Markus Pröll

USING A LOW-COST GYRO AND EEG-BASED INPUT DEVICE IN INTERACTIVE GAME DESIGN

Master's Thesis

GRAZ UNIVERSITY OF TECHNOLOGY
INSTITUTE FOR KNOWLEDGE DISCOVERY
LABORATORY OF BRAIN-COMPUTER INTERFACES

Graz, Austria, May 2012

Head of Department
Assoc.Prof. Dipl.-Ing. Dr.techn. Gernot Müller-Putz

Supervisor
Ass.Prof. Dipl.-Ing. Dr.techn. Reinhold Scherer

Evaluator
Ass.Prof. Dipl.-Ing. Dr.techn. Reinhold Scherer

Publications arising from this Thesis

Poster:
R. Scherer, M. Pröll, B. Z. Allison, and G. Müller-Putz. New input modalities for modern game design and virtual embodiment. *IEEE Virtual Reality*, 2012.

Video:
R. Scherer, M. Pröll, B. Z. Allison, M. Rapp, and G. Müller-Putz. BCI-based interaction with World of Warcraft. *IEEE Virtual Reality*, 2012.

Papers:
R. Scherer, E. Friedrich, B. Allison, M. Pröll, M. Chung, W. Cheung, R. Rao, and C. Neuper. Non-invasive Brain-Computer Interfaces: Enhanced Gaming and Robotic Control. *IWANN 2011*, Part I, LNCS, 6691:362-369, 2011.

R. Scherer, J. Faller, M. Pröll, D. Balderas, E. C. V. Friedrich, B. Z. Allison, and G. Müller-Putz. Towards Context-Sensitive Hybrid Brain-Computer Interfaces. *IWANN 2011 Soft-Computing Special Issue*, 2012. *(Submitted)*

Tobias G. Oesterlein, M. Pröll, S. Marko, G. Müller-Putz, and R. Scherer. Playing Popular Game Titles with a Hybrid Brain-Computer Interface. *IEEE Transactions on Computational Intelligence and AI in Games,* 2012. *(Submitted)*

3.4. User Study Design

This section describes in detail the design of the conducted user study. The major goals of this study are:

1. Capturing the wearing comfort of the consumer EEG headset EPOC over a long period of time.
2. Evaluating if it is possible for subjects to gain control over a virtual cube with the Emotiv Expressiv Suite.
3. Verifying if the Emotiv Affectiv Suite is capable of estimating the state of a player.
4. Assessing the reaction (premotor + motor) time and accuracy of facial expressions.
5. Validating if the interference matrix (explained in Chapter 3.4.10) can be applied in order to identify the most reliable and unreliable expressions.
6. Testing the Emotiv headset integration of the GBGC (see Chapter 3.1.2) with multiple users.
7. Capturing the user experience for the Emotiv EPOC as an input device for controlling a popular game application.
8. Assessing the usability of the gyro re-centering procedure using the bordered mode of the GBGC (described in Chapter 3.1.3).

The user study is not meant to be an exhaustive test that investigates the entire functionality provided by the Emotiv EPOC headset. Instead, a set of representative experiments was chosen to get an overview of the potentials of this EMG, EEG, and gyro sensor based input device. The study was designed with respect to the World Medical Association (WMA) declaration of Helsinki in the version of 2008 [1].

3.4.1. Structure

The study procedure is divided into five major parts (see Figure 3.4.1):

1. Introduction & Preparation.
2. Cognitiv Test.
3. Affectiv Test.
4. Expressiv Test.
5. Experience Test.

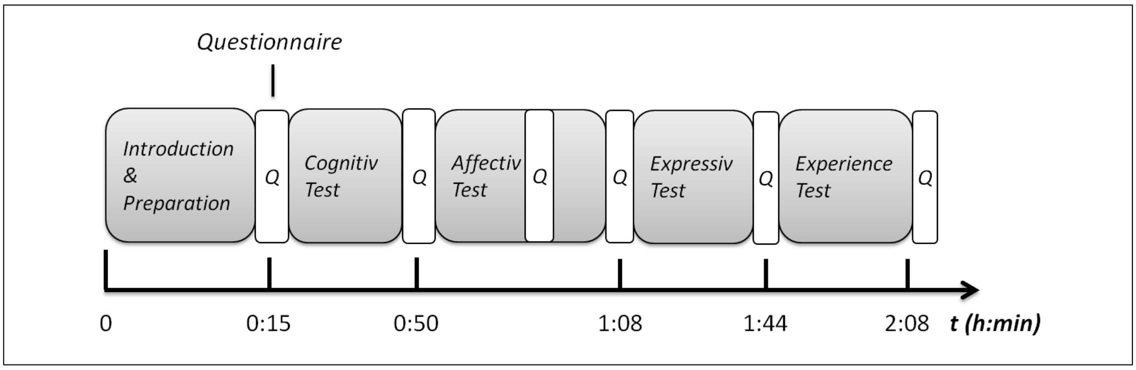


Figure 3.4.1.: The user study procedure, showing the average durations of each individual part of the user study.

In the first part, the participant is introduced to the study conditions. After mounting the EPOC on the subject's head, the Cognitiv Suite of the Emotiv Control Panel is used to determine if a new user can gain control over a virtual object. The Cognitiv Test is followed by the Affectiv Test. This part was designed to assess if the values for various emotions measured by the Affectiv Suite relate to the actual affective state of a subject. In the next part of the user study, the Expressiv Suite is used to investigate the detection accuracy for various facial expressions. To assess the reaction times of facial expressions, the premotor and motor times are recorded during the Expressiv Test. The study is concluded by an Experience Test which investigates if the Emotiv EPOC can be used to conveniently control a popular game application hands free–just using head movements and facial expressions. Between each part of the user study, a pen and paper questionnaire is used to assess the wearing comfort of the headset and the subject's state.

3.4.2. Questionnaires

The questionnaires used in this study are based on [26, 84]. The base questionnaire consists of seven different questions that had to be answered using a 10 cm long visual analog scale (VAS) ranging from 0 to 10 (see Figure 3.4.2). VASs were chosen over rating scales because of the findings of [28]. For a complete listing of the German questionnaires, see Appendix B.

The questions, which were given to the participants after each part of the study, are:

1. Do you feel relaxed? [0 = very excited; 10 = extremely relaxed]
2. Do you feel alert? [0 = very sleepy; 10 = extremely alert]
3. How is your motivation? [0 = not motivated at all; 10 = extremely motivated]
4. How is your mood? [0 = extremely bad mood; 10 = extremely good mood]
5. How does the headset feel on your head? [0 = extremely disturbing; 10 = not disturbing at all]

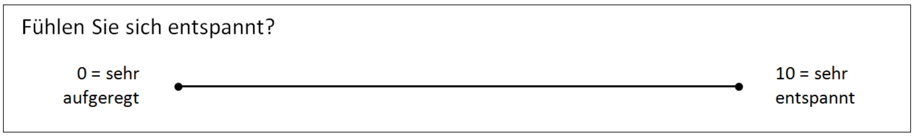


Figure 3.4.2.: The VAS style used in the questionnaires of the user study.

6. Does wearing the headset cause pain? [0 = extremely painful; 10 = not painful at all]
7. Was the task demanding? [0 = extremely demanding; 10 = not demanding at all] (not included in the Introduction & Preparation part of the study)

After each specific part of the user study the following additional questions were included:

1. Do you have the feeling that you could control the cube? [0 = no control; 10 = total control] (after Cognitiv Test)
2. Was the task hard for you? [0 = extremely hard; 10 = extremely easy] (after Experience Test)
3. How did you perceive the movement controls of your avatar? [0 = no control; 10 = total control] (after Experience Test)
4. How hard was it to re-center the gyro sensors using the border markers? [0 = extremely hard; 10 = very easy] (after Experience Test)
5. How did you perceive the control over the actions (target, attack and loot)? [0 = no control; 10 = total control] (after Experience Test)

3.4.3. Mehrdimensionaler Befindlichkeitsfragebogen

In the Affectiv Test (see Chapter 3.4.9) an additional questionnaire, called “Mehrdimensionaler Befindlichkeitsfragebogen” (MDBF) [84], was used to assess the affective state of a subject. The MDBF is a scientifically established questionnaire in German, which investigates three mood dimensions. These dimensions are:

1. “Gute-Schlechte Stimmung” (GS) (good-bad mood).
2. “Wach-Müdigkeit” (WM) (awake-tired).
3. “Ruhe-Unruhe” (RU) (calm-excited).

The possible values for all three dimensions range from 8 to 40. The MDBF used in this study is shown in Appendix B.

To quantify how playing two different games (QWOP and High Delivery, described in Chapter 3.4.9) affected the subjects, the questionnaire had to be answered three times: before playing the first game, between the two games, and after playing the second game.

CPU	Intel Core i5 661, quad core 3,33 GHz
Memory	8 GB, 1600Mhz
GPU	NVIDIA GeForce GTX 580
Display	Acer GD245HQ (1920x1080, 120 Hz)

Table 3.4.1.: The technical specifications of the PC used in the user study.

3.4.4. Subjects

For the study only healthy people were accepted. The age of the subjects ranged from 23 to 32 (mean age 28.6 years, 8 males, 2 females). No participant was taking medication that could affect his or her reaction time or cognitive function. None of the subjects drank alcohol 24 hours in advance of the study and only two out of ten subjects were smokers. Each subject stated that they had slept the usual (or more) amount the night before the study (mean hours of sleep 7.65). All subjects had normal or corrected-to-normal vision using contact lenses or glasses. Additionally, the subjects were asked not to use any hair styling products or lotions since these items can influence the EEG signal.

3.4.5. Room and Hardware Setup

The room setup used in this study (see Figure 3.4.3 (a)) was inspired by usability testing methods [3]. To avoid any distractions, the subject was seated in a position where only the required display could be seen and the window did not cause any glare on the screen. The investigator was able to see both displays that were used during the study. The subject's display was installed on a Neo-Flex LCD arm by Ergotron (see Figure 3.4.3 (b)). This LCD mount provides great flexibility and it is possible to adjust the LCD screen's position in any direction.

