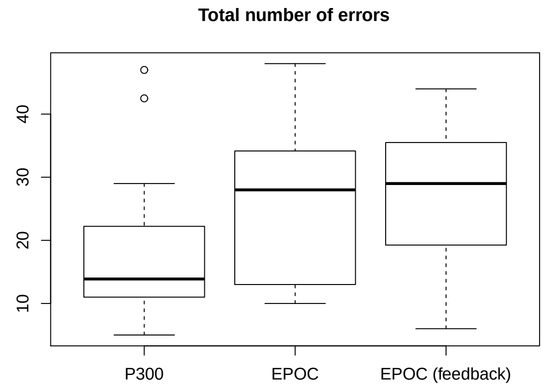


Figure 6.5: Comparison of Total number of errors by input method

Data presentation In P300 spelling, the subjects experienced less errors overall, reflecting their higher task completion rate with this input system (mean: 17.6, median: 13.9). In the cognitive switch cases, adding feedback did not result in a decrease of the high error rates (non-feedback mean: 26.1, non-feedback median: 28 ; feedback mean: 26.2, feedback median: 29). These findings will be discussed using the distribution of the error types.

6.2.2 Error types

The data representations in this section depict the total number of each error type during the individual evaluation tasks for the three system variants.

6.2.2.1 'Konzentration' errors

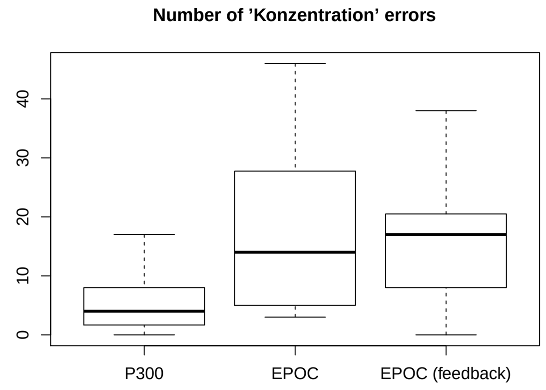


Figure 6.6: Comparison of concentration errors by input method

Data presentation The subjects attributed most errors to themselves during both cognitive switch variants (accumulated mean: 16.4, accumulated median: 15.9). During P300 spelling, users were less likely to regard their own focus as the source of the erroneous detection (mean: 4.9, median: 4). Input feedback slightly increased the number of concentration errors. It should be noted that the feedback variant (mean: 17, median: 15.9) also has a slightly increased total error rate in comparison to the cognitive switch without feedback (mean: 16.7, median: 14). The difference between the means of P300 input and cognitive switch input is significant ($p = 0.0001$).

Discussion In P300 spelling, input feedback is given at the very end of the stimulus presentations. Causes of concentration errors (e.g. fatigue, indifference towards stimuli, changes in reaction to stimuli) likely become unnoticed if there is no direct feedback of the current classifier results. Also, there might actually be errors in the classifier training. During the evaluation it could sometimes be noticed that certain cells in the matrix could not be selected with the trained classifier. There have recently be hints that the direction of the eye-gaze could be an influence the elicitation of P300 (refer to [BJB+10]).
To be able to use the cognitive switch, an active recall of the activation pattern is necessary. The high attribution of concentration errors shows that the subjects generally trusted the system detections, however had difficulties in recalling and withholding their imaginations. In the feedback variant, the visibility of the current classification result was one reason for the increased attribution of concentration errors. The subjects noticed when a false positive

arised. Additionally, it is possible that feedback influenced these error rates by unsettling the subjects. This was described by some participants and can be seen in the self-reported usage errors in section 6.3.4. We reject the null hypothesis for H7 since the users are indeed more likely to accredit errors to themselves during cognitive switch input.

6.2.2.2 'System' errors

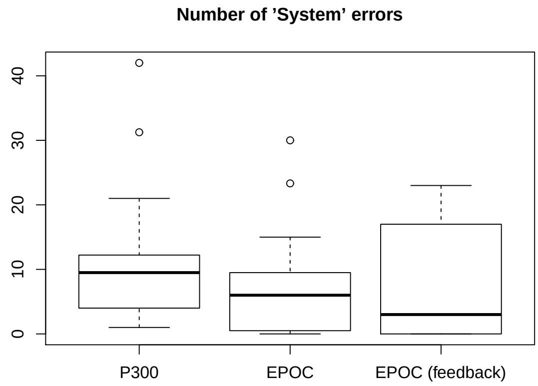


Figure 6.7: Comparison of system errors by input method

Data presentation In comparison to P300 input (mean: 10.46, median: 9.5), the subjects attributed less errors to the system in both cognitive switch variants (accumulated mean: 8.015, accumulated median: 6). System feedback (mean: 8, median: 3) reduced this error attribution slightly in our data sample in comparison to the variant without feedback (mean: 8.026, median: 6).

Discussion The attribution of system errors is reversing the concentration error distribution. The P300 input paradigm would not react immediately as the cognitive switch does, and does also not display intermediate classification results as the feedback variant. Some of the errors are attributable to a badly trained P300 classifier, to low signal quality or to the disadvantageously placed electrodes (actual causes within the system), while another part of them could also be caused by (unnoticed) increased fatigue or indifference towards the stimulus presentation.

6.2.2.3 'Antwort' errors

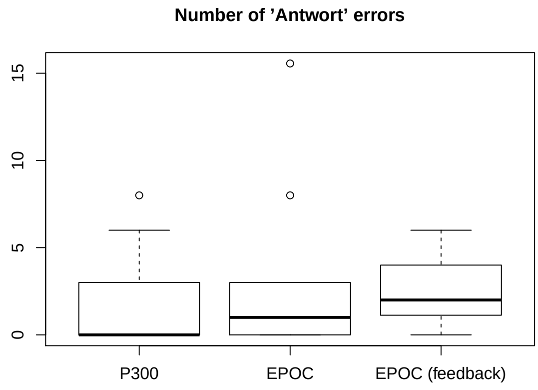


Figure 6.8: Comparison of answer errors by input method

Data presentation Wrong answers occurred most frequently in the cognitive switch task supported by feedback (mean: 2.6, median: 2), and least frequently in the P300 task (mean: 1.6, median: 0).

Discussion This distribution could be caused by decreased concentration and increased frustration during the cognitive switch tasks, which might result in less motivation for solving the language tasks. However, since this error type was very rare in general, such a conclusion would not be valid. The outlier in the non-feedback cognitive switch (12t, 16 answer errors) was deliberately aborting and restarting the scanning loop by selecting wrong answers.

6.2.3 Influence of matrix shuffling

A text entry system with low ITRs as in BCI would incorporate methods to predict what a user is about to spell, e.g. words or phrases. In such systems, the matrix would change the options after each input, forcing the user to re-read and re-orientate within the selectable set. To estimate the influence a constantly changing matrix on spelling accuracy, the matrix rows were shuffled in 50% of the tasks right when the BCI input started in this evaluation.

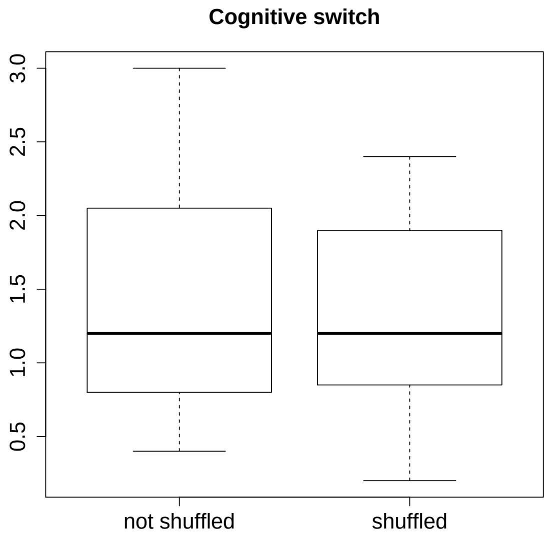


Figure 6.9: Influence of matrix shuffling on cognitive switch

During cognitive switch input, on average 1.41 errors per task occurred for the unshuffled and 1.3 for the shuffled cases. This result is negligible. Presumably the switch-scanning process provided enough time for re-orientation in the matrix.

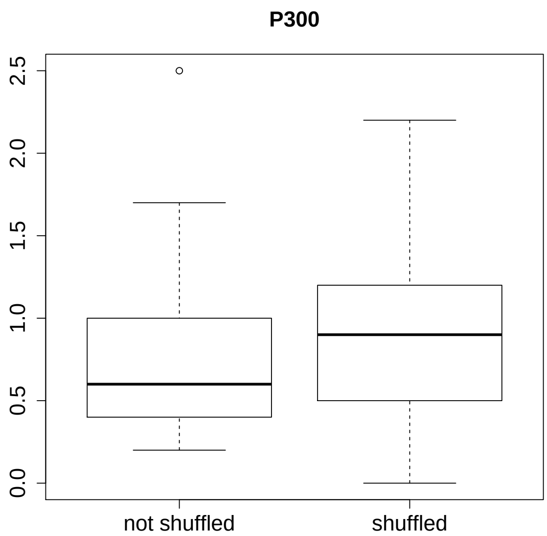


Figure 6.10: Influence of matrix shuffling on P300 input

The distribution for P300 input in figure 6.10 shows a slight difference in means, with 0.77 errors per task for the unshuffled and 0.83 for the shuffled cases. A possible explanation is that during P300 input, data for the on-line classification is already being acquired when the subjects were still searching for the solution in the new matrix. However, a t-test of the means shows that the difference of the means is not significant ($p = 0.381$).

Based on these findings, we fail to reject the null hypothesis for H6: On average there are more errors for P300 after shuffling the matrix in this evaluation, however these results are not significant.

6.3 Usability scales

The self-reported questionnaires were given to the subjects directly after completing the tasks for each evaluation system. As described ealier 4.3, the NASA TLX gives an impression on the workload the user felt during the evaluation tasks itself, while the modified version of the System Usability Scale tries to estimate how much acceptance the input paradigms systems receive during actual application situations. These questionnaires have been applied to BCI before in their original form by [PSG+11]. However, the report does not provide details for the individual scales.

Generally, the subjects were instructed to evaluate the method of choosing the answer (input method), not the evaluation task elements (the language tasks). In those cases where BCI illiteracy was attested for the evaluation situation, the users were not answering the corresponding questionnaires. However, they answered the questionnaires for the system that worked for them (mostly P300 input).

It is notable that most of the corresponding individual questions (performance and frustration in NASA TLX, "confidence" in the modified SUS) directly reflect the task performances given above. Thus, one can assume with more certainty that the user reports were accurate, and that the answers to the other items are comparably valid.

6.3.1 NASA Task Load Index

After each input method evaluation the subjects were first asked to answer the standardized NASA TLX. We advised them to review their experiences with the input method in the scales. We also told them that they should avoid trying to imagine being paralysed to rate this system, and to answer intuitively instead.

Since presenting a long questionnaire to subjects right after an exhausting evaluation task could decrease user motivation to answer it accurately, the shortened one-questionnaire version of NASA TLX was used. Also, according to literature [Har06], removing the weighting questionnaire from NASA TLX became common among researchers, commonly known as "Raw TLX". Refer to section 4.3.1 for more details on this.

6.3.1.1 Combined value

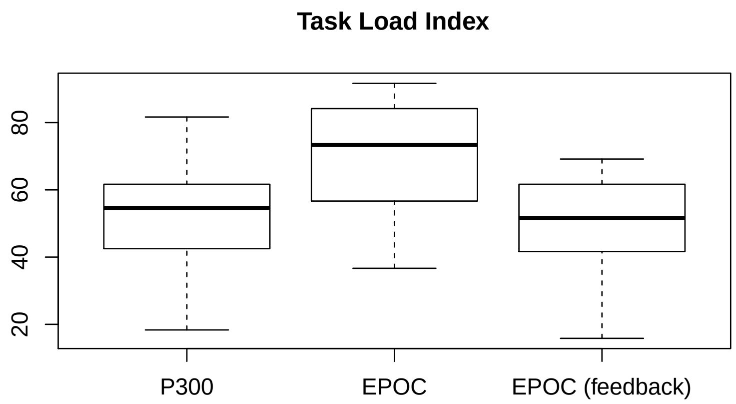


Figure 6.11: Comparison of NASA TLX by input method

The total workload is lowest for P300 input (mean: 51.6, median: 54.2, confidence interval $[46.5, 57.3]$). The EPOC without feedback is assessed as very demanding (mean: 70.1, median: 73.3, confidence interval $[60.54, 79.4]$), while the feedback variant's results (mean: 49.4, median: 51.7, confidence interval $[37.9, 60.9]$) are comparable to the workload of P300 spelling. The difference of the means between the cognitive switch variants is significant ($p = 0.006$). The difference between the feedback cognitive switch and P300 input is not significant ($p = 0.67$), while the difference between the non-feedback variant and P300 input is significant ($p = 0.0015$).

Discussion In comparison to [PSG+11], the TLX values for all input methods are very high. The TLX values for the P300 input are consistent with a similar experiment described in [PK10], where 5 subjects were evaluating the normal P300 speller implementation of BCI2000, resulting in an average TLX of 48.67. Surprisingly, the Task Load Index for the cognitive switch with feedback is on almost the same level as the one for P300 spelling. This combined value originates from different individual ratings, however, as can be seen from the next section. Based on the statistics about significance, we reject the null hypothesis for H3 when no feedback is provided. Users are likely to perceive different amounts of workload using these systems. However, we fail to reject the null hypothesis for H3 when the cognitive switch input with feedback is compared to P300 input. The workloads of these systems are likely to be similar.

6.3.1.2 Individual values

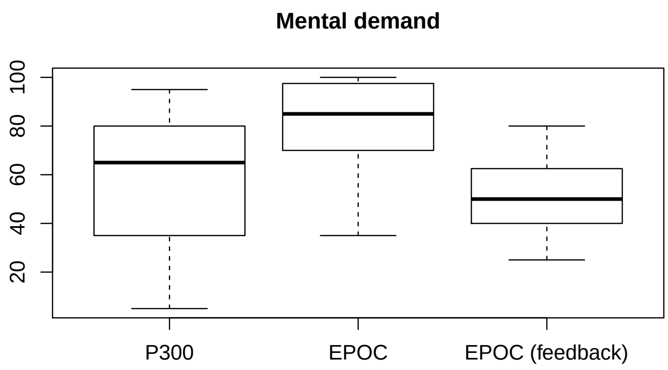


Figure 6.12: NASA TLX: Comparison of Mental demand

The subjects assigned high mental demand to all input systems. For our subjects it was lowest for the cognitive switch with feedback (mean: 51.4, median: 50.0). The cognitive switch without feedback reached the highest value (mean: 77.7, median: 85.0), declaring it as much more mentally demanding than the P300 input (mean: 57.0, median: 62.5, significant mean difference with $p = 0.0094$). The mean difference of the P300 input to the cognitive switch was not significant ($p = 0.376$).

Discussion The cognitive switch possibly received high ratings because it relies on repeatedly recalling a detailed input imagination. Although not requiring active modulation of brain activity, P300 spelling is still assessed as mentally demanding. The long continuous focusing periods the P300 input requires are the likely reason. It should be noted again that the subjects were asked to actively rejoice during the P300 input when their desired cell flashed, and to avoid indifference towards the stimulus presentation.