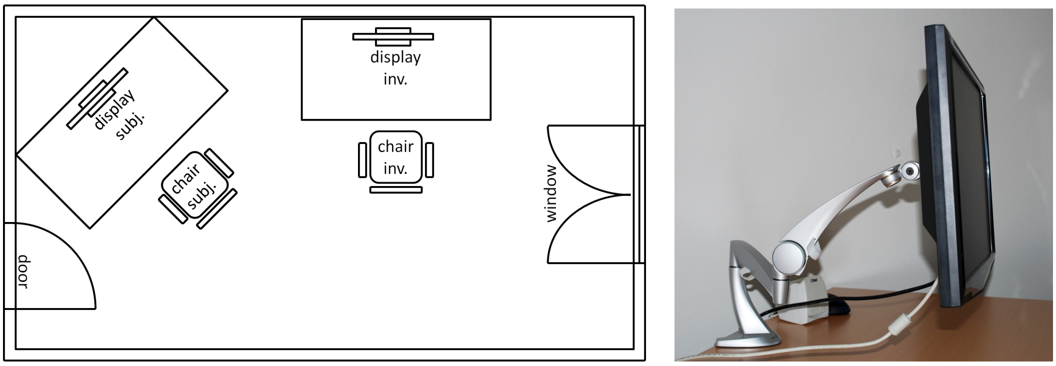
[10, 87] state that the Emotiv EPOC USB dongle uses a low power Bluetooth transmitter. This transmission technology can cause a bad connection even if the headset is fully charged. To avoid this problem, a USB extension cable was used to mount the dongle directly on the backrest of the subject's chair. This brings the dongle as close to the headset as possible.

The PC which used for testing ran a Microsoft Windows 7 (64-bit, Service Pack 1) operating system (OS). The detailed technical specifications of this PC are listed in Table 3.4.1.

3.4.6. Emotiv EPOC Test Suite

For providing a clean user interface and storing the measured data, a Test Suite for the Emotiv EPOC headset was implemented. This application runs on Microsoft Windows 7 and is written in Visual C++. The signals from the Emotiv EPOC headset are acquired by the Emotiv Research Edition SDK v1.0.0.5. The GUI uses the Windows Forms framework (see Figure 3.4.4).

For pilot testing purposes all timing parameters are adjustable via GUI entries. During



(a) (b)

Figure 3.4.3.: (a) A map of the room used in the user study, showing the locations of the door, window, chairs, and displays. (b) Shows the Neo-Flex LCD arm by Ergotron. This LCD mount provides great flexibility and it is possible to adjust the LCD screen's position in any direction.

a test the investigator receives live feedback from a message queue to check if the recorded data is valid.

3.4.7. Preparation & Introduction

The Emotiv headset was fully charged and the electrodes were mounted with a new set of felt sensors for each test session. To respect hygiene guidelines and to maintain the same study conditions for each participant, no felt pads were reused. The two rubber ground electrodes were cleaned and sterilized with surgical alcohol prior to each session. The felt sensors were moistened using the same multi-purpose contact lens cleaning solution as the manufacturer delivered in the original Emotiv EPOC package (Bausch&Lomb ReNu MPS Sensitive Eyes, see Figure 3.4.5).

The test subjects were welcomed and introduced to the testing location by the investigator. The experiment conductor was also in charge of providing small refreshments and non-caffeinated drinks during the study, so that the subjects could feel as comfortable as possible. To avoid interferences in the EEG signal or the wireless connection of the Emotiv EPOC mobile phones and other devices using WLAN or Bluetooth were turned off or removed from the testing location prior to the study. The headset was shown and explained to each subject before they received a written explanation of the user study. An informed consent form had to be signed by each participant in order to document that he or she understood the explanation document and agreed with the study conditions. The informed consent is included in Appendix A.

The display and chair were adjusted so that each subject could sit in a relaxed position and maintain a good view on the display. The headset was placed on the head of each subject according to the user manual [24] of the manufacturer. A new user profile was created with the subject's user code in the Emotiv Control Panel and the Emotiv EPOC

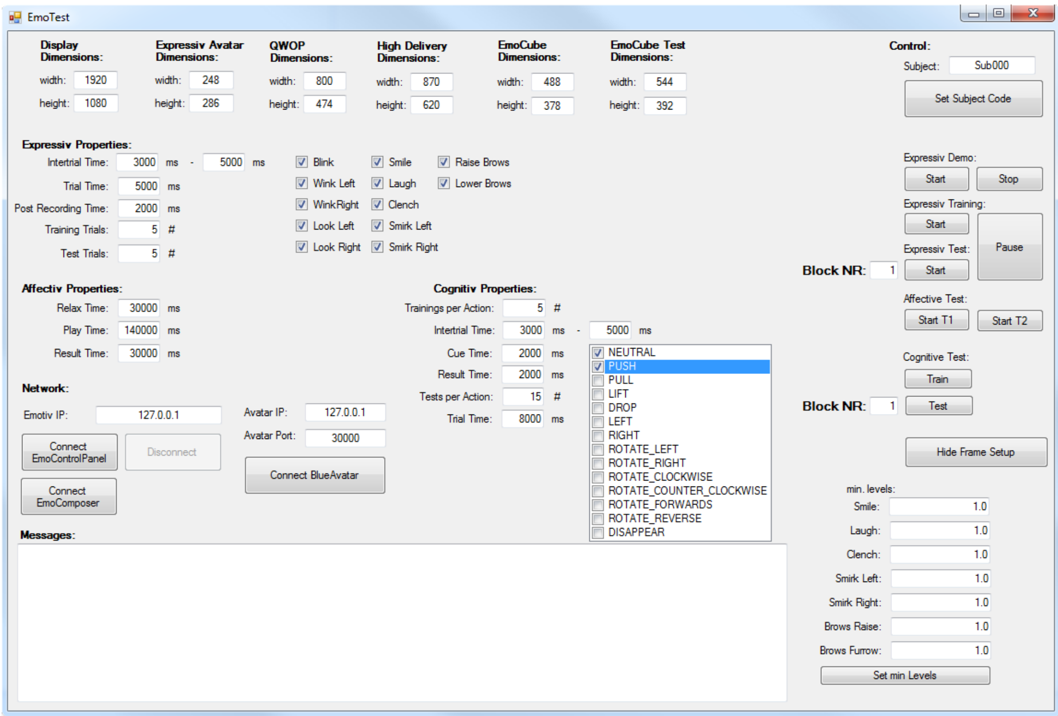


Figure 3.4.4.: The GUI of the Emotiv EPOC Test Suite. For pilot testing all time-based parameters are adjustable. The Test Suite provides clean user interfaces for the user study. Only interface elements that are used for the study are displayed on the subject's screen. On the top of the window, the screen resolution and the size of the external applications can be adjusted. The external applications are displayed in the middle of the screen and are framed with a clean white border by the Test Suite. Emotiv provided the Blue Avatar and EmoCube application. High Delivery and QWOP are Flash games running in a browser window which is also hidden by the Test Suite. The buttons to invoke the various tests are located to the right of the window.

(a) (b)

Figure 3.4.5.: The contact lens cleaning solution is used as a contact agent for the Emotiv EPOC headset. The felt pads of the electrodes need to be moistened using this saline, disinfectant solution. (a) Shows the solution provided by Emotiv and (b) shows the one used in the user study.

Test Suite (see Chapter 3.4.6), so that each user had his own personal profile in both applications. All 16 electrodes of the headset were readjusted until they had good contact with the subject's scalp. A good contact quality is indicated by a green dot in the Emotiv Control Panel (see Figure 3.4.6).

3.4.8. Cognitiv Test

The Cognitiv Test of the user study investigates how easy it is for new users to learn to control a virtual cube only using brainwave activity. The manual of the Emotiv Research Edition SDK version 1.0.0.5 [24] describes the Cognitiv Suite (see Figure 3.4.7) as follows:

> "The Cognitiv detection suite evaluates a user's real time brainwave activity to discern the user's conscious intent to perform distinct physical actions on a real or virtual object. The detection is designed to work with up to 13 different actions: 6 directional movements (push, pull, left, right, up, and down) and 6 rotations (clockwise, counter-clockwise, left, right, forward, and backward) plus one additional action that exists only in the realm of the user's imagination: disappear. Cognitiv allows the user to choose up to 4 actions that can be recognized at any given time. The detection reports a single action or neutral (i.e. no action) at a time, along with an action power which represents the detection's certainty that the user has entered the cognitive state associated with that action ... Almost all new users readily gain control over a single action quite quickly. Learning to control multiple actions typically requires practice and becomes progressively harder as additional actions are added."

For the study, only one action (lift) was selected for the cognitive training. The subjects were given the task to move a virtual cube floating in a box (see Figure 3.4.8) by just

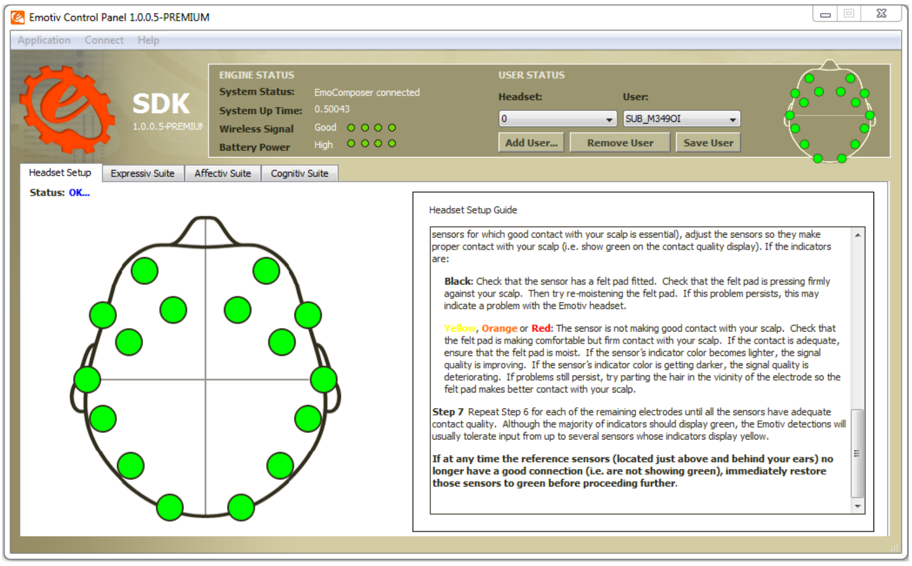


Figure 3.4.6.: The Emotiv Control Panel view, showing the contact quality of the 16 sensors. A green dot shows that the sensor has the best possible contact to the scalp of the subject's head.

using their brainwave activity.

Cognitiv Training and Test

Emotiv does not include any detailed information in the user manual of the EPOC headset on how the cognitive actions can be trained appropriately. The necessary information was extracted from postings of Emotiv officials in the support forum. Gmac (Emotiv support forum administrator) states [86] that at least five to six trainings per state are required to get a good distinction between two different states (neutral and lift).

To investigate if the performance increases with more training trials and practice, the Cognitiv Test was divided into two parts: training and test. In the first part, five trainings were performed for each state (10 trainings in total), followed by 15 test trials per state (30 in total). After clearing the training data, six trainings were used per action (12 in total) in the second part, which were again followed by 15 test trials per state (30 in total).

Although [55] states that at least 50 test trials are necessary for each state to calculate the performance of such a system, only 30 test trials per state were recorded due to overall time constraints. The detailed trial setup is depicted in Figure 3.4.9.

Instructions

For the Cognitiv Test, detailed information was extracted for the subjects from the Emotiv forums [85, 86]. The following instructions were given to the participants prior to the Cognitiv Test:

1. The EEG signal is used for detecting activations in certain brain regions.

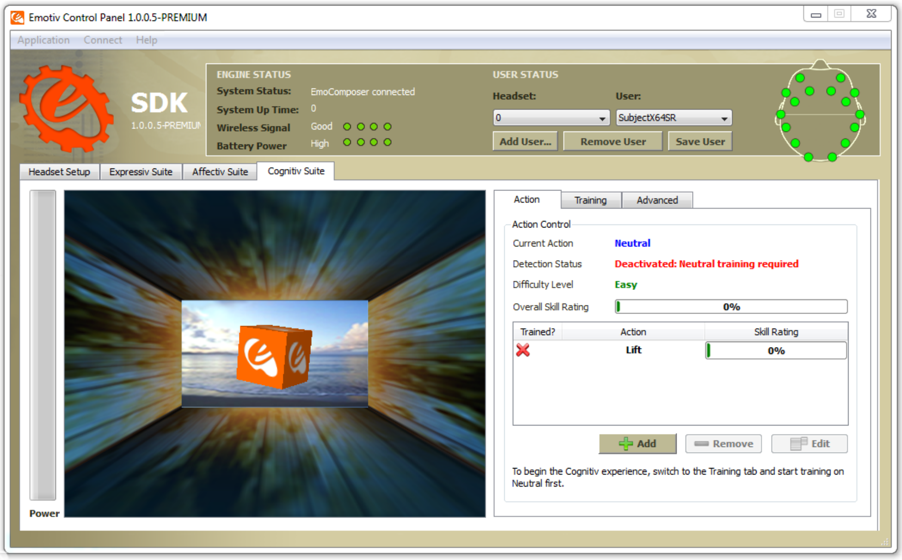
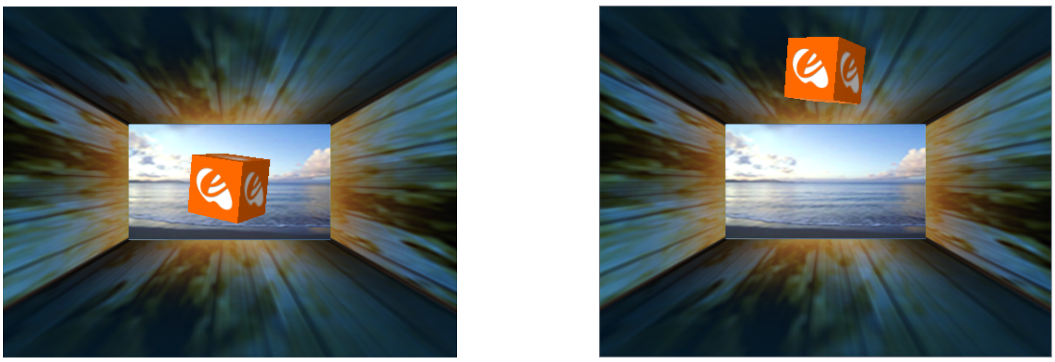


Figure 3.4.7.: The Emotiv Cognitiv Suite showing the actions, which can be applied to a virtual cube. A training progress for each activated action is shown in the right part of the window.



(a) (b)

Figure 3.4.8.: The virtual cube provided by Emotiv. It is used to get online feedback while performing cognitive actions. (a) shows the cube in a neutral position. The subjects were asked to try to lift the cube (b) by just using brainwave activity.

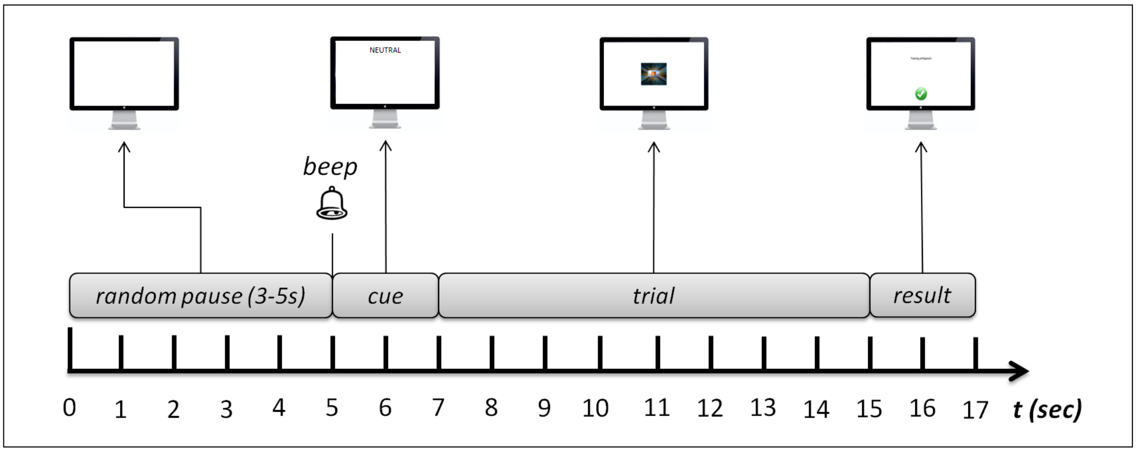


Figure 3.4.9.: The trial setup used in the Cognitiv Test for training and testing. The subjects were asked to relax during a random pause (3–5 sec). They should also try to swallow and blink only during this pause as much as possible. The cue, indicating the next cognitive action to perform, was chosen at random. During the trial time (8 sec), the virtual cube was displayed on screen, and depending on the previously given cue the subject can try to either lift the cube or hold it in a neutral position. During the result time (2 sec) a visual feedback was presented indicating if the training or test trial was successful or not.

2. To produce an appropriate activation you have to imagine a certain scene.
3. It is *not* important what this scene looks like. The system learns to recognize the activation pattern in a training phase. It is necessary that the imagined scene is constant (greater 8 sec) and reproducible.
4. Try not to flex any muscles! Especially avoid movements of the head, tongue, mouth, and jaw.
5. Try to truly relax during the imagination. Do not try to force it! This can cause muscle contractions which again will influence the EEG signal.
6. For the neutral state do not think of anything in particular. During the lift state, try to lift the virtual cube.
7. For lifting the cube you could think of ...
    - squeezing a fluid through a tube in front of your eyes.
    - your whole body beginning to float.
    - where the cube should be located.

After the instructions were explained to the subjects they were given two minutes time to relax. During this period the participants could choose and practice a good scene for their imagination to lift the virtual cube.

3.4.9. Affectiv Test

The user manual [24] for the EPOC headset describes the Affectiv Suite (see Figure 3.4.10) as follows:

> "The Affectiv Suite reports real time changes in the subjective emotions experienced by the user ... The Affectiv detections look for brainwave characteristics that are universal in nature and don't require an explicit training or signature-building step on the part of the user."

The Emotiv EPOC Affectiv Suite measures continuous values for five different emotions: engagement/boredom, frustration, meditation, instantaneous excitement, and long-term excitement. By frustrating the subjects on purpose the Affectiv Test examines if these measured values, in particular frustration, can reflect the affective state of a subject.

Frustration

[29] describes frustration in a gaming context as follows:

> "Frustration is that which arises when the progress a user is making towards achieving a given goal is impeded."

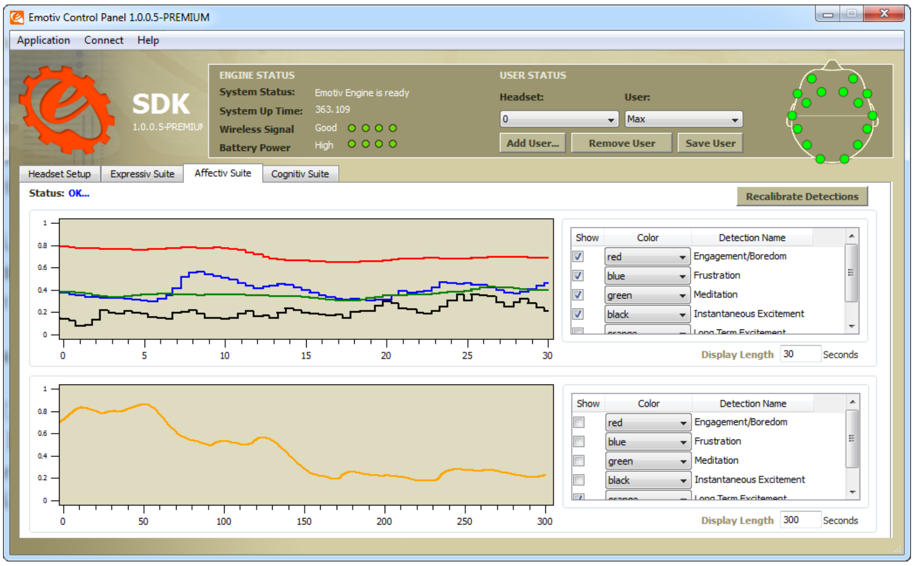


Figure 3.4.10.: The Emotiv Affectiv Suite showing the continuous graphs for engagement/boredom (red), frustration (blue), meditation (green), instantaneous excitement (black), and long-term excitement (orange).

To induce frustration the same approach as in [70] was used. Each participant was given a gaming task that could not be solved by the player. Since it was nearly impossible for the player to reach the goal in time it ensured frustration.

To get comparable results, the subject was "brought back" to the initial state by a second, manageable gaming task. This time the player could easily complete the task.