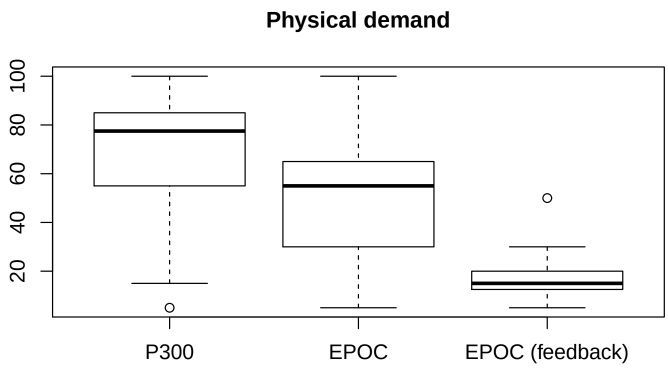


Figure 6.13: NASA TLX: Comparison of Physical demand

The P300 input was generally described as very physically demanding (mean: 68.3, median: 77.5). There was a large difference between the cognitive switch without feedback (mean: 51.3, median: 55.0) and with feedback (mean: 17.3, median: 15.0).

Discussion During P300 spelling input, subjects must sit still, suppress muscle movements and focus for a long period of time. During the calibration, they had to keep their eyes open for 45 seconds for each input letter. The eyes might become dry or start to water, resulting in actual physical pain. In comparison, the cognitive switch input requires attention only once in a fixed time frame. This is clearly reflected in the physical demand ratings. The high difference in the physical demand of the cognitive switch variants could be carryover-effects from the previous question.

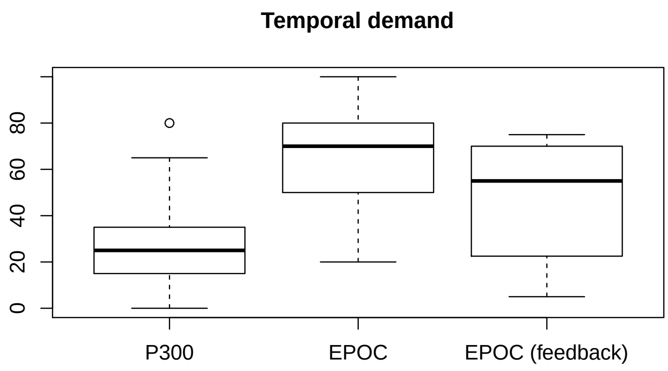


Figure 6.14: NASA TLX: Comparison of Temporal demand

Temporal demand was assessed low for P300 input (mean: 28.2, median: 25.0) and high for the cognitive switch input without feedback (mean: 65.0, median: 70.0). Feedback slightly lowered these values (mean: 46.4, median: 55.0), however the difference of the means is not significant ($p = 0.091$).

Discussion The cognitive switch paradigm required the subjects to actively react in a certain time slot. The subjects often reported that they could not recall their patterns in time in these systems (see section 6.3.4.2). The system-paced P300 input requires constant attention and is more forgiving to temporal errors, since the stimuli are presented multiple times.

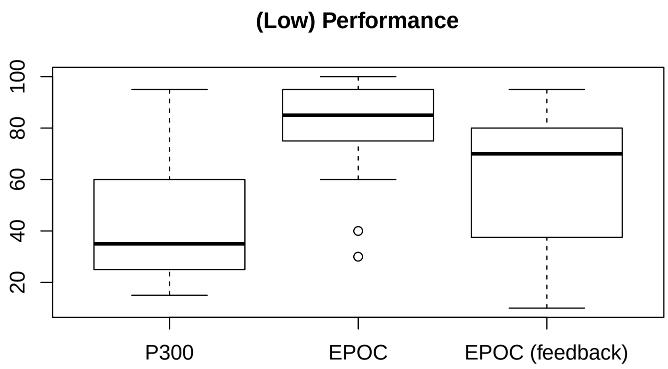


Figure 6.15: NASA TLX: Comparison of Performance

This item asked the subjects to assess to which extend they regarded their individual performance as a failure. P300 input reached lower results (mean: 43.5, median: 35.0), while they were very high for the cognitive switch without feedback (mean: 79.0, median: 85.0) and with feedback (mean: 60.0, median: 95.0). The difference between the cognitive switch variants is considerable, but not significant ($p = 0.091$).

Discussion The results of this item reflect the task performance and error ratings from the task performance measurements above. It also reflects the Frustration ratings, representing the self-experienced combination of these two.

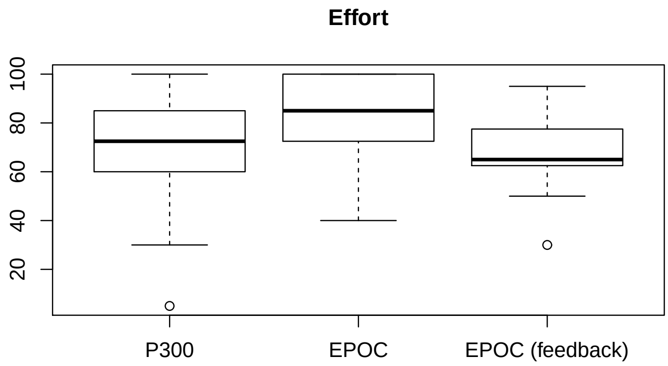


Figure 6.16: NASA TLX: Comparison of Effort

For this item, the users should estimate how much effort they needed to invest to achieve their results. The values for the different input methods are similarly high, with P300 input (mean: 68.0, median: 72.50) achieving slightly lower ratings than the cognitive switches (accumulated mean: 74.6, accumulated median: 75.0). Again, there are differences in the assessment of the non-feedback (mean: 80.0, median: 85.0) and feedback (mean: 67.7, median: 65.0) variants.

Discussion The cognitive switches both achieve the highest effort ratings and the lowest task performances. It can be concluded that in our evaluation situation it was very difficult for the subjects to achieve even a very low task completion rate. The P300 input required effort from the subjects as well, but they generally achieved much higher TCRs. feedback shows whether these accidental activations are currently approached. without feedback, the users are in uncertainty and might invest extra effort to avoid false positives.

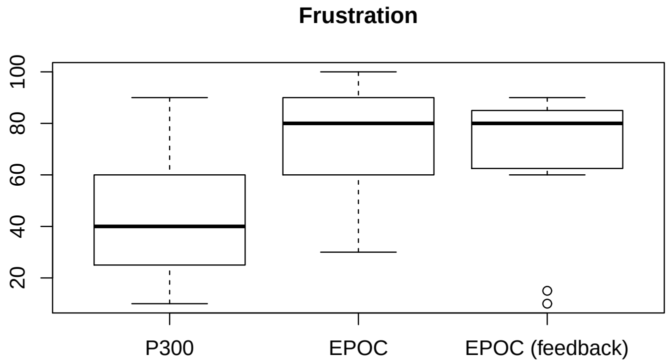


Figure 6.17: NASA TLX: Comparison of Frustration

This frustration rating reflects the error frequencies for both P300 input (mean: 44.7, median: 40.0) and the cognitive switch (accumulated mean: 69.2, accumulated median: 80.0). Providing feedback (mean: 71.3, median: 80.0) did not reduce the frustration of the cognitive switch input in comparison to the non-feedback variant (mean: 66.4, median: 80.0).

Discussion Every pattern activation creates an input for the cognitive switch method. Thus, false positives occur frequently, especially after the short training time in this evaluation. P300 input instead secures the input classification by presenting the stimuli multiple times.

6.3.2 System usability scale (modified)

Apart from the combined SUS metric, the answers to the individual statements are represented in histograms. This is more suitable to depict the distribution of answers on an ordinal scale. The results are shortly discussed after each item.

6.3.2.1 Combined value

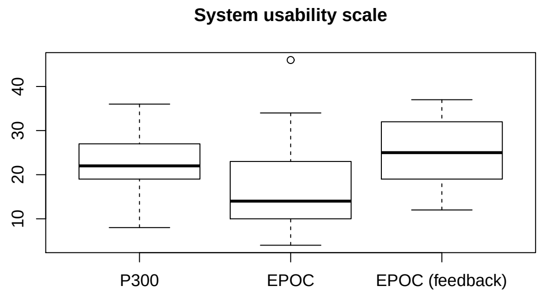


Figure 6.18: Comparison of modified SUS by input method

In total the SUS value of the P300 input (mean: 23.3, median: 22.0, confidence interval $[20.5, 26.0]$) was slightly higher than that of the cognitive switch variants (accumulated mean: 20.9, accumulated median: 20.5, confidence interval $[16.6, 25.2]$). This difference of the means is not significant, however ($p = 0.34$). The difference of the means between P300 input and cognitive switch does also not become significant if cognitive switch without ($p = 0.105$) and with feedback ($p = 0.57$) are regarded individually.

Feedback (mean: 25.0, median: 25.0, confidence interval $[19.1, 30.9]$) improved this score in comparison to providing no feedback (mean: 17.9, median: 14.0, confidence interval $[11.7, 24.1]$), however the difference was not significant ($p = 0.081$). In our data sample, the feedback variant is rated slightly more usable than P300 input.

Discussion As in TLX, it should be considered that the similarity of the values (i.e., for P300 input and the feedback variant of the cognitive switch) results from different answer distributions of the individual statements. However, it is notable that the task performance differences between P300 input and the feedback cognitive switch did not strongly influence the impression of the general usability of these input methods, at least based on this questionnaire.

Based on the findings, we fail to reject the null hypothesis for H4, meaning that there is no significant difference for the total SUS values of P300 input and cognitive switch (both variants).

6.3.2.2 Distribution of individual statements

For each individual item, the first two histograms compare P300 input to the accumulated answers of the cognitive switches. The second pair divides the cognitive switch evaluation data into the non-feedback and feedback variants. When comparing these last two, please keep in mind that the number of subjects in these groups was not equal (refert to section 5.2.2).

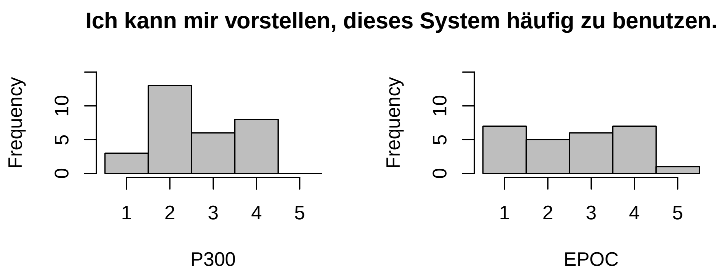


Figure 6.19: Modified SUS: Item 1 by input method

Item 1: "I think that I would like to use this system frequently." As for P300 spelling, one group (n=16, rating<3) assessed the system as unusable, while the other group (n=13 for rating≥3) could imagine using it often. The accumulated answers from the cognitive switch systems show a uniform distribution. However, dividing them into non-feedback and feedback variants shows that the subjects could rather imagine using the feedback variant for extended periods of time, while the non-feedback system is considered less usable.

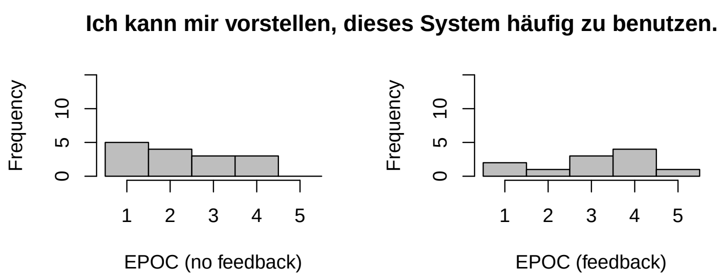


Figure 6.20: Modified SUS: Influence of feedback on item 1

Discussion Some of the users considered provided feedback as helpful. However, it is obvious that most subjects could imagine none of the systems for long-term usage. It should be noted that before answering to this item we did not ask the subjects to imagine being motor impaired, but to estimate whether communicating with this input method was possible at all.

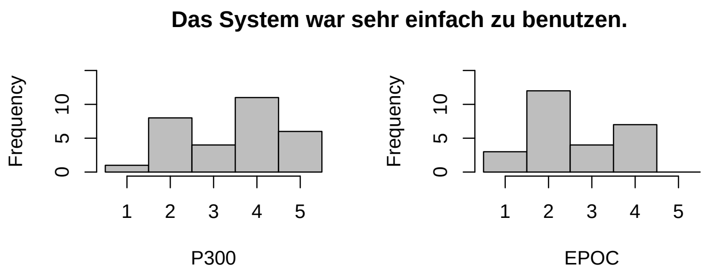


Figure 6.21: Modified SUS: Item 2 by input method

Item 2: "I thought the system was easy to use." P300 input is clearly considered less difficult (n=18 for rating≥4) than the cognitive switches (n=8 for rating≥4). By adding feedback to the cognitive switch, the task appears only slightly easier.

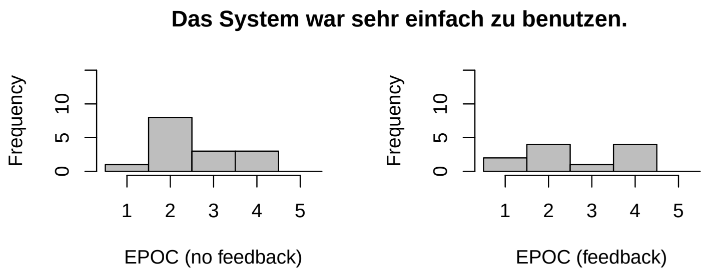


Figure 6.22: Modified SUS: Influence of feedback on item 2

Discussion Probably due to it being based on a physical reaction, the P300 spelling paradigm was not considered difficult by most subjects. Both variants of the cognitive switch equal in their answer distributions. It is notable that a small part of the subjects (rating=4, n=7) considered the cognitive input as simple. Two subjects in this group (mean=0.45) achieved a TCR of 100% (21t and 4t).

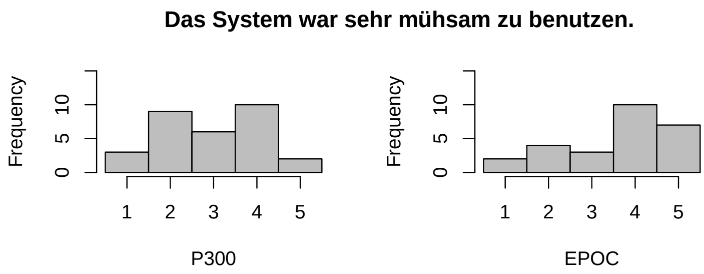


Figure 6.23: Modified SUS: Item 3 by input method

Item 3: "I found the system very cumbersome to use." For P300 input, again two groups can be seen: One considered the input method as cumbersome (n=12 for rating<3), while the second group disagreed to this assessment (n=12 for rating>3). These two groups can also be seen in the feedback variant of the cognitive input, while the non-feedback system was generally considered more cumbersome.