Procedure

The initial state of a subject was captured using a pen and paper questionnaire including the MDBF (see Chapter 3.4.3) and two VASs (see Chapter 3.4.2). After the subject answered the questionnaire, frustration was induced with the game QWOP. While the subject was playing the game, the Affective Suite was used to record the values of the five emotions. Afterwards, the subject was asked to answer the same questionnaire as in the beginning. Finally, the subject was given the second manageable gaming task in a game called High Delivery. The values for the emotions were also recorded during the play time, and the effect on the state of the players was again captured by a questionnaire.

The same trial setup was used for playing QWOP and High Delivery. It is shown in Figure 3.4.11.

QWOP

In the first game QWOP (see Figure 3.4.12(a)) the thighs and calves of an athlete had to be controlled by the player. The subjects were given the task to run ten meters of distance with their athlete in 2:20 min (time determined by the fixed duration of High Delivery). The athlete is very fragile and easily falls to the ground. If any other body parts other than the feet of the athlete hit the ground, the game is over and has to be

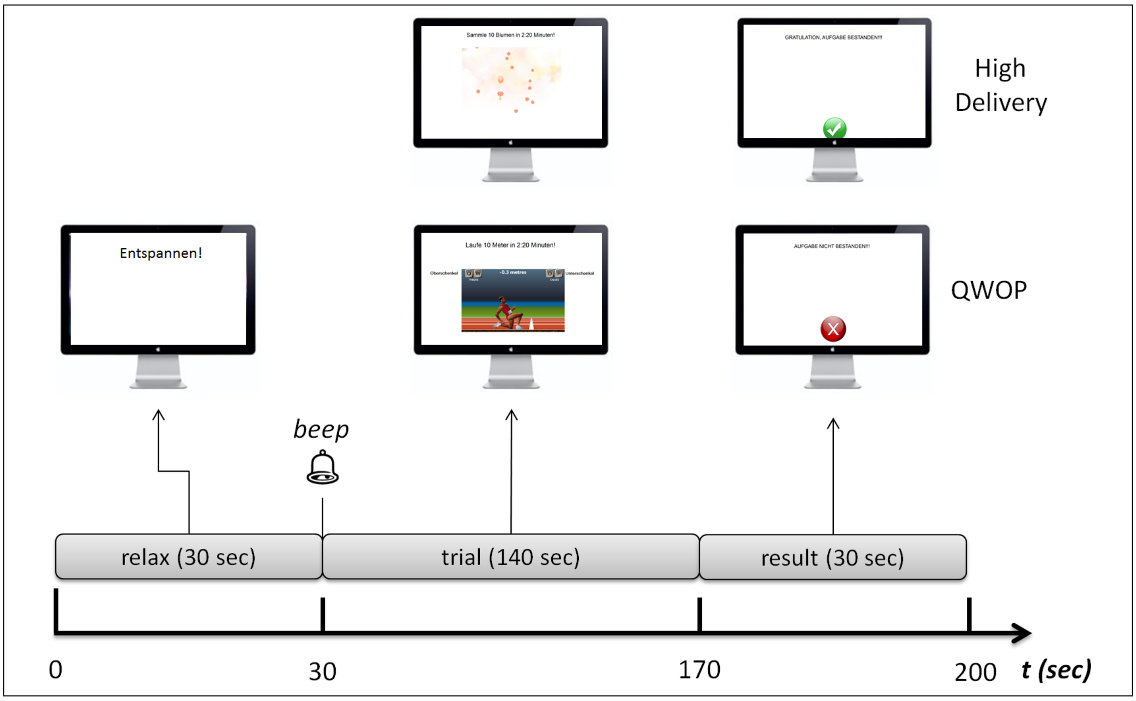
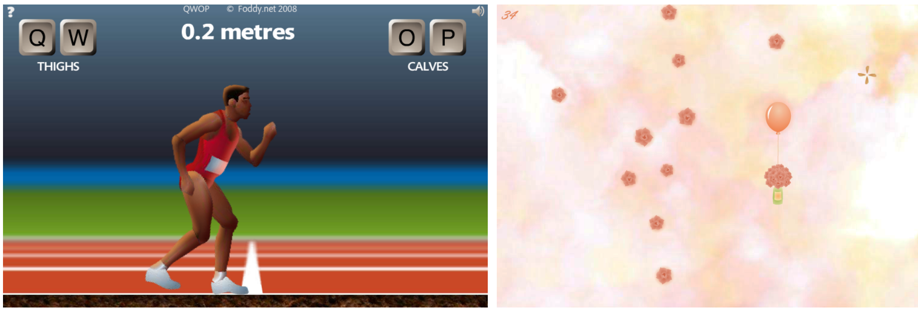


Figure 3.4.11.: The trial setup used in the Affectiv Test of the user study. After a 30 sec relax time the subjects had 2:20 min time to complete the given game task. The subjects received negative feedback after the QWOP task and positive feedback after the High Delivery task. During the result time, the subjects were advised to relax again, but this time they should think about their performance in the game task.



(a) (b)

Figure 3.4.12.: The games used in the Affectiv Test of the user study. (a) Shows a game called QWOP. The player controls the thighs (buttons Q, W) and calves (buttons O, P) of an athlete to run down a track as far as possible. The achieved distance is shown on the top of the screen. In the second game, High Delivery (b), the player has to catch flowers with a bottle hanging from a balloon. The mouse pointer is represented by a small fan which can blow the balloon in different directions. The amount of flowers caught by the player is displayed in the upper left hand corner of the screen.

restarted by pressing the space bar. For a player that has never played this game before it is very hard to learn how to make the athlete run the distance within the designated time frame. The timing and sequence of button presses has to be trained to get better control over the athlete.

High Delivery

In the second task of the Affectiv Test, the subjects had to collect ten flowers within the game High Delivery (see Figure 3.4.12(b)). The flowers drop from the top of the screen and have to be caught with a bottle hanging from a balloon. The balloon can be blown in any direction with a small fan. The player controls the position of the fan with the computer mouse. The task was chosen in a way that each subject could reach the given goal easily. The game ends automatically after 2:20 min.

Compared to QWOP, High Delivery has very easy controls. It is also accompanied by a very relaxing tune which plays throughout the whole game. The flowers that have to be caught by the player drop at a very slow speed.

3.4.10. Expressiv Test

In the Expressiv Test the accuracy of recognizing facial expressions using the Emotiv EPOC headset was investigated. Additionally, the premotor and motor times of facial expressions were recorded. The Expressiv Test was again divided into three parts:

1. Demo.

2. Training.

3. Test.

Approach
The Emotiv Expressiv Suite (see Figure 3.4.13) can detect 12 different facial expressions (for a full listing, see Table 3.4.2). Since the headset is able to detect many expressions the accuracy for each expression is very low. It would not be feasible to calculate the accuracy using a cue-paced experiment where the requested expression has to occur as the next classification.

To get better accuracy for the facial expressions, [9] combined multiple expressions into groups: eyeball movements (look left, look right), frontal muscle movements (raise brows, blink, wink left, wink right), and jaw muscle movements (smile, clench). Their results show that the EPOC device is able to distinguish between groups of muscles, but the classifications of the exact muscle motion and corresponding facial expression can be inaccurate.

Accuracy can also be raised by finding the optimal settings for the sensibility sliders (see Figure 3.4.13). These sliders need to be adjusted to reduce the miss-classification rate for some highly detected expressions. According to [9], the average classification rate can be increased 12% by adjusting these sliders. To find the optimal settings, a user can only use a trial and error approach.

In this work an automated approach to find these interfering expressions is proposed with the concept of an interference matrix (see Table 3.4.5). The highly detected expressions can easily be identified and then be either excluded, adjusted in sensitivity or grouped to get a better accuracy.

Demo
Before the Expressiv Test the subjects were introduced and familiarized with the set of facial expressions, which can be detected by the Emotiv EPOC headset. The Emotiv EPOC Test Suite (see Chapter 3.4.6) was set to a demonstrative mode where only the avatar, representing the user's facial expressions, is displayed (see Figure 3.4.14). The investigator showed the subject a card with the name of an expression and the corresponding avatar face (see Figure 3.4.15). The subject was then asked to try to recreate the expression shown on the card. After three to four attempts to recreate the expression, the investigator drew the next card and the procedure was repeated until all of the 12 expressions (for a complete list see Table 3.4.2) were performed by the subject.

Training
In the next step of the Expressiv Test, the subjects were asked to try to recreate the facial expressions indicated by a randomly chosen cue. The subjects were aware that speed was relevant but not as important as accuracy. The results of [31, 55] indicate that at least 10 practice trials are necessary for healthy subjects to learn a movement procedure in simple electromyographic reaction time (ERT) studies. Based on these results, the Expressiv Test was divided into one training block and two testing blocks. Each block consisted of five trials per facial expression, making up a total of 60 trials for one block.

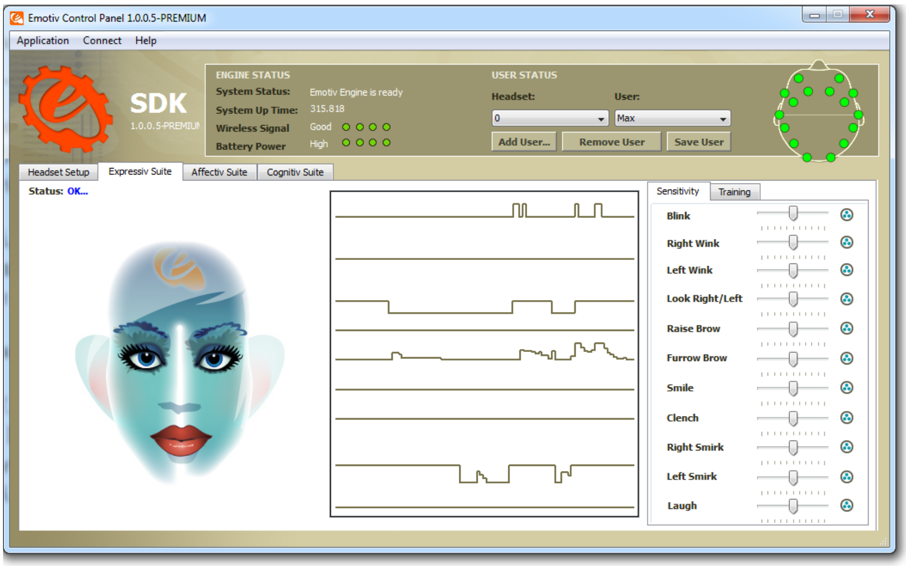
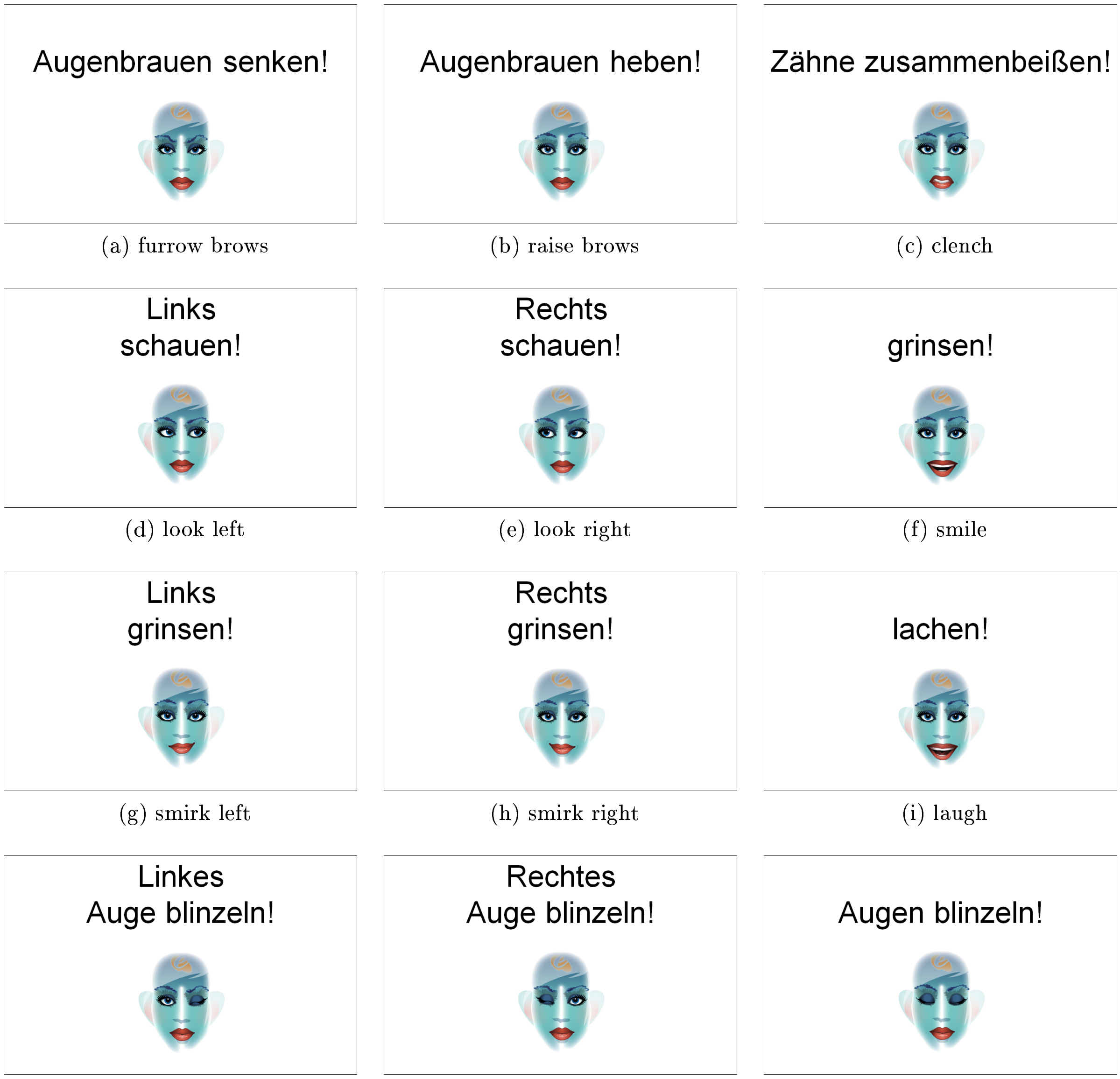


Figure 3.4.13.: The Emotiv Expressiv Suite showing the graphs for the detection and power of facial expressions over time. The sliders to the right adjust the sensitivity for the detections. In the user study the sensitivity sliders were set to their default position. To capture the detection accuracy with the default settings of the Emotiv EPOC Control Panel it is necessary to leave the sliders at the default settings. This also ensures that the results are comparable between subjects.

Figure 3.4.14.: The screen setup of the demonstration mode. The Test Suite shows only the avatar, which represents the subject's facial expressions currently detected by the Emotiv Expressiv Suite.



(a) furrow brows (b) raise brows (c) clench
(d) look left (e) look right (f) smile
(g) smirk left (h) smirk right (i) laugh
(j) wink left (k) wink right (l) blink

Figure 3.4.15.: The cards used for familiarizing the subjects with the 12 different facial expressions (a–l).

eyes	upper face	lower face
blink	furrow brows	smile
wink left	raise brows	laugh
wink right		clench
look left		smirk left
look right		smirk right

Table 3.4.2.: List of facial expressions detectable by the Emotiv EPOC Expressiv Suite grouped by their API function.

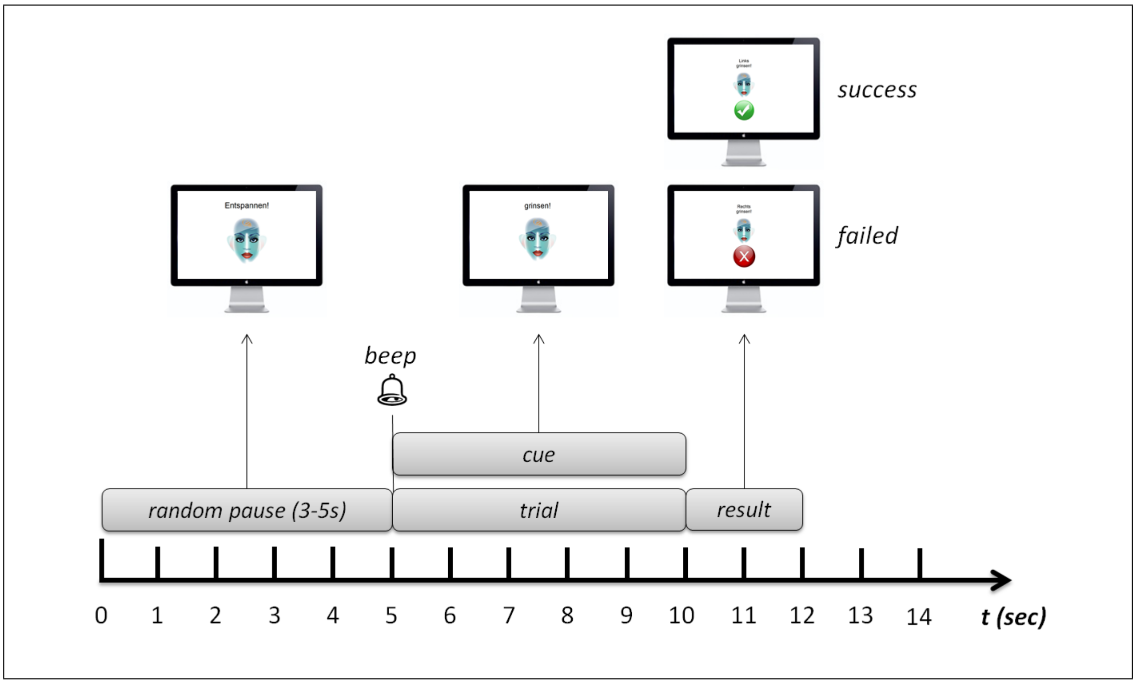


Figure 3.4.16.: The Expressiv Test training and test trial setup, showing the timings and screens of each trial phase.

Trial Setup

The training trial setup is shown in detail in Figure 3.4.16. At the beginning of each trial the subject was asked to relax the facial muscles for a random period of three to five seconds. The participants were advised to try to return to a neutral state with their avatar during this time. The timeout was randomized so that a subject could not prepare for the upcoming cue. The random pause was followed by a beep (700 Hz, 100 ms) and a cue indicating the name of a randomly chosen facial expression. During the trial time the subject could try to recreate the given facial expression. If the subject succeeded in creating the right signal during a five-second trial timeout, the recording of the motor time was started. The time between the trial start and the motor time was considered as premotor time. As long as the power of the signal for the requested facial expression was on the rise or equal to the previous sample, the time was considered as motor time. As soon as the signal power was lower than the previous sample, the recording of the motor time was stopped and a positive feedback was displayed on the screen. After a two-second result timeout (or post-recording time) the Expressiv trial was finished. A negative feedback was displayed if the subject was not able to perform the right facial expression or the headset was not able to detect it during the five-second trial time.

Minimum Power Levels

The values provided by the API of the Emotiv headset (see Listing 3.3) for the power level of the facial expressions are either binary (blink, wink left, wink right, look left, look right) or of floating point precision between 0.0 and 1.0 (clench, smile, laugh, smirk left, smirk right, brows furrow, brows raise). During the training procedure, the lowest maximum power level for each expression was determined. These minimum levels were used later on in the testing portion of the Expressiv Test to detect interferences of other signals. Table 3.4.4 and Table 3.4.3 show sample results of the expressive training.

trial no.	type	complete	premotor time (ms)	motor time (ms)	max. power
0	wink left	false	5000	0	0
1	blink	true	795	177	1
2	look left	true	1026	482	1
3	laugh	true	2553	148	0,698156
...	...	...	...	...	...

Table 3.4.3.: The data format of the recorded test results for each individual trial. The trial number indicates the order of the randomly chosen trial types. The premotor and motor times are recorded in milliseconds. The maximum power is the highest power level reached during the motor time.

```
//! Returns the detected upper face Expressiv action power of
    the user
/*!
     \param state - EmoStatehandle
     \return power value (0.0 to 1.0)
     \sa ES_ExpressivGetUpperFaceAction
*/
EMOSTATE_DLL_API float ES_ExpressivGetUpperFaceActionPower(
    EmoStateHandle state);

//! Returns the detected lower face Expressiv action power of
    the user
/*!
     \param state - EmoStatehandle
     \return power value (0.0 to 1.0)
     \sa ES_ExpressivGetLowerFaceAction
*/
EMOSTATE_DLL_API float ES_ExpressivGetLowerFaceActionPower(
    EmoStateHandle state);
```

Listing 3.3: The documentation of the ES_ExpressivGetLowerFaceActionPower and the ES_ExpressivGetUpperFaceActionPower API functions from the EmoStateDLL.h of the Emotiv EPOC SDK.