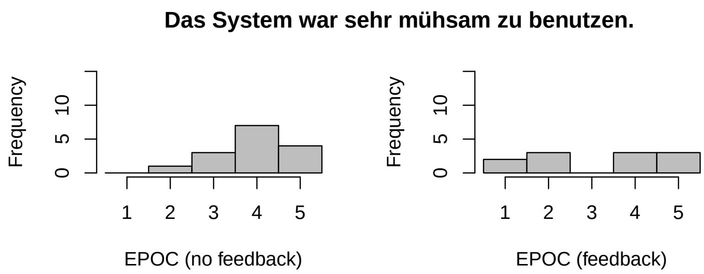


Figure 6.24: Modified SUS: Influence of feedback on item 3

Discussion One group probably assessed the physical effort of the P300 spelling as more significant, resulting in a higher agreement to this item. The cognitive switch was partly experienced considerably less cumbersome if input feedback was provided, although recalling the input imagination still required the same mental effort.

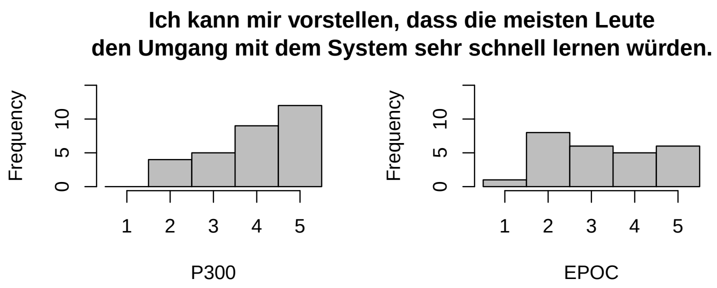


Figure 6.25: Modified SUS: Item 4 by input method

Item 4: "I would imagine that most people would learn to use this system very quickly." 26 subjects considered the P300 input as fast to learn (rating$\geq$3), while the cognitive switch has a more uniform distribution with a tendency to the disagreeing dimension of the scale. More subjects agreed to this statement when feedback was provided for the cognitive switch.

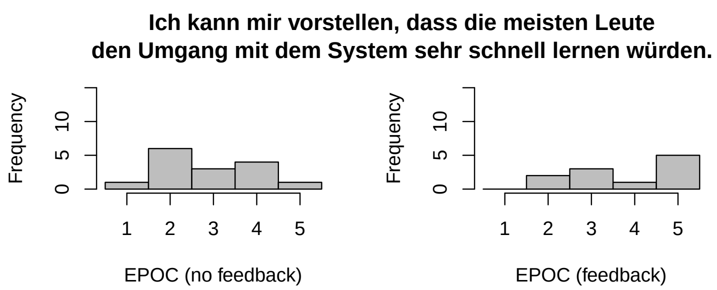


Figure 6.26: Modified SUS: Influence of feedback on item 4

Discussion To use the cognitive switch, users have to learn how to actively modulate their brain activity by repeatedly and exactly recalling an activation pattern. Also, the cognitive switch possibly required them to learn how to actively avoid false positives, which might be caused by the short training time. Again, the P300 input being mainly based on a physical reaction is possibly the cause for this result. The learning time of this system is therefore considerably reduced.

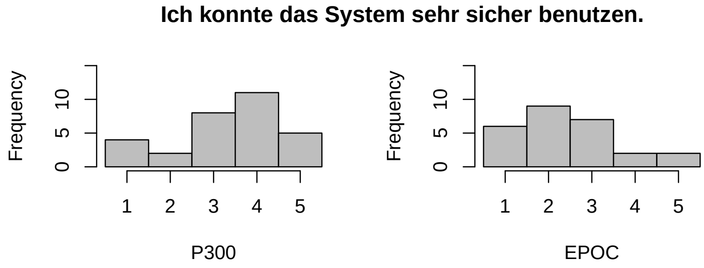


Figure 6.27: Modified SUS: Item 5 by input method

Item 5: "I felt very confident using the system." Subjects agreed to feel more, however not absolutely confident using the P300 input for the selection of answers. The distribution is reversed for the cognitive switch.

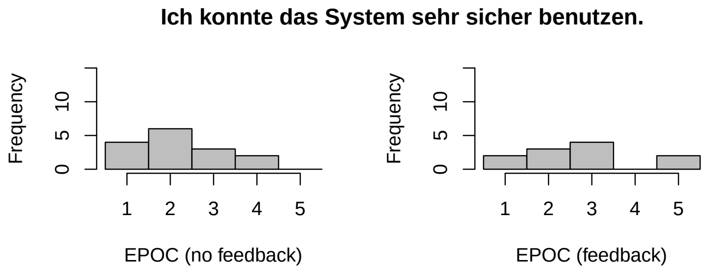


Figure 6.28: Modified SUS: Influence of feedback on item 5

Discussion The original version used "felt confident" in this SUS statement. The German "sicher benutzbar" means that both the numbers of false positives and false negatives would be low, which would ensure a reliable form of communication. Again, this item therefore reflects the error and task completion rates of the individual systems, with the non-feedback cognitive switch being the least reliable system of the three.

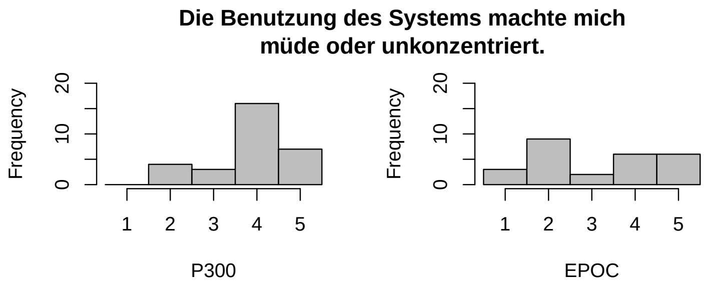


Figure 6.29: Modified SUS: Item 6 by input method

Item 6: "Using the system made me feel tired or unfocused." Most subjects (n=23 for rating>3) assessed the P300 input as exhausting. For the cognitive switches again two groups can be seen, with one of them with low (n=12 for rating<3) and one with high agreement (n=14 for rating>3) to this statement.

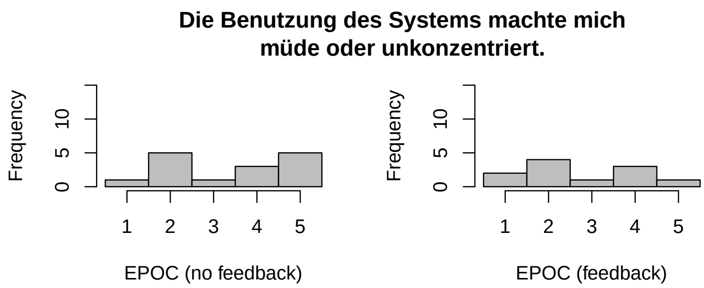


Figure 6.30: Modified SUS: Influence of feedback on item 6

Discussion This item reflects the physical and mental effort required for using the systems. The necessary constant focus required for P300 input probably was the highest influence. Adding feedback to the cognitive switch obviously also made this task appear less strenuous for the subjects.

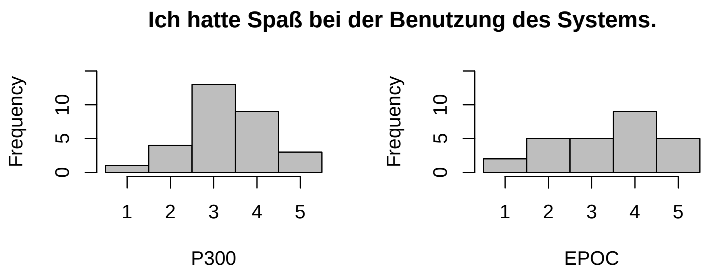


Figure 6.31: Modified SUS: Item 7 by input method

Item 7: "I had fun using the system." After the cognitive switch input the subjects were more likely to completely agree to this statement (n=14 for rating>3). When assessing the P300 input, subjects were most likely to answer in the mid range (n=13 for rating=3).

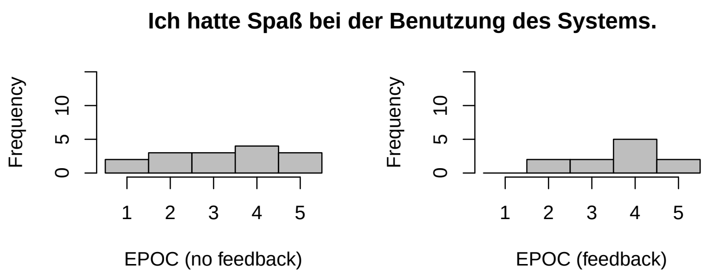


Figure 6.32: Modified SUS: Influence of feedback on item 7

Discussion Despite the frustrating usage of the systems, obviously most subjects were rather enjoying the input process during the evaluation. This (new) item should be regarded critically, though, because it probably rather reflects the subject's curiosity and the novelty of the situation.

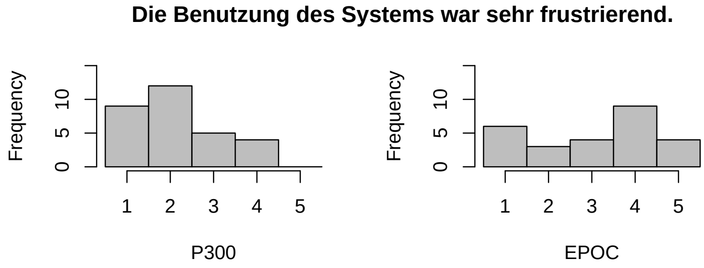


Figure 6.33: Modified SUS: Item 8 by input method

Item 8: "Using the system was very frustrating." After the P300 task, the subjects – in favour of this system – generally did not agree to this statement (n=21 for rating<3). The answer distribution of the cognitive switch was rather uniform, with a tendency for high agreement, however.

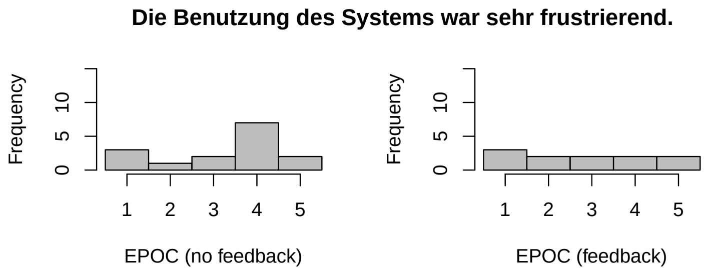


Figure 6.34: Modified SUS: Influence of feedback on item 8

Discussion This item reflects (and validates) the frustration rating of the NASA TLX, and therefore also the error and task completion rates. The subjects again stated that they estimated the cognitive switch input as far more frustrating than the P300 input.

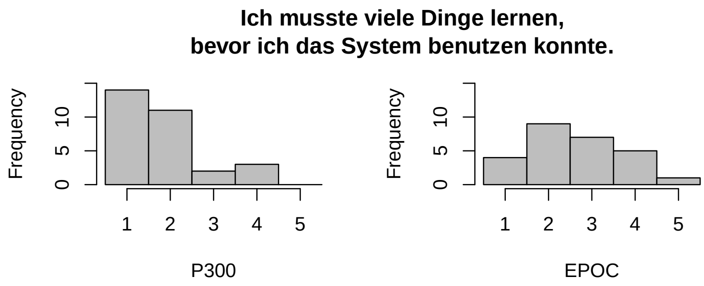


Figure 6.35: Modified SUS: Item 9 by input method

Item 9: "I needed to learn a lot of things before I could get going with this system." Most subjects disagreed to this statement both for the cognitive switch and for the P300 spelling input. The distribution of the non-feedback and feedback condition can be called almost equal.

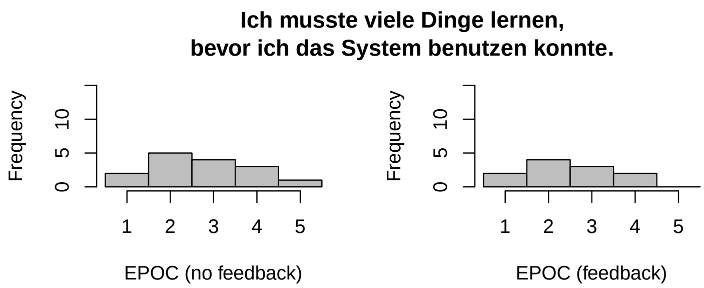


Figure 6.36: Modified SUS: Influence of feedback on item 9

Discussion This is similar to the positive statement about the learnability above, but not directly reverse. For both systems, most users did not think that there was particularly much to learn before being able to use them.

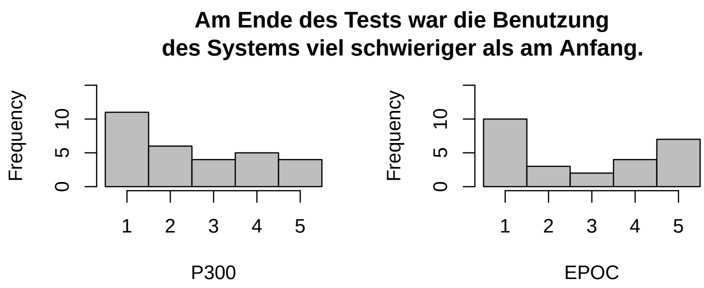


Figure 6.37: Modified SUS: Item 10 by input method

Item 10: "At the end of the test using this system was much more difficult than at the beginning." Again, the sample is divided in an agreeing (n=17 for rating<3) and a disagreeing (n=12 for rating>3) group for the cognitive switch input. For the P300 spelling paradigm, many subjects (n=17 for rating<3) would rather disagree, however not all of them did so (n=9 for rating>3).

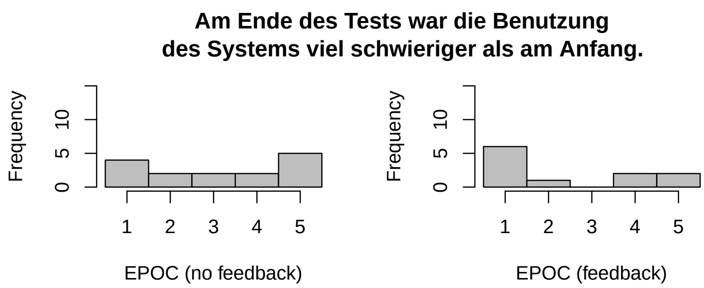


Figure 6.38: Modified SUS: Influence of feedback on item 10

Discussion This item reflects how strongly the repeated usage of the system influenced the perceived difficulty. Both the higher mental workload of the cognitive switch input and the physical effort of the P300 input probably made many subjects agree to this statement.

6.3.3 Preferred system

In the last question the subjects should state which system they would prefer. They were also asked to shortly describe the reasons for their decision.

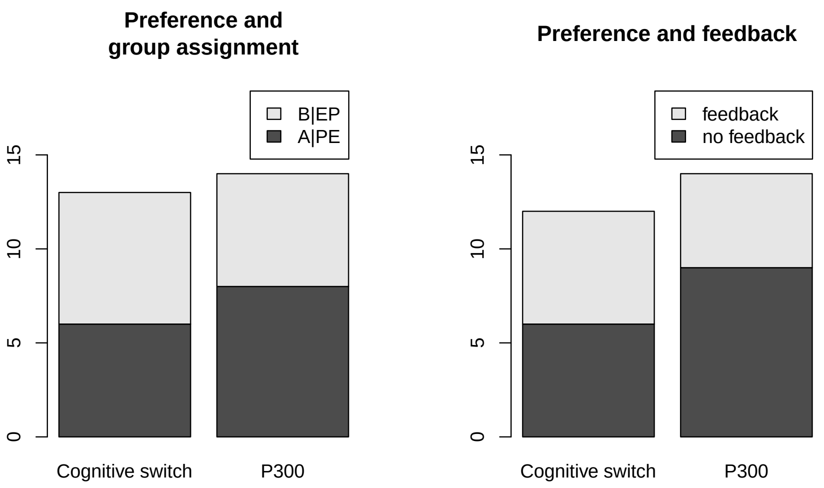


Figure 6.39: Preferred system and group assignment

As seen from figure 6.39, this question resulted in an almost equal distribution of the preferences. The five persons that were categorized as BCI-illiterate are not included in this figure. The group assignment had no influence on the subjects' choices, however, if no feedback was provided the subjects were more likely to prefer P300 input. This distribution could be considered as unexpected regarding the low TCR and high frustration level of both cognitive switch variants.

In the questionnaire the subjects were asked to give reasons for their choice. Refer to Appendix C.1 for the transcribed answers. Figure 6.40 below attempts to identify and categorize similar statements. Please keep in mind that such a categorization is prone to subjectivity.

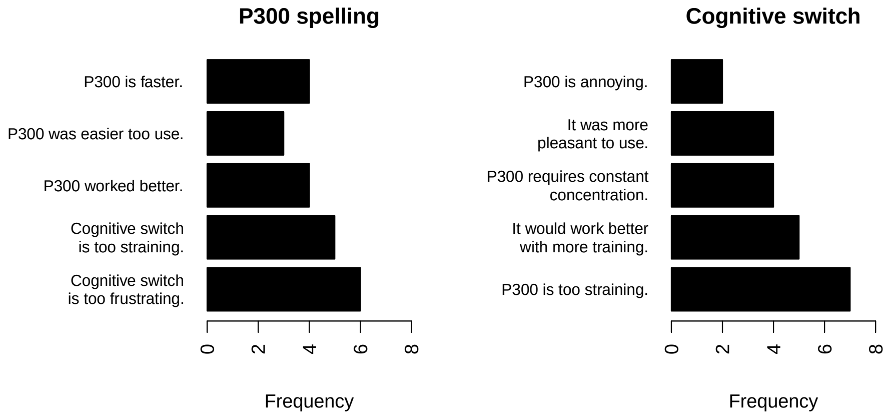


Figure 6.40: Rationale for system preference

The given reasons (refer to figure 6.40) reflect the questionnaire parameters from the usability scales: The physical demand is the main cause for rejecting P300 spelling. The subjects often reasoned the preference of the cognitive switch by assuming that it would work better if there was more time to practice. The cognitive switch input was mainly rejected due to the frustration the difficult input caused, and due to the mental demand it required.
Other notable, but infrequently given reasons were:

Cognitive switch
The user has more control over the cognitive switch input.
In the long run, by training more activation patterns the cognitive switch has a larger input alphabet and is therefore the more powerful input system.

P300 input
P300 input appears to be more exact.

One subject gave a conditional preference: In his opinion the P300 input is more suitable for short texts, while the cognitive switch should be used for longer writing.

6.3.4 Self-reported error types

After completing each evaluation system, the subjects were asked to select the most frequently occurring errors from a list of possibilities in the final questionnaire (see Appendix B.4). The available items do not have a theoretical foundation. They were arbitrarily derived from advice for ERP and BCI experiments, experiences with the system after the first two evaluations, and

from personal experience. They are also not directly comparable, since items from the different sets do not reflect each other. In future research it might be advisable to ask subjects for their own experience with system errors in descriptive questions.
The results should be regarded as the distribution of errors the subjects were most likely to select from the given choices. The distribution gives an impression of possible user interaction improvements.

6.3.4.1 P300 error types

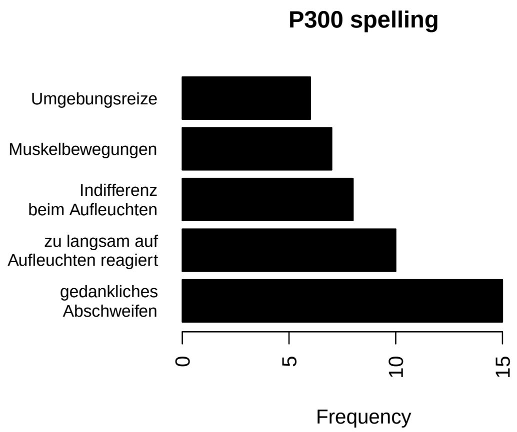


Figure 6.41: Self-reported error types – P300 spelling

The results show that all of the available error types frequently occurred during P300 input, with "Gedankliches Abschweifen" (mind wandering off) being the most frequently selected, and "zu langsam reagiert" (reaction to stimuli too slow) the second-most frequent. One frequently selected type was "Indifferenz beim Aufleuchten", which means perceiving the stimulus, but not being attentive in a way that would create the P300 ERP (meaning, being indifferent towards it).

Discussion Input using the P300 paradigm requires maintaining the same high level of concentration for an extended period of time. Also, a user is required to react uniformly and frequently to stimuli, in disregard of naturally decreasing concentration. With extended length of the input stimulus presentation, thoughts are also likely to wander off. The three most frequently selected error types base on these behaviours and requirements.
A reason for the high frequency of selections of "zu langsam auf Aufleuchten

reagiert" could be that the given P300 paradigm does not prevent the occurrence of double stimuli, that are difficult to react to. As for "Umgebungsreize" (environmental stimuli), it should be noted again that they were constant during the evaluation, to the extend that one of them (the sound of a closing door) was created by intention if it did not occur because there were people entering the room.

6.3.4.2 Cognitive switch error types

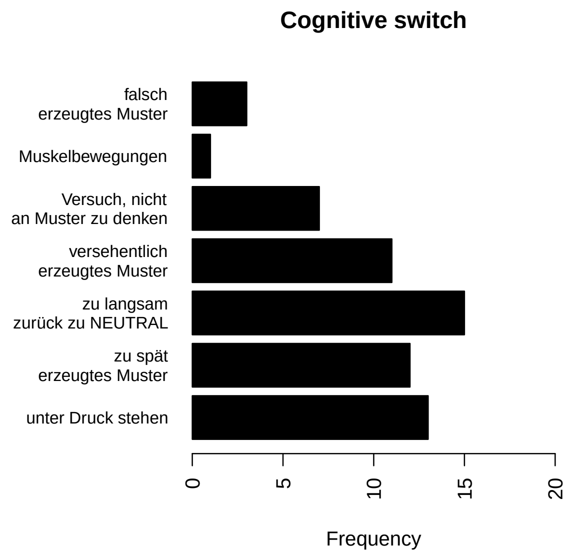


Figure 6.42: Self-reported error types – cognitive switch

The results reflect that obviously timing created the most prominent errors during the cognitive switch input. The release time ("zu langsam zurück zu NEUTRAL") and the switch activation time ("zu spät erzeugtes Muster") were the most frequently selected timing problems. This was a result of many false positives occurring right after a row was selected, when the activation pattern was still being detected by the system. The release time was then too short for the subject to return to a neutral state. Also, the subjects often attributed errors to their ability to recall and prevent the activation pattern. Mental pressure ("unter Druck stehen") and randomly occurring false positives ("versehentlich erzeugtes Muster") were often selected, too.

Discussion The distribution shows that false positives occurring just after selecting a row were highly prominent errors for the subjects during the evalu-

ation. Both these and activation patterns that were recalled too late could be regarded in attempts of user interaction improvements of the cognitive switch. Other types of false positives could possibly be addressed by other changes in the interaction system, e.g. an input verification system to prevent accidental activations from happening.
The subjects experienced that mental pressure highly influenced their ability to recall their activation pattern. During the cognitive switch paradigm, the users only have a limited time period for activation. In case they do not succeed to recall it, they would have to wait for the scanning cursor to return to the desired option again. The self-reported error results reflect this behaviour. Muscle movements do have a high influence on the EEG data and the success of pattern activation. However, they were not regarded as a severe problem by the subjects. A reason could lie in the fact that users do not have to suppress them for an extended period of time (as in P300 input).

6.4 Midas touch problem

Please refer to 4.2.2 for a description of the Midas touch problem during asynchronous BCI interaction. As mentioned before, one reason for including the language tasks was to examine whether this type of false positive input would occur during concentrated thinking periods. Figure 6.43 depicts the occurrence of false positives during the language quizzes and their relation to the TCR and total errors.

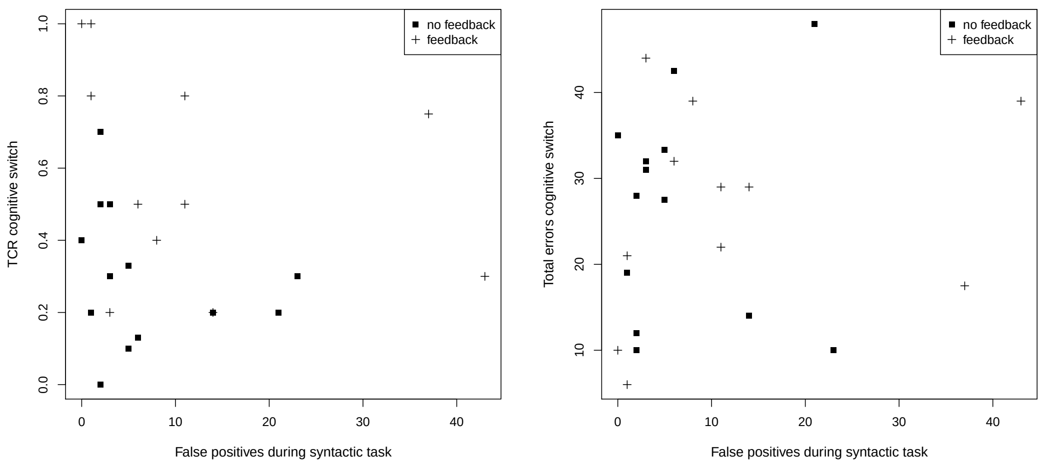


Figure 6.43: False positives during syntactic task and TCR / Total errors

The total number of false positives during the syntactic tasks has a mean of 7.4 and a median of 3.0. There were only 2 subjects who created no false positives during the syntactic tasks. This clearly indicates that the Midas touch problem would have created severe usability problems if the interface would have been active during this the syntactic task. It is notable that the two subjects with the highest task completion rate (4t and 21t) had a low number of false positives (1 resp. 0) during the syntactic task.

The distributions in the figure 6.43 were tested for correlations: However, the small correlations between the number of syntactic task false positives and TCR ($\rho = 0.013$, $p = 0.95$) and syntactic task false positives and total errors ($\rho = -0.054$, $p = 0.8$) were not significant. The Midas touch problem therefore occurs for all subjects regardless of their TCR, and we fail to reject the null hypothesis for H5.
It should be kept in mind that this evaluation is based on a very short training period. It would be necessary to determine to which extend this problem occurs if the users are well trained before predicting these occurrences in a real application situation. During extended usage periods, users of the system will not tolerate even a low number of regularly occurring false positives.

6.5 P300 data properties

In this section, notable properties of the P300 signal are presented additionally to the evaluation data. They can serve as a foundation to assess the capabilities of the EPOC device.

6.5.1 P300 epochs and correlations

As described earlier 5.2.1.1, the number of stimulus intensifications (or "epochs") during the P300 input was adjusted to the user's performances during the calibration. The chosen number is the one necessary for the classifier to detect an input with 100% accuracy three times in a row. The distribution of epochs can be seen in figure 6.44. The bin size of this histogram is set to 3:

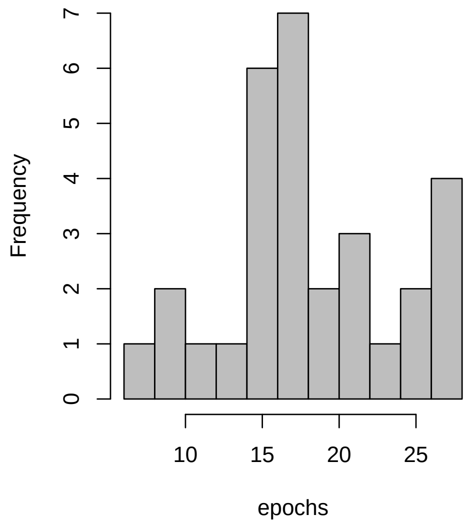


Figure 6.44: Distribution of epochs necessary for input

Discussion The number of intensifications balances two effects: On the one hand, the P300 classifier will need a certain amount of user data to clearly capture the current shape of the signal and make a prediction. On the other hand, while a large amount of averaged signal data will help the online input classification, the user will become tired when concentrating on long periods of stimulus presentations. At the extreme values of the distribution, the subjects had to concentrate for 13 seconds to make an input if they needed only 7 epochs, and for 47 seconds if they needed 25 epochs. In the latter case it is likely that increased fatigue or indifference blurred the online signal data, raising the probability of errors. Note that those subjects who needed fewer intensifications also had less errors and higher task completion rates, as can be seen from figures 6.45.

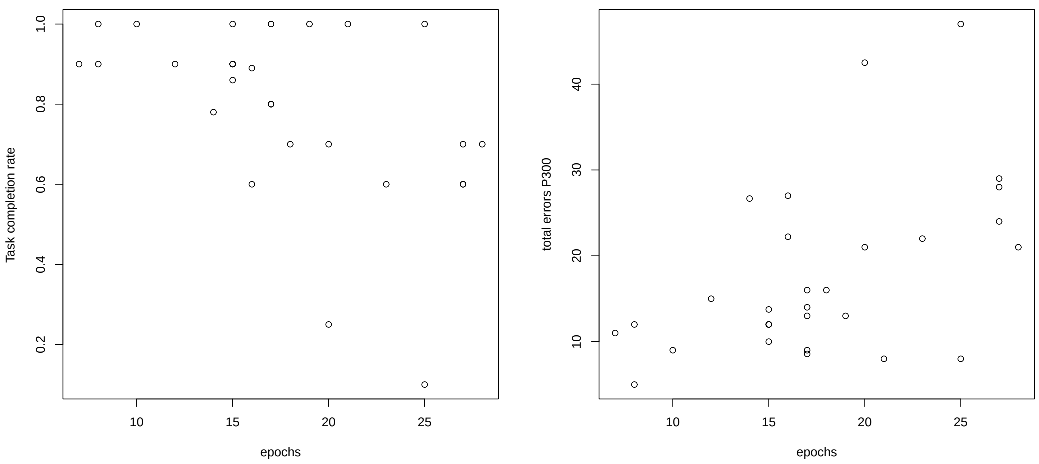


Figure 6.45: Relation of epochs and TCR / Total errors

Spearman's rank correlation coefficient shows a negative correlation between the necessary epochs and the TCR ($\rho = -0.497$, strong correlation according to Cohen's recommendations, statistically significant with $p = 0.0052$), and a positive correlation between the epochs and the total number of errors ($\rho = 0.46$, strong correlation, significant with $p = 0.0105$). Another statistically significant ($p = 0.0002$) and strong negative correlation ($\rho = -0.63$) exists between the number of epochs and the overall SUS rating. We can conclude that a low number of necessary epochs made the P300 spelling less error-prone and increased user acceptance in this evaluation. However, it should be noted that high individual ability for using this type of system decreased the necessary number of epochs, which would be a logical explanation for higher SUS ratings.