Test

In this part of the Expressiv Test, the subjects were again asked to recreate facial expressions chosen at random by the Emotiv EPOC Test Suite. This time, the subjects were informed that they should perform the requested expressions as fast as possible, but still focus on accuracy rather than speed. The trial setup was the same as the Expressiv Test training trial setup (see Figure 3.4.16). The test part included two blocks, each consisting of five test trials per expression, making up a total of 60 test trials per block.

expression	**completed**	**min. power**
blink	5	1
wink left	5	1
wink right	5	1
look left	5	1
look right	4	1
smile	1	0.123391
laugh	5	0.802884
clench	5	0.0862337
smirk left	5	0.766055
smirk right	5	0.435859
raise brow	5	1.0
furrow brow	5	0.258221

Table 3.4.4.: The data format of the recorded test results for all trials showing the number of successfully completed trials for each expression. Additionally, the lowest maximum power for each expression is determined. This minimum power is the lowest maximum power of all completed trials for one facial expression type. For example, the expression “smile” has a minimum power of 0.340529, which means that this was the lowest maximum power occurring during the five completed trials.

A short break between each block gave the subjects the opportunity to refresh and focus on the next task.

The results of the test consist of an individual file per trial showing all recorded input signals, a result table equal to Table 3.4.3 and an interference matrix for each block. The interference matrix is very much like a confusion matrix and shows how often another signal type interferes during a test trial. In case the power of a signal, other than the one currently requested by the cue, became greater than or equal to the minimum level (calculated during the training) of the requested expression, it was counted as interference and the counter in the matrix increased by one. These counters were stored in a 12 x 12 matrix, with each line and column representing one facial expression. The requested expressions are displayed in the lines of the matrix, and the expressions interfering the requested expressions are displayed in the columns of the matrix (see Table 3.4.5).

	blink	wink left	wink right	look left	look right	raise brow	fur. brow	smile	clench	laugh	smirk left	smirk right	∑
blink	-	0	2	0	0	0	1	0	0	0	0	0	3
wink l.	2	-	0	0	0	0	1	0	0	0	0	0	3
wink r.	5	2	-	0	0	0	2	0	0	0	0	0	9
look l.	3	1	2	-	3	0	7	0	0	0	0	0	16
look r.	6	2	9	1	-	0	20	3	0	0	0	0	41
r. brow	1	0	0	0	0	-	0	0	0	0	0	0	1
f. brow	1	0	0	0	0	0	-	0	0	0	0	0	1
smile	6	0	0	0	0	0	0	-	0	6	6	1	19
clench	2	0	0	1	0	0	0	1	-	0	0	0	4
laugh	3	2	1	0	0	0	1	1	0	-	0	1	9
smirk l.	1	0	1	0	0	0	1	0	0	0	-	0	3
smirk r.	2	0	0	0	0	0	2	0	0	1	1	-	6
∑	32	7	15	2	3	0	35	5	0	7	7	2	

Table 3.4.5.: The sample interference matrix of subject EE01F (block 2). It is a result of the test part in the Expressiv Test. It depicts the reliability of each facial expression and how they are influenced by other expressions. The summed-up values in the last line show how often the expressions interfered other expressions. A high value indicates that the expression was detected very often although its performance had not been requested. The summed-up values in the far-right column indicate how often an expression has been interfered by other expressions, even it was requested by the cue. A high value indicates that many other expressions were detected before the requested expression was recognized. In this example, look right received the most interference. Raise brow and furrow brow were interfered only once each. In contrary to raise brow, furrow brow interfered with many other expressions. Based on these results, raise brow is the most reliable expression for this particular subject.

3.4.11. Experience Test

The final portion investigated the possibility to play a modern game title hands–free by just using the Emotiv EPOC headset and the GBGC. This assessment was designed to answer the question whether performance can be increased by selecting the most reliable facial expressions identified by the interference matrix (see Chapter 3.4.10). User feedback was solicited regarding the usability of the gyro re-centering procedure that relied on the bordered mode of the GBGC and the visual feedback (described in Chapter 3.1.3).

World of Warcraft

In this part of the user study, the subjects had to control the popular MMORPG World of Warcraft (Blizzard, Inc., see Figure 3.4.17 and Figure 3.4.18) by using the gyro sensors and a selection of three facial expressions. The gyro sensors were used to control the movements of the avatar. By turning the head to the right, the subject could turn the avatar to the right and when turning the head to the left, the avatar also turned left. Depending on how far the subject rotated the head from a neutral position, the avatar rotated slowly or quickly. To run forward, the subject had to lift his head. There was no command for going backwards with the avatar since the same effect can be achieved by turning around.

(a) (b) (c)

Figure 3.4.17.: The logos of the successful MMORPG World of Warcraft (a) and the producer and publisher, Blizzard, Inc. (b). (c) Shows the box arts of the first version of WoW released on November 23, 2004.

WoW Add-ons

For user feedback, two WoW add-ons were created to extend the default UI of the game (see Figure 3.4.18). One of them displays the current position of the gyro sensors in the virtual 2D space (see Figure 3.4.19, described in detail in Chapter 3.1.3), and the second interface extension displays the facial expressions (see Figure 3.4.20) used to trigger three different in-game actions (target, attack, and loot).

Control Groups

To investigate if the interference matrix is capable of detecting the most reliable expressions, the subjects were divided into two groups: A and B. To perform three different

(a)

(b)

Figure 3.4.18.: The WoW user interface. (a) Shows the UI without any modifications and additional add-ons. (b) Shows the UI with the add-ons for the Emotiv EPOC headset loaded. The add-on below the avatar's feet displays the current position of the gyro sensors and the add-on to the right shows the facial expressions that trigger the in-game actions (target, attack and loot).

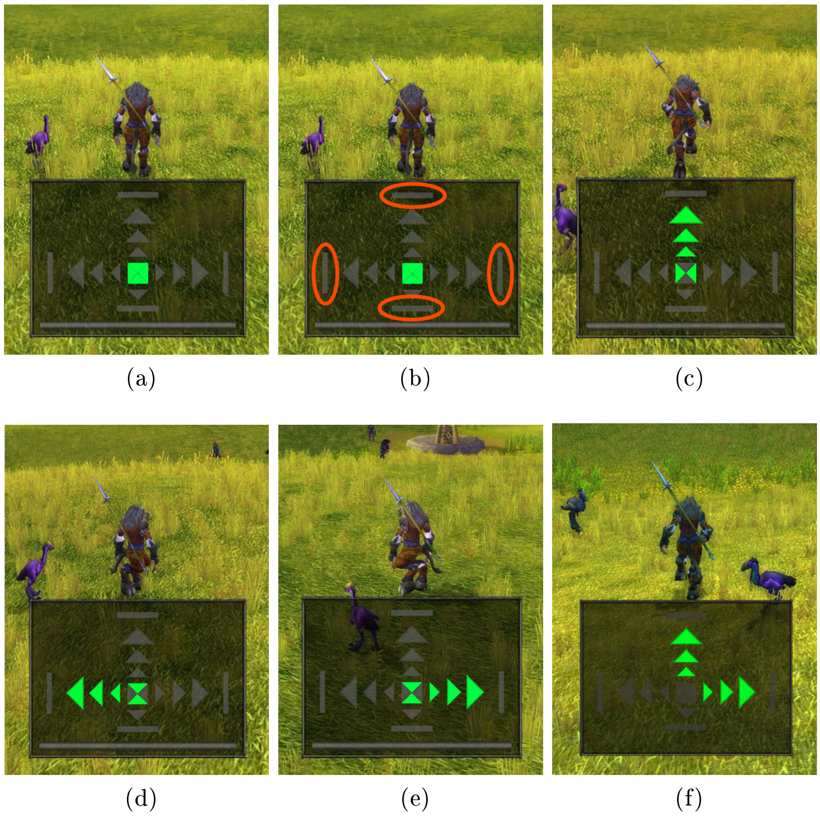


Figure 3.4.19.: The WoW add-on, displaying the current position of the gyro sensors. (a) Shows the gyro sensors in a centered position. (b) Shows the border markers (circled in red) that turn red in case the subject reaches the borders of the virtual gyro box (described in Chapter 3.1.3). These markers are helpful to re-center the virtual gyro position without an extra command. (c) Shows the add-on indicating that the users's head is lifted upwards and the avatar is moving forwards. (d–e) Show the add-on displaying that the avatar is rotating left/right. The subject can turn the character slowly by gently rotating the head. The further the head is rotated, the faster the character will turn and more arrows will turn green. (f) Represents a forward and right movement of the avatar. The rotations can be combined with a forward movement by turning and lifting the head at the same time.

Figure 3.4.20.: The second WoW add-on, showing the names of the actions (target, attack, and loot) and the icons for the facial expression, which were selected individually for the subject to trigger the actions. In this example, the player could target an enemy by raising his eyebrows, he could attack his target by looking to the right, and when he moved his avatar close to the corpse he could loot the items by smirking right.

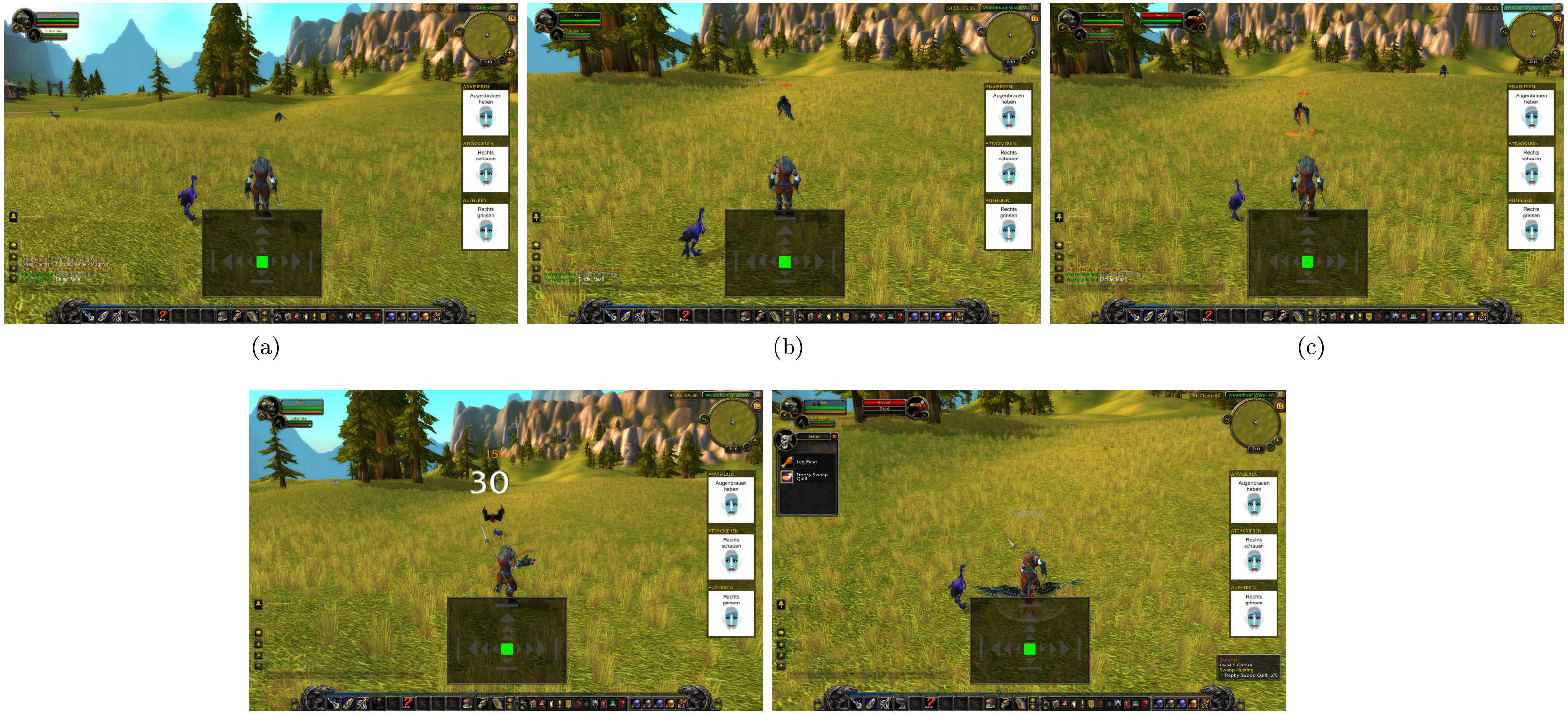
actions in the game (target, attack, and loot) three facial expressions were selected depending on the results of the Expressiv Test and the group the subject fell into. The expressions were chosen based on the results of the interference matrix resulting from the test block 2 of the Expressiv Test. For group A, three expressions were selected– those that were least interfered by other expressions. For group B, the expressions were selected with the highest interference count, but only if the subject was able to perform at least 60% of the trials successfully. Blink was selected for none of the groups as a controlling action since it is a naturally occurring involuntary expression.

Activity Score

To get a rough performance measure for the activity of a subject in the game, a scoring system was introduced. A subject received five points for killing an enemy and 10 points for looting an enemy corpse. The subjects were asked to collect as many points as possible.

To attack an enemy in WoW, the player has to run close to a creature and perform the target action (see Figure 3.4.21). Once the target was selected it could be attacked. In this user study, all the special attacks were triggered automatically by a GBGC input macro. Once the enemy was defeated, the user could collect loot from the corpse of the killed creature. To do this, the player had to move close to the dead body. Since the looting procedure is more tedious and requires more accurate movement than the killing they were awarded double the points.

The subjects had five minutes to get used to the game, the controls, and the re-centering procedure using the GBGC bordered mode. During this practice time no points were awarded. After the first five minutes each participant had 10 minutes to kill and loot as many enemies as possible. During this time each successful action was awarded with the corresponding points.



(a) (b) (c) (d) (e)

Figure 3.4.21.: WoW action score system. To score points, the subjects had to kill and loot creatures within the game WoW (a). To perform an attack, the player has to approach an enemy (b). Next, he has to target (select) the creature by performing the corresponding facial expression (c). Then the player can give the command to attack the creature (d). When the target is dead, he can move close to the corpse and collect the loot from the corpse of the recently killed opponent (e).

4.2. User Study

This section describes the detailed results of the conducted user study. The VAS questionnaire results for the wearing comfort of the headset are assessed first, followed by the measurement and questionnaire results raised by the Cognitiv Test, Affectiv Test, and Expressiv Test. The results of the Experience Test conclude this chapter.

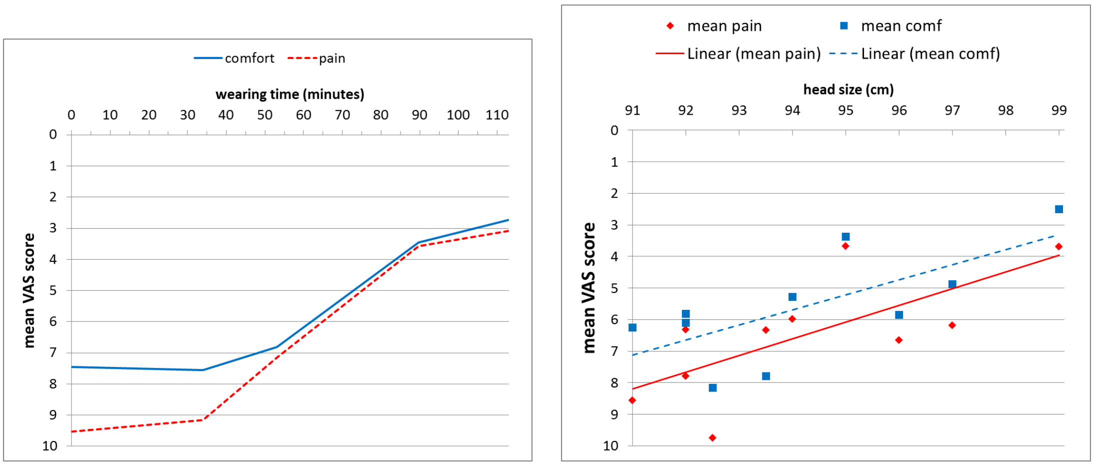
4.2.1. Wearing Comfort

The mean wearing times assessed by the questionnaires are 0, 34, 53, 90, and 113 minutes. Each part lasted about the same time for each subject due to the design of the user study. The clock was started immediately after the headset was mounted and all electrodes were adjusted to good contact. Figure 4.2.1(a) and Table 4.2.1 show the results of the questionnaires.

The relation between the head size and the mean comfort/pain VAS values is depicted in Figure 4.2.1(b). The head size was calculated by adding the nasion-inion distance and the head circumference of a subject. The means for comfort and pain were calculated using all scores marked by a subject during the whole study.

Wearing Time:	**0 min**	**34 min**	**53 min**	**90 min**	**113 min**
Mean Comf.:	7.45 ($\pm$2.00)	7.56 ($\pm$1.65)	6.82 ($\pm$3.38)	3.46 ($\pm$2.58)	2.73 ($\pm$2.31)
Mean Pain:	9.53 ($\pm$0.96)	9.15 ($\pm$1.48)	7.15 ($\pm$3.53)	3.58 ($\pm$3.03)	3.08 ($\pm$2.98)

Table 4.2.1.: The mean VAS values ($\pm$STD) for the wearing comfort and pain of the Emotiv EPOC. A value of 10 indicates that the subjects did not perceive the headset as disturbing or painful. Zero denotes that wearing the headset was considered disturbing or painful.



(a) (b)

Figure 4.2.1.: The questionnaire results for the wearing comfort of the Emotiv EPOC. (a) Shows the mean values from the VAS, assessing wearing comfort and pain for all subjects over time. A value of 10 on the vertical axis indicates that the headset was not perceived as disturbing or painful at all, whereas 0 denotes that wearing the headset was very disturbing or painful. (b) Shows the relation between the mean comfort/pain VAS values and the head size of a subject. The head size was calculated by adding the nasion-inion distance and the head circumference. The vertical axis depicts the mean VAS values for comfort and pain which are calculated over all scores of the study for each single subject. The linear tendency lines show that subjects with a bigger head size generally perceived the headset more disturbing and painful than subjects with a smaller head size.

4.2.2. Cognitiv Test

This cue-paced experiment measured how long the subject lifted the cube during the trial time. If the cube was lifted during a neutral trial, the lift time was considered as a false positive (FP). Conversely, if the cube was lifted during a lift trial, the lift time was considered as a true positive (TP).

A comparison between the mean lift time during the neutral trials (MLTN) and the mean lift time during lift trials (MLTL) reveals if a subject was able to influence the position of the cube according to the given cue. The comparison between the MLTN (mean FP) and MLTN (mean TP) is shown in Figure 4.2.3 and Figure 4.2.4.

Accuracy

Accuracy is generally calculated as:

$$ACC = \frac{TP + TN}{P + N}$$

Where TP is the number of true positive and TN is the number of true negative (TN). P is the number of all positive instances of an experiment and N the number of all negative instances. To calculate the accuracy for each block of the Cognitiv Test, the formula for accuracy was adapted to:

$$ACC = \frac{MLTL + (TT - 1500 - MLTN)}{2(TT - 1500)}$$

Where MLTL is considered as TP and TT is the trial time (8000 ms). MLTN is considered as FP, therefore, TN can be calculated by (TT - MLTN).
1500 ms were subtracted from the overall trial time since the subjects were asked to start with their imagination as soon as the trial time started. The delay, which is needed for activation, is therefore not part of the accuracy calculation. 1.5 seconds were chosen based on the resulting graph from a time-based evaluation of the mean lift time energy (see Figure 4.2.5(d)). Additionally, [88](Chapter 10.4.3 BCI as Game Controller) confirms that all BCI types (SSVEP, ERD/ERS and P300) have a latency higher than 1.5 seconds. The resulting accuracies for each Cognitiv Test block are shown in Figure 4.2.5(a).
Another representation for investigating the lift and neutral behavior over time is shown in Figure 4.2.5(b–d). The mean lift energy was calculated once for all neutral trials and once for all lift trials (block 1, block 2 and all trials). The results of this evaluation for each individual subject are shown in Appendix E.

To assess if the subjects improved in accuracy using more training trials in advance to the test trials, the accuracy improvements were calculated using the following formula:

$$ACC\_IMPRV = ACC(block\,2) - ACC(block\,1)$$

A positive ACC_IMPRV value states that the subject's accuracy improved from block 1 to block 2. A negative ACC_IMPRV value indicates that the accuracy was lower in block 2 than in block 1. The results for the accuracy improvements are shown in Figure 4.2.5(e).