6.5.2 P300 signal on the EPOC

The data visualizations in this section are derived from the P300 calibration application of BCI2000. They give an impression of the data acquisition properties and the ERP capabilities of the EPOC device. Since the signal acquisition capabilities of the device should demonstrated, results of well-performing subjects with 100% TCRs and a clearly visible signal were chosen. The few low performing subjects had high amounts of visible noise in the r^2 matrices.

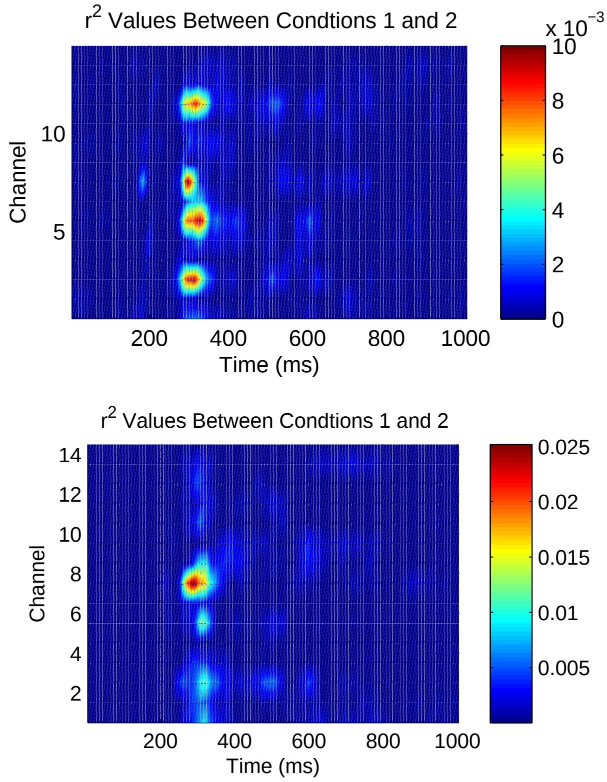


Figure 6.46: P300 $\mathbf{r^2}$ matrix on the EPOC for two subjects (00t and 25t).

The $\mathbf{r^2}$ matrices in figure 6.46 highlight the most relevant features of the ERPs for the separation of the potentials that were caused by the frequent or by the rare stimuli. The vertical axis represents the 14 channels of the EEG device. The horizontal axis shows the time that passed since a rare stimulus occurred. The rare stimuli are the current letters the subject had to spell. They can also be denoted as “Attended stimuli”, where a P300 wave is created due to the subject focusing on the desired letter. From these visible elicitations around 300ms after these stimuli it becomes evident that a P300 signal can be detected using the signal quality and electrode arrangement of the EPOC device.

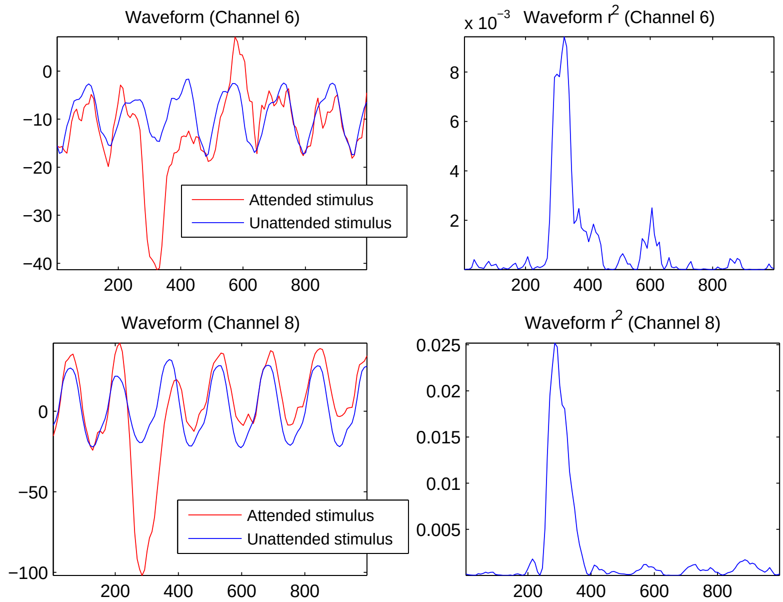


Figure 6.47: P300 amplitude waveforms and r^2 time course on the EPOC for two subjects (00t and 25t).

Figure 6.47 depicts the exact waveforms of these elicitations, based on the amplitudes (left) and the r^2 values (right) of the calibration process. In the waveform of 00t (upper figure), other ERP components can be seen very clearly, while the waveform of 25t achieved high amplitudes. 25t was one of the fastest subjects during the evaluation.

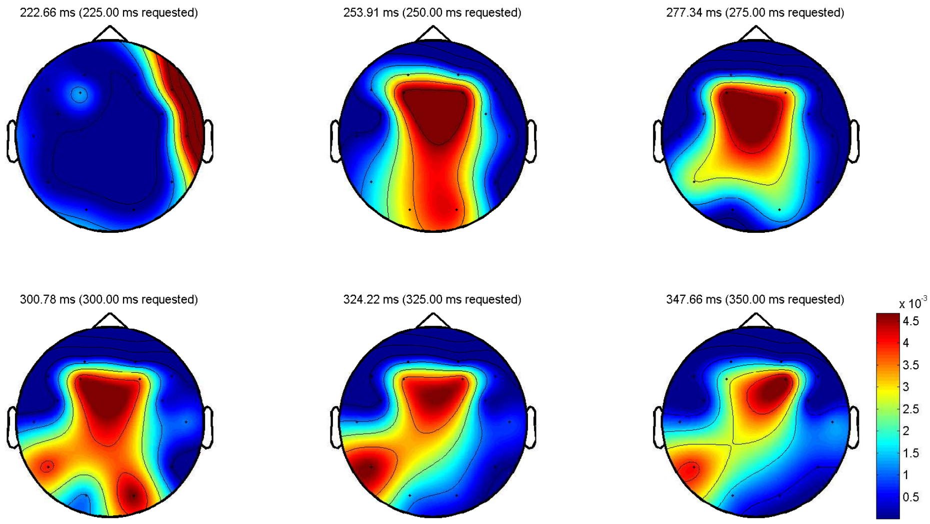


Figure 6.48: P300 topographic plots and time behaviour on the EPOC for subject 00t.

Figure 6.48 represents the scalp distribution of the r^2 values. The EPOC electrodes are marked in this plot. The electrodes of the device are not evenly distributed, and almost all electrodes around the midline of the 10-20 System are missing. Since the BCI2000 application interpolates between the channels to create a continuous bidimensional map, these graphs should be interpreted with caution.

The plot outcome in the figure was an exception during the evaluation: For most subjects, only the electrodes O1 and O2 at the occipital lobe captured sufficient data for P300 detection. This resulted in a plot where the P300 signal appeared to be elicited around the inion at the back of the head.

Chapter 7

Conclusions and future research

There are 230.000 new patients diagnosed with PVS all over Europe each year. Up to 40% of these could actually be in the completely conscious “total” locked-in condition (refer to [AMML96]). Any way to enable them to communicate would dramatically increase their life quality and autonomy. However, rehabilitation centres often can neither afford nor set-up their only remaining way to interact with their surroundings, which is BCI technology. The affordable and easy-to-use EEG technology that has recently entered the consumer electronics market might be capable of closing this gap.

The objective of this thesis was to evaluate the performance and user acceptance of two non-invasive BCI input methods on a recent consumer EEG. A sample of 30 healthy subjects was participating in this evaluation, allowing statements about statistical significance. Based on the results, a suggestion about the paradigm that should be investigated further with paralysed patients can be made. Also, more specifically, it allows estimations about which of the two variants should be obtained and on how much additional implementation work needs to be done.

This thesis represents the first work approaching the area of BCI at the Bauhaus-University of Weimar. Thus, the first period of the thesis working time was spend for a literature review on BCI, EEG, experimentation methodology and usability issues in the available interaction and communication paradigms. Before starting this work, the consumer EEG was selected, acquired and its capabilities investigated. During the initial working time there was also the idea to use the EPOC for a VJ system, based on on-line analysis of EEG patterns during music perception. Another initial direction of the work was developing a simple valence-arousal space on-line classifier based on the EPOC. Based on this, a research question was developed. Approaches such as setting up and evaluating a fully-functional system with phrase prediction were declined due to a lack of comparability and existing commercial communication solutions.

We decided to compare two existing text input solutions in an initial evaluation with short training times: The cognitive switch paradigm is used in the existing EPOC keyboard solutions, while the P300 paradigm represents an input system that has been well investigated in research for decades. System limits where comparisons are possible were developed and implemented mostly identically for both system types. Hypotheses were derived from literature and could mostly be confirmed with the evaluation data.

The results, both the quantitative and the descriptive data can provide a descriptive basis for the estimation of performance and user acceptance using the consumer EEG device. Main problems of both input variants are revealed, which provide a basis for future developments and usability research on the investigated input paradigms. Additional data, such as hints about which types of imaginations work best were collected from the subjects. The data can also be transferred to other selection-based user interfaces.

7.1 Findings and recommendations

The evaluation has shown that, while users generally estimate it as physically demanding, P300 input should be the preferred system during further evaluations (e.g. with a sample of paralysed participants). Since it relies on an almost automatic physical reaction, it is more stable and requires no mentally strenuous and error-prone user contribution. Also due to its short training time, it might be less straining for motor-impaired people. P300 input achieved significantly higher task completion rates, a higher bitrate, and similar to lower TLX and SUS ratings. A bitrate of 10.25 bit/min has been achieved by one subject although electrode positions and signal quality are suboptimal for P300 detection. The number of errors was also not vulnerable to re-orientation in the input matrix, which makes the P300 paradigm a candidate for effective predictive text entry systems.

The classifying system of the EPOC device is a black box, so no practical advice on the training procedure can be derived from previous research into imagery-based input. The resulting explorative form of calibration can therefore turn out to be strenuous, especially for patients, and might not work at all due to difficult-to-detect BCI illiteracy. Based on the results, it is certain that the existing keyboard solutions based on the Cognitive suite input are not usable with short training times. Also, they are likely to remain highly error-prone due to the Midas Touch problem and false positives/negatives with prolonged training time. The Midas Touch problem occurred for all participants regardless of their TCR.

However, due to less physical strain and no demand of constant concentration, the cognitive switch paradigm should be re-evaluated with addressed usability problems, i.e. the Midas touch problem and the high number of false positives. Also, the subjects equally preferred the cognitive switch and the P300 input, although the latter worked much better during the evaluation. It is worth investigating how the preferences turn out if there is more time for training the cognitive switch classifier.

The findings can be summarized to the following insights:

The Emotiv EPOC is a suitable for text selections using P300 input with bit rates at the lower average range of research grade devices. The average ITR of the subjects was 1.28 bit/trial (max. 3.62 bit/trial) and 2.85 bit/min. This corresponded to 2.1 selections per minute on average (max. 4.6). With 10.25 bit/min, the best performing subject in this study achieved an ITR around the average of recent research-grade devices (refer to [EM09]). When calculating based on the original BCI2000 calibration suggestions, theoretically 3.83 bit/min on average (max. 16.5 bit/min) could have been possible, corresponding to 2.7 (8.1) selection per minute.

With the switch-scanning input variant of the available software keyboards, a text selection bit rate of up to 1.69 bit/trial is possible with a single trained Cognitive suite imagery. Due to many false positives, this bit rate is in unsuitable ranges for input, however. This finding is based on a very short training time and initial user experience. Examining prolonged training times is necessary to determine the actual values.

Feedback on current classification results of the Cognitive suite does increase task completion, SUS and decrease TLX and most of its sub ratings. The subjects are also more likely to prefer the cognitive switch with feedback to the P300 input.

A continuously changing matrix (e.g. in text prediction systems) does not interfere with the cognitive switch and P300 input paradigms.

A searching process at the beginning of BCI input had not significant effect on the frequency of errors. If the approach in this work indeed reflects difficulties of text prediction systems, it is possible to use both cognitive switch and P300 input for their control.

False positives will occur frequently when using the currently available software keyboards for the Cognitive suite of the EPOC. For most novice users, additional learning effort in comparison to P300 input is required for imagery input in the Cognitive suite. This input method is likely to create frustration for these users otherwise. The number of users who can immediately control the Cognitive suite with high accuracy was small during the evaluation.

Concentrating on language processing and text formulation is likely to result in false positives during the imagery-based cognitive switch input. The Midas Touch problem is likely to result in unacceptable false positives during focus changes. It will be necessary to halt the system during non-input periods, which is resulting in less efficient text entry.

In comparison, physical demand is rated higher for P300 input, and mental demand for cognitive switch input. This could negatively affect extended usage times, i.e. when writing long texts.

The P300 input and the cognitive switch with feedback are accepted equally by the subjects. This is reflected in the similar SUS ratings and the system preferences.

Subjects are more likely to accredit errors to themselves during the cognitive suite input in comparison to P300 input. The subjects obviously rated the classifiers of the Emotiv EPOC as accurate, but their own skill to modulate their EEG and focus on their imagination as low and vulnerable.

7.2 Critical view of the process

Before listing possible future research approaches, critical views on the evaluation design and process will be stated. This section can provide improvements for future evaluations using the device.

Short calibration time The evaluation was designed as an initial training and usage of the input systems. Especially for the cognitive switch evaluation, more training time could improve both performance and user acceptance. Often during motor-imagery evaluations, subjects are trained for several days or weeks before evaluation data is recorded. However, for these prolonged evaluation periods less subjects would be interested in participation.

Free cognitive switch training The participants were free to choose and test which input imagination worked best for them during the training of the cognitive switch. This was due to the fact that there was no prior experience with the "black box" EPOC classifiers and well-working input imaginations for the electrode set the device provides. Also, the "skill" value in the cognitive suite proved to be no reliable indicator of input ability. Therefore, the estimation of the subject's cognitive switch ability was based on a list of criteria derived from experience, and on the self-reporting of the subjects.

Basic interface Both evaluation systems should mainly test basic selection using the BCI paradigm, and therefore were kept simple without additional functionality such as phrase prediction. The cognitive switch paradigm was derived from existing EPOC keyboard solutions.

Matrix shuffle It is not certain that shuffling the matrix rows can reflect the respective difficulty of re-orientation in an input matrix when using a word- or phrase prediction system. Originally the evaluation system additionally changed the words in the input matrix for this reason; however this process was expected to confuse the subjects too much.

Transferability It remains unclear to which extend the results are transferable to actual writing systems. The results of the system model including the syntactic tasks were not compared to a system were copy inputs should be made in a sequence without interruptions. Also, no actual free writing task, such as a description of a scene by word or phrase selections, was evaluated. Therefore, the validity of the utilised approach remains unclear.

No paralysed participants Only healthy subjects were evaluated. It remains unclear to which extend paralysed patients and their caregivers are capable of using the consumer EEG. Critical factors such as the dependency on gaze direction or the ability to imagine movement could not be evaluated. Also, EMG signals can not be avoided in these settings, which makes them prominent confounding variables.

7.3 Future research

Based on the declined evaluation variants and on the experiences during the evaluation, several approaches for future research studies can be stated.

Prolonged training The performance of the cognitive switch for text input should be evaluated during multiple training sessions lasting over several days or weeks. Refer to [LWE11, p. 75] for first experimental experiences with training in the Cognitive suite on consecutive days. Prolonged training might also allow using multiple input patterns during the evaluation.

SSVEP evaluation SSVEP is a promising new type of EEG based BCI (refer to section 2.3.1). OpenViBE provides a basic module for this type of interaction, that could theoretically be used with the EPOC. So far the performance of the EPOC device was not investigated for this input paradigm. The device contains the electrodes O1 and O2 close to the primary visual cortex, which may result in good performance ratings of SSVEP.

Input imaginations Participants should receive instructions on which input pattern to use for the cognitive input, and performances should be evaluated comparatively. During the evaluation we noticed that activation patterns recalled from a repetitive or a rich-in-emotion memory performed well. This type of input imagination could e.g. be compared to motor imagery that is typically used in asynchronous interfaces.

Functional prototype A text input system that would be applicable in real-world settings, such as a predictive text input system including phrase prediction should be evaluated descriptively in a free writing task (e.g. by letting the users describe an image). A possible implementation approach would be [Hoe11], where the service NetSpeak which was developed at Bauhaus-University of Weimar was used to implement a phrase prediction system. For the cognitive switch, a feasible clustering system must be developed.

7.4 Overall conclusion

This work first collected data on TLX, SUS and TCR for a high number of participants based on the potentially clinically applicable Emotiv EPOC device. Within the limits of the evaluation system, first estimations of the ITRs of P300 and single-switch input imagery based on this device were provided within a mimicked writing task. We examined the influence of two possible usability problems (Midas touch problem during language tasks, sudden changes of matrix row order). It was shown that adding simple feedback to the input imagery system can improve user acceptance and task performance. The text additionally attempted to provide a comprehensive overview of BCI literature at the time of writing.

The initial question was which of the input paradigms is more suitable for writing tasks. The findings of this evaluation clearly suggest that P300 input should be investigated further, within more efficient text entry systems before the usability problems of the cognitive switch are solved. More specifically, the higher-priced developer's edition that is necessary for P300 classification should be considered for paralysed patients, as well as the development of a predictive P300 text entry system.

Appendix A

Evaluation instructions

A.1 Introduction to the evaluation

BCI-Texteingabesysteme

Ein Mensch im vollständigen Locked-In-Zustand kann bei vollem Bewusstsein keinen Muskel mehr bewegen. Das Syndrom tritt unter anderem durch Stammhirninfarkte oder den Muskelschwund amyotrophe Lateralsklerose (ALS) auf.

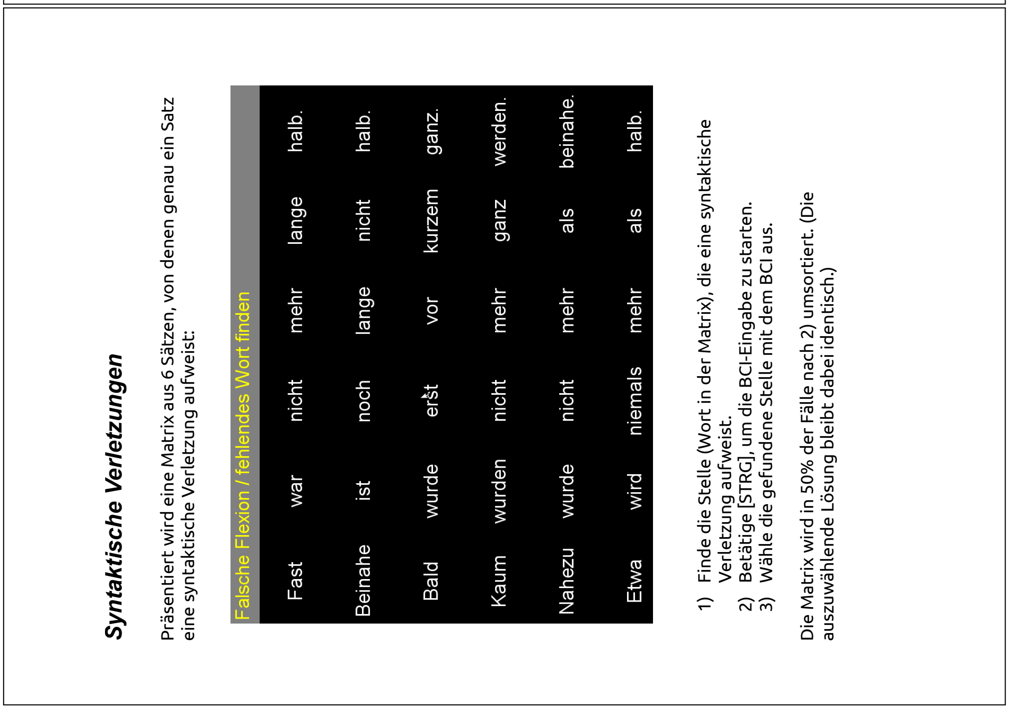
Jean-Dominique Bauby

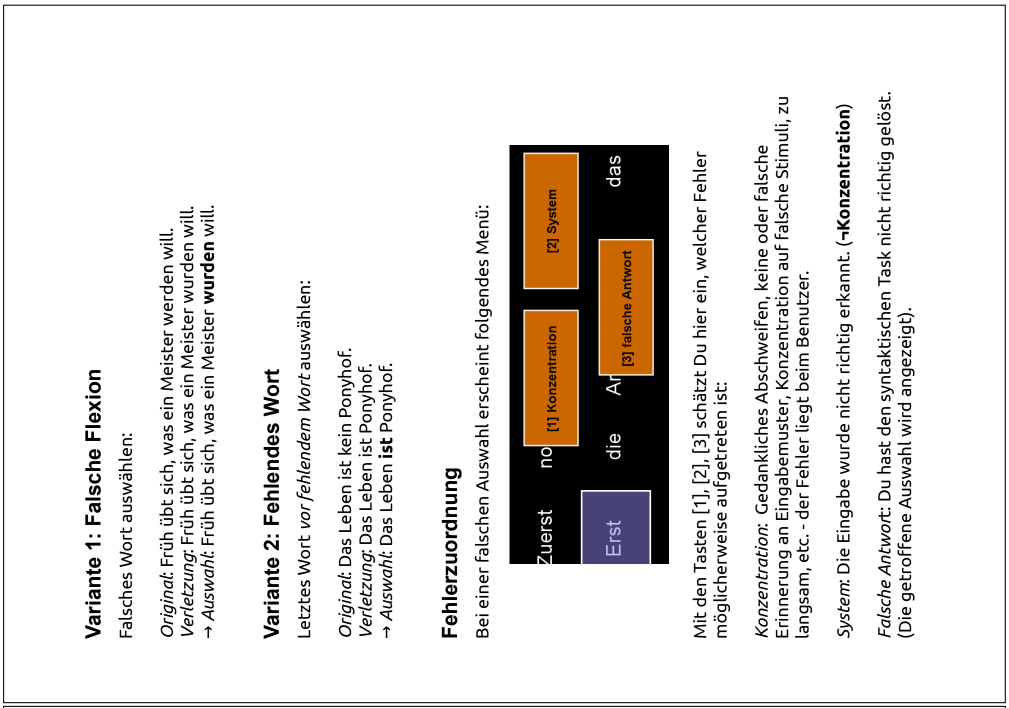
Jean-Dominique Bauby konnte 1997 seinen Krankenschwestern mit Hilfe seines Lidschlags und eines aufgesagten Alphabets das Buch „La scaphandre et le papillon" diktieren. Er erreichte mit Ja/Nein-Entscheidungen eine Eingaberate von etwa einem Wort in zwei Minuten.

Vollständig gelähmte Patienten haben auch diese Möglichkeit nicht. Für diese Menschen könnten Brain-Computer-Interfaces in Zukunft die einzige Möglichkeit darstellen, mit der Außenwelt zu kommunizieren.

In der Evaluation werden zwei Brain-Computer-Interface-Texteingabevarianten auf Basis von EEG-Messungen (Spannungsschwankungen an der Kopfoberfläche) gegeneinander evaluiert.

A.2 Syntactic task instructions

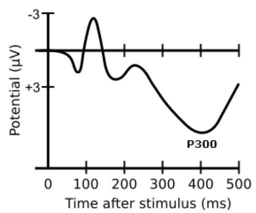




A.3 P300 usage instructions

P300-Spelling

Bei überraschenden, neuartigen oder als selten empfundenen Ereignissen findet man etwa 300ms nach dem Ereignis einen positiven Signalausschlag im EEG:



Da diese Signale sehr stabil sind, kann man sie für BCI-Eingaben verwenden. Der P300-Speller ist eine mögliche Variante: Der Benutzer sieht eine Matrix, deren Reihen und Spalten in zufälliger Reihenfolge aufleuchten. Er kann das gewünschte Element auswählen, indem er seine Aufmerksamkeit darauf richtet.

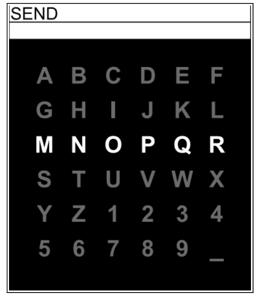
Anleitung

1) Du siehst eine Buchstabenmatrix. In der Zeile am oberen Bildschirmrand befindet sich das einzugebende Wort.
2) Richte Deine Aufmerksamkeit auf den nächsten Buchstaben.
3) Bei jedem Aufleuchten: Zählen *oder* den Buchstaben denken *oder* „diesen“ bzw. „ja“ denken.

Während Zeilen und Spalten aufleuchten, vermeide:

- Zwinkern, Augenbewegungen, Sprechen, Schlucken
- Muskelkontraktionen, z.B. im Gesicht oder am Kopf
- Bewegungen der Gliedmaßen
- gedankliches Abschweifen

Schritt 1: Kalibrierung



Die P300-Signale verschiedener Personen unterscheiden sich in Form, Amplituden und Zeitverlauf. Wir stimmen deshalb zunächst die Erkennung des P300-Signals auf Dein spezifisches Signal ab.

Das System zeigt Dir in der oberen Zeile ein Wort, das mit P300-Speller Buchstabe für Buchstabe eingegeben wird. Insgesamt werden bis zu fünf Worte eingegeben.

Schritt 2: Evaluation

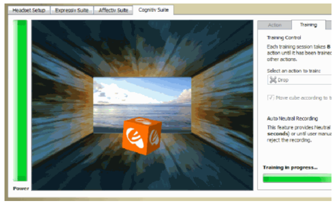
Die ermittelten Systemparameter werden in das Evaluationssystem geladen. Wenn das System gut kalibriert wurde, wirst Du bei der Evaluation insgesamt weniger Einzeleingaben machen als bei der Kalibrierung.

Hinweise

- ein gutes P300-Signal ist spürbar, z.B. als *neugierig*, *leicht aufgeregt* , und ein wenig *durchfließend*
- glücklich und entspannt 

A.4 Cognitive switch usage instructions

EPOC cognitive input



Die Software des EPOC kann Vorstellungen erkennen und diese in externen Programmen zur Eingabe bereit stellen. Das System muss dazu auf eine bestimmte Eingabevorstellung hin kalibriert werden.

Schritt 1: Kalibrierung

Trainiere in der Cognitive Suite, bis Du mit dem Eingabemuster 85% Genauigkeit für die Eingabe DROP erreichst. Neben der aktiven Eingabe muss auch der neutrale Zustand NEUTRAL als Basiszustand trainiert werden.

Wähle eines der folgenden Eingabemuster:

- eine *aktive* Bewegung (bspw. Handtuch auswringen, Hand zur Faust ballen)
- einen Klang (bspw. Sänger bei Kammerton, Gitarrenriff)
- andere Vorstellung, die subjektiv besser funktioniert

Bitte teile mir am Ende der Kalibrierung Dein Eingabemuster mit.

Vermeide Folgendes in Zusammenhang mit der Kalibrierung:

- Zwinkern, Augenbewegungen, Sprechen, Schlucken
- Muskelkontraktionen, z.B. im Gesicht oder am Kopf
- Bewegungen der Gliedmaßen

Das heißt: Trainiere das Gerät nicht versehentlich auf eine Muskelbewegung.

Schritt 2: Aktivierungszeit

Ermittelt werden hier Aktivierungs- und Deaktivierungszeit für das Eingabemuster.

activate: DROP
release: NEUTRAL

Schritt 3: Evaluation

Über die Eingabematrix läuft ein Cursor. Mit dem Cursor wird zuerst eine Zeile, dann eine Spalte ausgewählt. Erzeuge das Eingabemuster, sobald der Cursor das gewünschte Feld markiert.

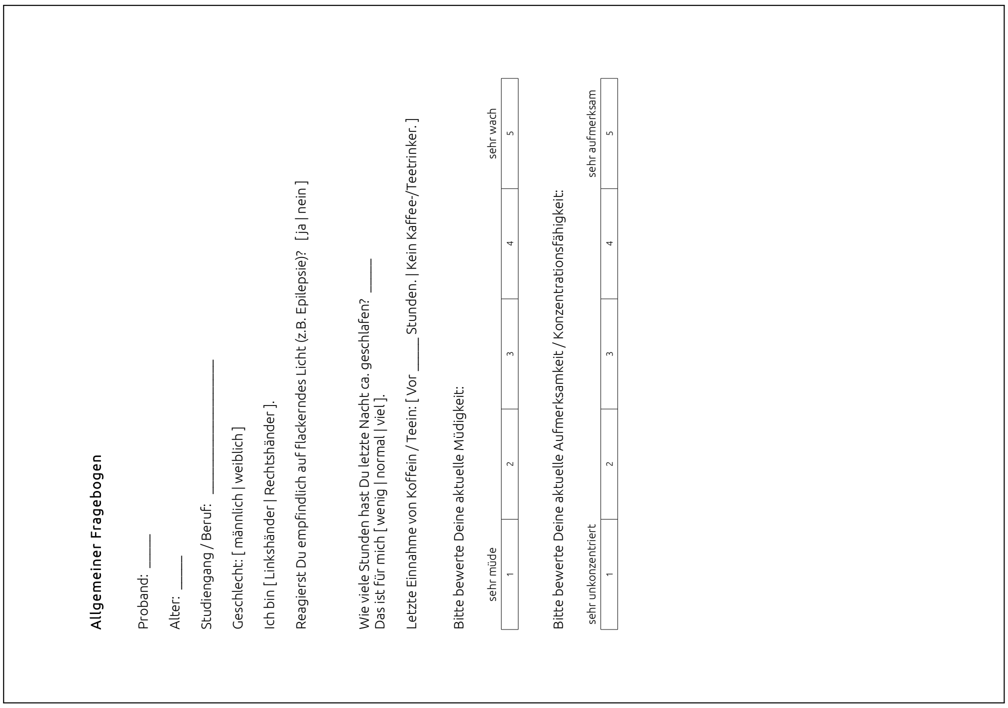
Hinweise

- Der Übergang vom neutralen Zustand zum Muster ist fließend / leicht und erfordert Aufmerksamkeit und Konzentration, aber nicht „Angestrengtheit".
- Eine genaue Vorstellung führt zu einer besseren Erkennung.
- Die Latenz bei der Erkennung kann 1-2 Sekunden betragen.

Appendix B

Questionnaires

B.1 Demographic questionnaire



Allgemeiner Fragebogen

Proband: _____

Alter: _____

Studiengang / Beruf: ____________________

Geschlecht: [männlich | weiblich]

Ich bin [Linkshänder | Rechtshänder].

Reagierst Du empfindlich auf flackerndes Licht (z.B. Epilepsie)? [ja | nein]

Wie viele Stunden hast Du letzte Nacht ca. geschlafen? _____
Das ist für mich [wenig | normal | viel].

Letzte Einnahme von Koffein / Teein: [Vor _____ Stunden. | Kein Kaffee-/Teetrinker.]

Bitte bewerte Deine aktuelle Müdigkeit:

sehr müde				sehr wach
1	2	3	4	5

Bitte bewerte Deine aktuelle Aufmerksamkeit / Konzentrationsfähigkeit:

sehr unkonzentriert				sehr aufmerksam
1	2	3	4	5

B.2 NASA Task Load Index (TLX)

Proband: ________ System: ________

Wie geistig anstrengend war die Aufgabe?

sehr wenig | sehr stark

Wie körperlich anstrengend war die Aufgabe (z.B. für die Augen)?

sehr wenig | sehr stark

Wie stark empfandest Du bei der Aufgabe Zeitdruck?

sehr wenig | sehr stark

Wie erfolgreich hast Du die Aufgabe Deiner Einschätzung nach durchgeführt?

perfekter Erfolg | Misserfolg

Wie sehr musstest Du Dich anstrengen, um Deine Leistung zu erreichen?

sehr wenig | sehr stark

Wie verunsichert, entmutigt, gereizt oder verärgert warst Du?

sehr wenig | sehr stark

1

B.3 Modified System Usability Scale (SUS)

System usability scale (*mod.*)

Proband: _____ System: _____

	trifft gar nicht zu				trifft völlig zu
	1	2	3	4	5
1. Ich kann mir vorstellen, dieses System häufig zu benutzen.	1	2	3	4	5
2. Das System war sehr einfach zu benutzen.	1	2	3	4	5
3. Das System war sehr mühsam zu benutzen.	1	2	3	4	5
4. Ich kann mir vorstellen, dass die meisten Leute den Umgang mit dem System sehr schnell lernen würden.	1	2	3	4	5
5. Ich konnte das System sehr sicher benutzen.	1	2	3	4	5
6. Die Benutzung des Systems machte mich müde oder unkonzentriert.	1	2	3	4	5
7. Ich hatte Spaß bei der Benutzung des Systems.	1	2	3	4	5
8. Die Benutzung des Systems war sehr frustrierend.	1	2	3	4	5
9. Ich musste viele Dinge lernen, bevor ich das System benutzen konnte.	1	2	3	4	5
10. Am Ende des Tests war die Benutzung des Systems viel schwieriger als am Anfang.	1	2	3	4	5

B.4 Final questionnaire

Abschließender Fragebogen

Proband: _____

Welche Unterstützung hättest Du Dir bei der Eingabe gewünscht? Durch welche Erweiterungen könnten die Systeme leichter zu benutzen sein?

P300-Spelling

____

____

____

EPOC cognitive input

____

____

____

Welche Eingabefehler traten am häufigsten auf (zwei auswählen)?

EPOC cognitive input

- ☐ unter Druck stehen
- ☐ zu spät erzeugtes Muster
- ☐ zu langsam zurück zu NEUTRAL
- ☐ versehentlich erzeugtes Muster
- ☐ sonstiges: ____________________
- ☐ “Denke nicht an Eisbären“
- ☐ Muskelbewegungen
- ☐ falsch oder unvollständig erzeugtes Muster

P300-Spelling

- ☐ gedankliches Abschweifen
- ☐ zu langsam auf Aufleuchten reagiert
- ☐ Indifferenz beim Aufleuchten
- ☐ sonstiges: ______________________
- ☐ Muskelbewegungen
- ☐ Umgebungsreize

Welches Eingabesystem bevorzugst Du?

[P300-Spelling | EPOC cognitive input]

Bitte kurz begründen:

____

____

____

Appendix C

Descriptive answers

C.1 Rationale for system preference

00t (P300 | cognitive switch without feedback)
"EPOC war anstrengender und entmutigender"

01t (P300 | cognitive switch without feedback)
"P300, fehlende Willensanstrengung, stressärmer, direkter"

02t (P300 | cognitive switch without feedback)
"P300 ist zwar 'leichter' zu bedienen, allerdings auf Dauer anstrengend. Ich könnte mir EPOC mit mehr Training effektiver vorstellen."

03t (P300 | cognitive switch with feedback)
"Eindeutig. EPOC. P300 war mir vom Cursor zu hektisch. Bei weiterem Training mit EPOC kann man bestimmt sehr gute Ergebnisse erzielen."

04t (P300 | cognitive switch with feedback)
"Ein starkes Pattern ist relativ einfach zu erzeugen. Ein positives Pattern macht sogar Spaß, somit ist auch der gesamte Eingabeprozess nicht so langweilig. Weiterhin strengt EPOC, im Gegensatz zu P300, die Augen nur minimal an und man ermüdet – ich jedenfalls! - gar nicht."

05t (P300 | cognitive switch without feedback)
"es funktioniert besser"

06t (P300 | cognitive switch without feedback)
"Solange man noch nicht zu müde ist, gab es weniger Fehleingaben, was das Benutzen motivierender machte."

07t (<u>P300</u> | cognitive switch without feedback)
“Die nötige Konzentration ist nicht so hoch, bzw. muss der Zustand nicht gehalten werden im Gegensatz zum EPOC.”

08t (P300 | <u>cognitive switch without feedback</u>)
“Ich glaube durch Training kann ich besser werden. Das ‘Blitzen’ hat bei P300-Spelling genervt.”

09t (P300 | <u>cognitive switch with feedback</u>)
“Gefühlt weniger anstrengend, angenehm für die Augen, visuelles Feedback war hilfreich”

10t (P300 | <u>cognitive switch without feedback</u>)
“EPOC, da ich da mehr Einfluss hatte. Ich selbst bin verantwortlich, falls ich mein Pattern nicht ‘richtig’ anwende. Bei P300 kann ich nur versuchen konzentriert zu bleiben, habe aber keinen bewussten Einfluss.”

11t (<u>P300</u> | cognitive switch without feedback)
“Weil der andere Test nicht so gut funktioniert hat =) Man hat bei P300 das Gefühl bei einem Fehler nicht schon verloren zu haben (von neu beginnen zu müssen)”

12t (<u>P300</u> | cognitive switch without feedback)
“Es war leichter die Reize oft/schnell zu erzeugen. Für mich weniger anstrengend”

13t (P300 | <u>cognitive switch without feedback</u>)
“nicht so anstrengend”

14t (P300 | <u>cognitive switch with feedback</u>)
“Angenehmer in der Benutzung, weniger anstrengend, da neutrale Haltung möglich ist während der Benutzung”

15t (P300 | <u>cognitive switch without feedback</u>)
“Mit mehr Training würde es vermutlich besser funktionieren, P300 ist für Augen zu anstrengend”

16t (<u>P300</u> | cognitive switch without feedback)
“EPOC war für mich unbenutzbar”

17t (<u>P300</u> | <u>cognitive switch without feedback</u>)
“Kurze Sätze / Eingaben: P300, Längere Sätze: EPOC”

18t (<u>P300</u> | cognitive switch without feedback)
“Nicht so anstrengend / geht schneller”

21t (P300 | cognitive switch with feedback)
"bessere Erfolge; subjektiv erfolgreicher angewandt, freie Patternwahl vereinfacht 'Schalten' (bewusstes Schalten; einmalig darauf vorbereitet, und nicht ständig)"

22t (P300 | cognitive switch with feedback)
"EPOC ist mächtiger, da sich verschiedene Pattern parallel erlernen lassen (größeres Eingabe-Alphabet)."

23t (P300 | cognitive switch with feedback)
"Mehr Treffer, intuitiver, aber anstrengend, gefühlt langsamer."

24t (P300 | cognitive switch with feedback)
"Weil ich damit besser klar kam, d.h. erfolgreicher war und deutlich weniger frustriert; allerdings empfand ich es auch als deutlich anstrengender, da konstante Aufmerksamkeit erforderlich war"

25t (P300 | cognitive switch with feedback)
"Auslöseimpuls ist eher binär, wobei ich bei EPOC eine größere Dauer benötige, um den Impuls (DROP) auszulösen"

26t (P300 | cognitive switch with feedback)
"viel einfacher zu bedienen, nicht so ermüdend, fast gar nicht frustrierend"

27t (P300 | cognitive switch with feedback)
"intuitiver und augenfreundlicher, insgesamt entspannter, mit Übung schneller"

28t (P300 | cognitive switch with feedback)
"P300 war weniger (mental) anstrengend"

29t (P300 | cognitive switch with feedback)
"wesentlich weniger frustrierend, 'gefühlt' genauer und schneller"

C.2 Improvements of P300

00t
"Erkennung ob falsches Wort nur möglich bei eindeutiger Kennzeichnung des Wortes"

01t

"Unterscheidung falsch-fehlend notwendig? Evtl. soll Position Fehler kenntlich machen, scheinbar zwei Helligkeitswerte bei Hervorhebung? "Echoeffekt'?"

02t

"Ebenso, maximal Unterstützung durch Eye-Tracking"

03t

"Das Aufleuchten war zu kurzfristig. Interface war ansonsten okay."

04t

"endgültige Selektion dauert sehr lange - Augen ermüden |Selektionsphase sollte einfach zu unterbrechen sein, zum Ausruhen"

08t

"Ein visuelles Feedback. Irgendwie zu sehen, wie ich mich konzentriere."

09t

"'ruhigeres' Flackern, Wort leicht vergrößern, wenn es aufblinkt"

10t

"Feedback zur Patternaktivierung"

12t

"Zur Bestätigung visuelles Feedback über Erfolg und Misserfolg"

14t

"geringere Frequenz des Aufleuchtens, weniger Begriffe oder größere räumliche Trennung um weniger von den daneben Liegenden beeinflusst zu werden"

16t

"Langsameres Flashen der Worte"

17t

"Visuelle Unterstützung des Favoriten (Feedback)"

18t

"Zwischenklassifizierung (momentane Auswahl)"

19t

"Anzeigedauer und Wechselgeschwindigkeit scheinen teilweise zu hoch, um das Signal abzusetzen"

26t
“Feedback, ob eine Eingabe richtig war”

27t
“Grün auf Schwarz”

28t
“weniger kontrastreiche Darstellung der Buchstaben wäre angenehmer für die Augen (Kontrast: Hintergrund-Buchstaben)”

C.3 Improvements of cognitive switch

00t
“Visuelles Feedback: des Pattern, der Zeilenaktivierung”

01t
“transparente Bedienerführung, Zuverlässigkeit Bedienen (Testen)”

02t
“Möglicherweise Tracking der Augen um eine ‘mögliche’ Auswahl zu forcieren.”

03t
“Konkrete Verbesserungsvorschläge: Keine. Ich musste immer die letzte Sekunde des vorangegangenen Feldes für die Mustererzeugung mit nutzen.”

06t
“Ein Pattern zum abbrechen”

08t
“Visuelles Feedback wäre auch hier hilfreich. Ich kann mir vorstellen, dass man über einen längeren Zeitraum besser wird.”

11t
“Kalibrierung sehr wichtig, hohe Konzentration erforderlich (kann nicht lange aufrecht erhalten werden)”

12t
“Probleme: zu langes Warten, dadurch hohe Konzentrationsschwäche → keine Lösung denkbar, da Auslösung zu lange braucht.”

17t
“Schwellwertanzeige (Feedback)”

18t

"Je nach Sicherheit auch True Negatives klassifizieren (und ausschließen)"

23t

"Response bei Neustart, starte ich evtl. schon mit false positive?"

24t

'Bessere Anpassung (Adaption, da sich Grundzustand im Verlauf ändert), intensiverer Auswahlvorgang, um ein eindeutiges Pattern zu finden"

25t

"Der Cursor sollte nie versteckt werden, sollte immer präsent sein. Dynamischer Release Threshold, nach einer Auswahl den Threshold höher setzen, damit das Release nicht gestört wird"

27t

"dynamische Schwellwertanpassung—bei großer Auswahl bzw. langen Sätzen → Auswahl, ob man von oben/unten bzw. rechts/links anfängt, auszuwählen → falls mehrere Pattern → Pattern für switch oben/unten bzw. rechts/links"

28t

"zweites Pattern für 'zurück"'

29t

"Pause nach einer Eingabe, um mögliche Falscheingaben (immer noch aktives Pattern) zu umgehen"

C.4 Cognitive switch activation imaginations

At the end of the evaluation, the users described their input patterns as detailed as possible to the experimenter. The experimenter asked questions on certain properties of the imagination, e.g. whether muscle movements or sensory perception was included. This section will list these descriptive answers of those participants that could complete the training process.

00t

"Kampfschrei von links"

01t

"Griffbewegung"

02t

"mentales Fixieren / Ertasten von Gegenständen"

04t

"Emotionen bei einem kleinen Wettkampf mit einem Freund (Erinnerung vom Morgen)"

05t

"'runter, runter, runter' denken, mit leichter Wut"

06t

"Handtuch auswringen"

07t

"Geräusch einer Kreissäge"

08t

"Vorstellung, einen Ton zu singen (aktiver Chorsänger)"

09t

"mit den Zehen wackeln, Gefühl"

10t

"fröhlich Spaghetti mit Gabel aufrollen"

11t

"Gefühl: 'Geh runter, Du böser Kubus!"'

12t

"Marimba-Klänge"

13t

"Freund von mir umarmen"

14t

"Nadel auf einen Plattenspieler setzen, Kratzen der Nadel beim Aufsetzen, Geräusch der Platte am Anfang (Knistern), erste Töne der Musik, Vorstellung der Bewegung der Nadel mit der Hand, Erinnerung an Tastgefühl vorhanden, Gefühl der Faszination über Analogheit"

15t

"eigene Fußmatte auswringen (Erinnerung vom Morgen)"

16t

"Kartentrick-Bewegung (Erinnerung an regelmäßige Bewegung)"

17t

"Topf mit Schwamm reinigen"

18t

"Handtuch auswringen"

19t

"Police Academy-Titelmelodie und Intro-Video"

21t

"(aktiver Fußballspieler) kurz vorm Strafraum gegen Ball treten, Gefühl des Balls, Visualisierung, Trigger: Schuss, Bahn verfolgen, Ball zappelt im Netz"

22t

"bestimmte Übung mit Bo (Kampfsport)"

24t

"Schwimmszene am Meer, emotionsreich, ruhig"

26t

"Klettern, bestimmter schwieriger/anstrengender Handgriff, aus Hocke beim Klettern, mit Füßen abdrücken und mit Händen hochziehen, gut in Erinnerung, "Triumph" beim ersten Gelingen, Hochziehvorgang"

27t

"Einfangen einer Frisbee-Scheibe aus der 3. Person gesehen, positive Emotion, Ablauf, Druck auf Fingern – Trigger: Scheibe fangen"

28t

"Gitarrengriff: Langsames Anschlagen der Saiten, Saite für Saite gespürt, Klang vorgestellt, mentales Durchführen der Bewegung"

List of Figures

List of Tables

Acronyms

A|PE Group A|PE (with 1. P300 input, 2. Cognitive switch).

ALS Amyotrophic Lateral Sclerosis.

AT Assistive technology.

B|EP Group B|EP (with 1. Cognitive switch, 2. P300 input).

BCI Brain-Computer Interaction.

EEG Electroencephalography.

EMG Electromyogram.

ERD Event-related desynchronization.

ERP Event-related potential.

fMRI functional Magnetic Resonance Imaging.

HCI Human-Computer Interaction.

INRIA Institut national de recherche en informatique et en automatique.

ITR Information Transfer Rate.

PVS Persistent Vegetative State.

Pyff Pythonic feedback framework.

SCP Slow Cortical Potentials.

SMR Sensorimotor rhythms.

SSVEP Steady state visual evoked potential.

SUS System Usability Scale.

TCR Task Completion Rate.

TLX NASA Task Load Index.

Bibliography

[ABJ+01] Kuebler A, Kotchoubey B, Kaiser J, Wolpaw JR, and Birbaumer N. Brain-computer communication: unlocking the locked in. *Psychological Bulletin*, 127(3):358–375, 2001.

[AMML96] Keith Andrews, Lesley Murphy, Ros Munday, and Clare Littlewood. Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. *British Medical Journal*, 313:13–16, 1996.

[AN08] Kuebler A. and Birbaumer N. Brain-Computer Interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? *Clinical Neurophysiology*, 119(11):2658–2666, 2008.

[ANO06] Dixit A, Vaney N, and Tandon OP. Evaluation of cognitive brain functions in caffeine users: a P3 evoked potential study. *Indian Journal of Physiology and Pharmacology*, 50:175–180, 2006.

[BAB+11] C Brunner, G Andreoni, L Bianchi, B Blankertz, C Breitwieser, S Kanoh, CA Kothe, A Lecuyer, S Makeig, J Mellinger, P Perego, Y Renard, G Schalk, IP Susila, B Venthur, and GR Mueller-Putz. BCI Software Platforms, 2011.

[Bau97] Jean-Dominique Bauby. *Le Scaphandre et le Papillon*. Editions Robert Laffont, 1997.

[BCGM09] T.W. Berger, J.K. Chapin, G.A. Gerhardt, and D.J. McFarland. *Brain-computer interfaces: an international assessment of research and development trends*. Springer, 2009.

[BDK+06] B. Blankertz, G. Dornhege, M. Krauledat, M. Schröder, J. Williamson, R. Murray-Smith, and K.R. Muller. The Berlin brain-computer interface presents the novel mental typewriter hex-o-spell. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, pages 108–109, 2006.

[Ber29] Hans Berger. Über das Elektrenkephalogramm des Menschen. *European Archives of Psychiatry and Clinical Neuroscience*, 87:527–570, 1929.

[BFC+11] Pavel Bobrov, Alexander Frolov, Charles Cantor, Irina Fedulova, Mikhail Bakhnyan, and Alexander Zhavoronkov. Brain-Computer Interface Based on Generation of Visual Images. *PLoS ONE*, 6(6), 06 2011.

[BJB+10] P Brunner, S Joshi, S Briskin, J R Wolpaw, H Bischof, and G Schalk. Does the 'P300' speller depend on eye gaze? *Journal of Neural Engineering*, 7(5):056013, 2010.

[Bra] AutoNOMOS: BrainDriver. URL: `http://www.autonomos.inf.fu-berlin.de/subprojects/braindriver`.

[BSJ+10] Venthur B, Scholler S, Williamson J, Daehne S, Treder MS, Kramarek MT, Mueller KR, and Blankertz B. Pyff - a pythonic framework for feedback applications and stimulus presentation in neuroscience. *Frontiers in Neuroscience*, 4:179, 2010.

[Bus08] Ernesto A. Bustamante. Measurement invariance of the NASA TLX. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 2008.

[CCH+10] Andrew Campbell, Tanzeem Choudhury, Shaohan Hu, Hong Lu, Matthew K. Mukerjee, Mashfiqui Rabbi, and Rajeev D.S. Raizada. NeuroPhone: brain-mobile phone interface using a wireless EEG headset. In *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*, MobiHeld '10, pages 3–8, New York, NY, USA, 2010. ACM.

[Cec10a] H. Cecotti. A Self-Paced and Calibration-Less SSVEP-Based Brain-Computer Interface Speller. *Neural Systems and Rehabilitation Engineering*, 18 (2):127–133, 2010.

[Cec10b] H. Cecotti. Spelling with Brain-Computer Interfaces: Current trends and prospects. In *Cinquieme conference pleniere francaise de Neurosciences Computationnelles*, 2010.

[COEK11] Daniel Cernea, Peter-Scott Olech, Achim Ebert, and Andreas Kerren. EEG-based Measurement of Subjective Parameters in Evaluations. In *Proceedings of the 14th International Conference on Human-Computer Interaction*, 2011.

[CSE+09] Guger C, Daban S, Sellers E, Holzner C, Krausz G, Carabalona R, Gramatica F, and Edlinger G. How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462(1):94-8, 2009.

[D'A09] T. D'Albis. *A predictive speller for a brain-computer interface based on motor-imagery.* PhD thesis, University of Milan, 2009.

[DKV+10] Arnaud Delorme, Christian Kothe, Andrey Vankov, Nima Bigdely-Shamlo, Robert Oostenveld, Thorsten Zander, and Scott Makeig. MATLAB-Based Tools for BCI Research. In Desney S. Tan and Anton Nijholt, editors, *Brain-Computer Interfaces*, Human-Computer Interaction Series, pages 241–259. Springer London, 2010.

[Dor07] G. Dornhege. *Toward brain-computer interfacing.* Neural information processing series. MIT Press, 2007.

[DSM+09] Thorsten Dickhaus, Claudia Sannelli, Klaus-Robert Muller, Gabriel Curio, and Benjamin Blankertz. Predicting BCI performance to study BCI illiteracy. *BMC Neuroscience*, 10(Suppl 1):P84, 2009.

[Dum44] Alexandre Dumas. *The Count of Monte Cristo.* Journal des Debats, 1844.

[EM09] J. Elshout and G.G. Molina. Review of Brain-Computer Interfaces based on the P300 evoked potential. Technical report, Philips Research Europe, 2009.

[Emo11] Emotiv. Emotiv EPOC published papers, 2011. URL: `http://emotiv.com/researchers/`.

[Emo12] Emotiv. Emotiv.com: Researchers, 2012. URL: `http://emotiv.com/researchers/`.

[FD88] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prothesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70:510–523, 1988.

[FH81] Linda Flower and John R Hayes. A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4):365–387, 1981.

[GAP11] B. Graimann, B. Allison, and G. Pfurtscheller. *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction.* The Frontiers Collection. Springer, 2011.

[GHP11] H. Gürkök, G. Hakvoort, and M. Poel. Evaluating user experience with respect to user expectations in brain-computer interface games. In G. R. Müller-Putz, R. Scherer, M. Billinger, A. Kreilinger, V. Kaiser, and C. Neuper, editors, *Proceedings of the 5th International Brain-Computer Interface Conference, BCI 2011, Graz, Austria*, pages 348–351, Graz, Austria, September 2011. Verlag der Technischen Universität Graz.

[GPv+11] H. Gürkök, D. Plass-Oude Bos, B. L. A. van de Laar, F. Nijboer, and A. Nijholt. User Experience Evaluation in BCI: Filling the Gap. *International Journal of Bioelectromagnetism*, 13(1):54–55, July 2011.

[GUM11] Pires G., Nunes U., and Cestelo-Branco M. GIBS Block Speller: Toward a Gaze-Independent P300-based BCI. In *Proc. IEEE Eng. Med. Biol.*, 2011.

[Har88] S. G. Hart. Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research. *Human Mental Workload*, 1988.

[Har06] Sandra G. Hart. Nasa-Task Load Index (Nasa-TLX); 20 Years Later. In *Human Factors and Ergonomics Society Annual Meeting*, volume 50, 2006.

[Hoe11] Felicitas Hoebelt. MessageAssistant: Ein grafischer, haeufigkeitsbasierter Ansatz zur Komposition von Kurznachrichten. Master's thesis, Bauhaus-Universitaet Weimar, 2011.

[Jac91] Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9:152–169, 1991.

[Jas58] Herbert H. Jasper. Report of the committee on methods of clinical examination in electroencephalography - appendix: The ten-twenty electrode system of the International Federation. *Electroencephalography and clinical neurophysiology*, 10:371–375, 1958.

[JSS+10] Muenssinger JI, Halder S, Kleih SC, Furdea A, Raco V, Hoesle A, and Kuebler A. Brain Painting: First Evaluation of a New Brain-Computer Interface Application with ALS-Patients and Healthy Volunteers. *Frontiers in Neuroscience*, 4, 2010.

[KKH+99] Andrea Kuebler, Boris Kotchoubey, Thilo Hinterberger, Nimr Ghanayim, Juri Perelmouter, Margarete Schauer, Christoph Fritsch, Edward Taub, and N. Birbaumer. The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Experimental Brain Research*, 124:223–232, 1999. 10.1007/s002210050617.

[KSM+08] D.J. Krusienski, E.W. Sellers, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw. Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, 167:15 – 21, 2008.

[LWE11] Robert Lievesley, Martin Wozencroft, and David Ewins. The Emotiv EPOC neuroheadset: an inexpensive method of controlling assistive technologies using facial expressions and thoughts? *Journal of Assistive Technologies*, 5:67 – 82, 2011.

[Mac10] Geoff Mackellar. Emotiv forums: Emotiv detections and disabled users, Check here first, then contact us if you have more questions, 2010. URL: `http://emotiv.com/forum/messages/forum3/topic1016/message6121/#message6121`.

[MBH+08] E. Mugler, M. Bensch, S. Halder, Wolfgang Rosenstiel, M. Bogdan, N. Birbaumer, and A. Kuebler. Control of an Internet Browser Using the P300 Event Related Potential. *International Journal of Bioelectromagnetism*, Vol. 10, No.1:56–63, 2008.

[MP95] J Malmivuo and R Plonsey. *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields.* Oxford University Press, 1995.

[NDBD11] Le Nguyen, Paul Du, Kim-Mai Bui, and Lap Duong. UCEEG - Project Brain Speller, Microsoft Imagine Cup, 2011. URL: `http://ise.canberra.edu.au/dtran/activities/imagine-cup-2011/`.

[NYY+09] C. S. Nam, J. Yongwoong, L. Yuequing, Y.J. Kim, and H.Y. Yoon. Usability of the P300 Speller: Towards a More Sustainable Brain-Computer Interface. *International Journal on Human-Computer Interaction*, 1(5), 2009.

[PAA+03] M.A. Pastor, J. Artieda, J. Arbizu, M. Valencia, and J. C. Masdeu. Human Cerebral Activation during Steady-State Visual-Evoked Potentials. *The Journal of Neuroscience*, 23(37):11621–11627, 2003.

[Pas11] E. Pasqualotto. *Usable Communication: Usability Evaluation of Brain-Computer Interfaces.* PhD thesis, Sapienza University of Rome, 2011.

[PGv+11] D. Plass-Oude Bos, H. Gürkök, B. L. A. van de Laar, F. Nijboer, and A. Nijholt. User Experience Evaluation in BCI: Mind the Gap! *International Journal of Bioelectromagnetism*, 13(1):48–49, July 2011.

[PK10] Matthias Pfeiffer and Detlef Krömker. The emotiv EPOC BCI as inexpensive solution for the P300 spelling task. In Ralf Dörner and Detlef Krömker, editors, *Self Integrating Systems for Better Living Environments: First Workshop, Sensyble 2010*, number 1, pages 29–35. Shaker Aachen, November 2010.

[PSG+11] E. Pasqualotto, A. Simonetta, V. Gnisci, S. Federici, and M. O. Belardinelli. Toward a usability evaluation of BCIs. *International Journal of Bioelectromagnetism*, 13(3):121 – 122, 2011.

[SC07] Saeid Sanei and J. A. Chambers. *EEG Signal Processing.* Wiley-Interscience, 2007.

[SKM+06] Eric W. Sellers, Dean J. Krusienski, Dennis J. McFarland, Theresa M. Vaughan, and Jonathan R. Wolpaw. A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73(3):242–252, 2006.

[SM10] G. Schalk and J. Mellinger. *Introduction to Brain-Computer Interfacing Using BCI2000.* Human-Computer Interaction Series. Springer, 2010.

[SMH+04] G Schalk, D McFarland, T. Hinterberger, N Birbaumer, and JR Wolpaw. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.

[Sut11] Craig Sutherland. Plug in the Brain: Evaluating the Usability of Brain Computer Interfaces, 2011.

[THM+09] M Turunen, J Hakulinen, J Melto, T Heimonen, T Laivo, and J Hella. SUXES - User Experience Evaluation Method for Spoken and Multimodal Interaction. In *Proceedings of INTERSPEECH 2009*, 2009.

[Til12] Tilvus. Tilvus Assistive Interface, 2012. URL: `http://tilvus.net/About.html`.

[TLB+10] G. Townsend, B.K. LaPallo, C.B. Boulay, D.J. Krusienski, G.E. Frye, C.K. Hauser, N.E. Schwartz, T.M. Vaughan, Wolpaw J.R., and Sellers E.W. A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clinical Neurophysiology*, 121(7):1109–20, 2010.

[TN10] D.S. Tan and A. Nijholt. *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*. Human-Computer Interaction Series. Springer, 2010.

[vGP+11] B. L. A. van de Laar, H. Gürkök, D. Plass-Oude Bos, F. Nijboer, and A. Nijholt. Perspectives on User Experience Evaluation of Brain-Computer Interfaces. In C. Stephanidis, editor, *Universal Access in Human-Computer Interaction. Users Diversity. 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011, Orlando*, volume 6766 of *Lecture Notes in Computer Science*, pages 600–609, Berlin, June 2011. Springer Verlag.

[Vid73] J J Vidal. Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, 2(1):157–180, 1973.

[vNG+11] B. L. A. van de Laar, F. Nijboer, H. Gürkök, D. Plass-Oude Bos, and A. Nijholt. User Experience Evaluation in BCI: Bridge the Gap. *International Journal of Bioelectromagnetism*, 13(3):157–158, September 2011.

[Vol11] Ivan Volosyak. SSVEP-based Bremen-BCI interface - boosting information transfer rates. *Journal of Neural Engineering*, 8(3):036020, 2011.

[Wat11] Scott Watter. Emotiv forums: Test Bench marker timing accuracy-please fix, 2011. URL: `http://emotiv.com/forum/messages/forum22/topic1797/message10441/#message10441`.

[WBM+00] JR. Wolpaw, N. Birbaumer, D. McFarland, PH Peckham, G. Schalk, E. Donchin, LA Quatrano, CJ. Robinson, and

T. Vaughan. Brain-Computer Interface Technology: A Review of the First International Meeting. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 8:161–163, 2000.

[WBM+02] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[YFG+10] Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy, Bertrang, and Lecuyer A. OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments. *Presence*, 19(1):35–53, 2010.