Receiver Operating Characteristic (ROC) Curves

The Receiver Operating Characteristic (ROC) curve plots the relation between the true positive rate (TPR) and the false positive rate (FPR) of a binary classified system. Whereas the TPR and FPR is calculated by:

$$TPR = TP/P$$

$$FPR = FP/N$$

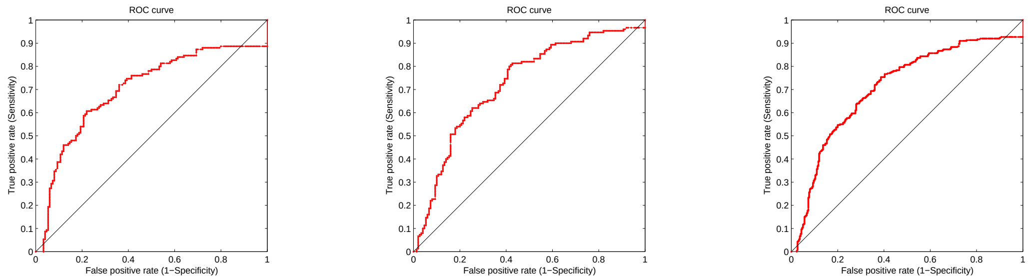
A threshold value is used to divide the test trials of the Cognitiv Test into two classes (neutral and lift). All trials that have a lift time below the threshold level are considered as neutral, and all trials with a greater lift time are considered as lift trials. By comparing the estimated results to the trial type a number of TP and TN can be calculated. In both cases the total number of neutral and lift instances is 15 (P = N = 15). By shifting the threshold from 0 to TT (8000 ms) a set of TPR/FPR pairs can be calculated. Each pair represents a point in the 2D ROC space. The FPR is drawn on the horizontal (x) axis, and the TPR is shown on the vertical (y) axis. The values on the x and y axes range from 0.0 to 1.0.

A random system would be represented in a ROC curve as a diagonal line from the origin [0, 0] to the upper right corner [1, 1]. A ROC curve that runs above this random line (close to the upper left corner) states that the system performs better than random. A perfectly accurate system would be shown by two straight lines running from [0, 0] to [0, 1] to [1, 1]. The calculated ROC curves for all trials in block 1, block 2, and both are shown in Figure 4.2.2. The ROC curves for each subject are shown in Appendix D.

A performance indicator, which can be derived from a ROC curve, is the area under the curve (AUC). A random system has an AUC of 0.5, whereas a perfect system has an AUC of 1.0. The AUC for block 1 is 0.7 and for block 2 the AUC is 0.724, which results in a difference of +0.024 (+2.4%).

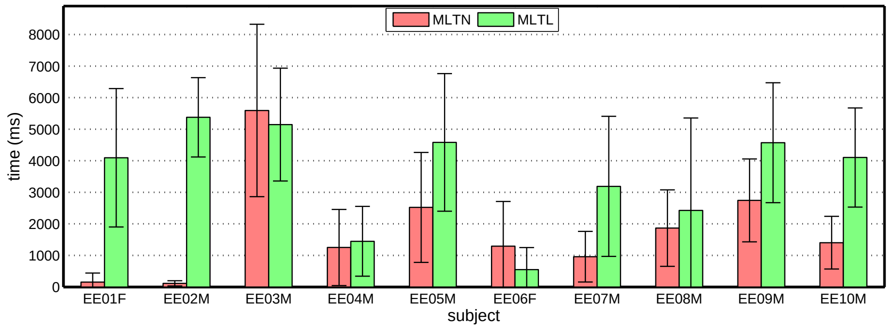
Perceived Cube Control

A VAS was included in the questionnaire after the Cognitiv Test, assessing the subjects' perceived control over the virtual cube. The results of the VASs are shown in Figure 4.2.6.

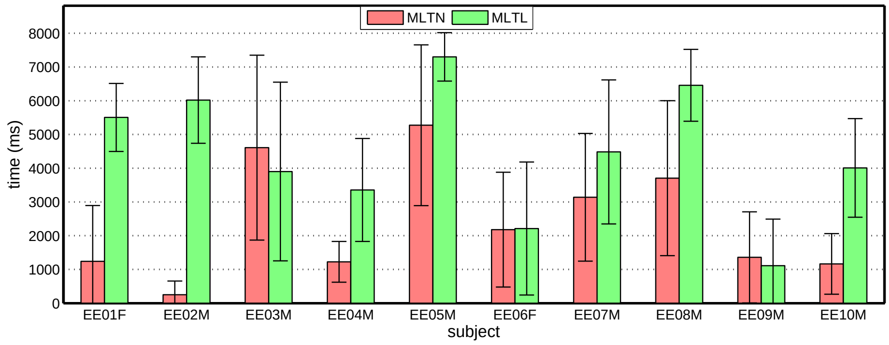


(a) (b) (c)

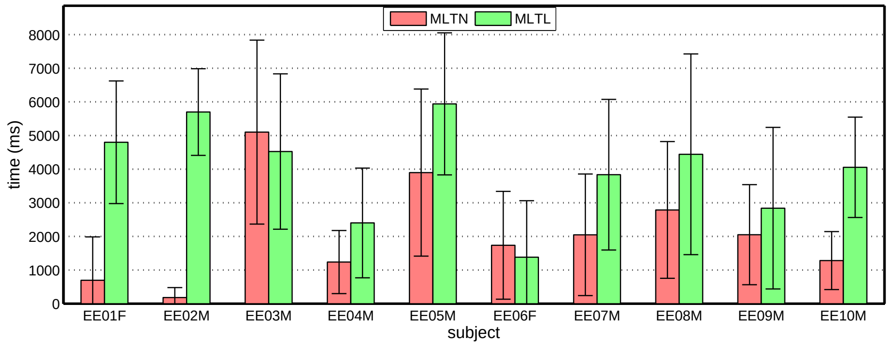
Figure 4.2.2.: The ROC curves for the Cognitiv Test. (a) Shows the ROC curve for all test trials in block 1 with an AUC of 0.7. (b) Uses all samples from all subjects of block 2 and has a AUC of 0.72. (c) Shows the ROC curves resulting from all (block 1 + block 2) test trials of the Cognitiv Test with an AUC of 0.71.



(a)



(b)



(c)

Figure 4.2.3.: The mean lift times (±STD) for each individual subject. The red bars show the MLTN (mean FP) and the green bars show the MLTL (mean TP). If the MLTL is significantly higher than the MLTN, it can be assumed that a subject had control over the cube. (a) Shows the MLTN and MLTL for the Cognitiv Test block 1 where five trials per state have been used in advance for training. (b) Shows the same values for Cognitiv Test block 2 where six trainings per state were performed prior to the test trials. (c) Shows the MLTN and MLTL for all test trials performed by a subject.

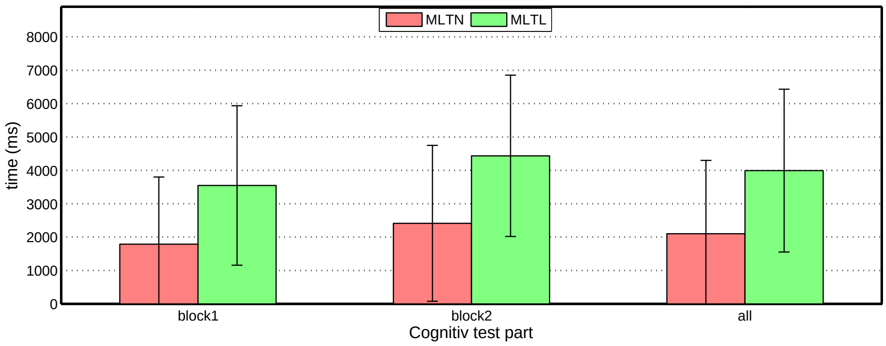
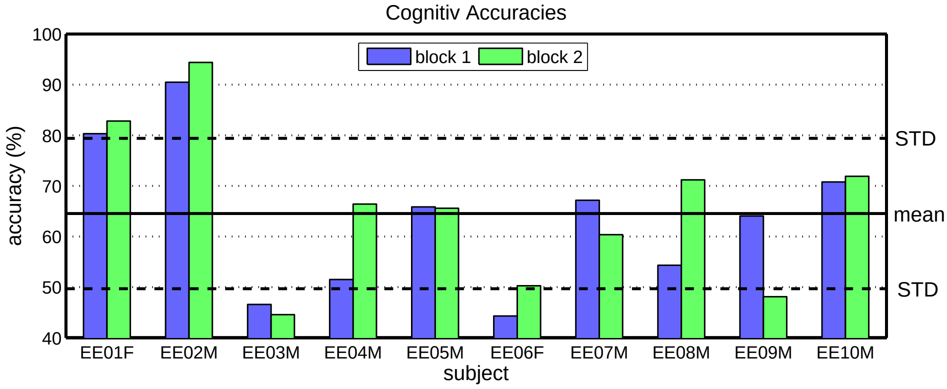
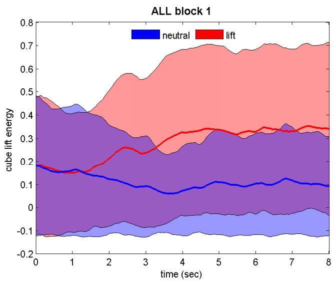


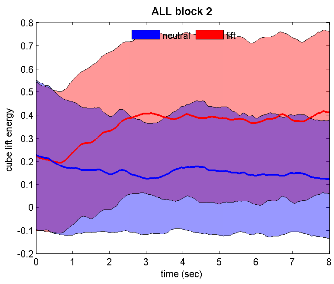
Figure 4.2.4.: The mean lift times (±STD) for all subjects combined. The red bars show the MLTN (mean FP) and the green bars show the MLTL (mean TP). The left group shows MLTN/MLTL for block 1, the middle bars represent the MLTN/MLTL for block 2, and the right bars show these values for both blocks combined.



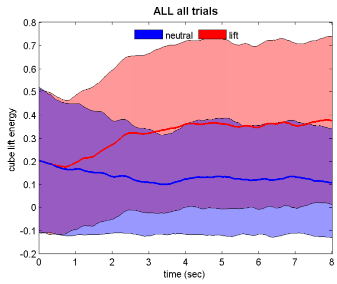
(a)



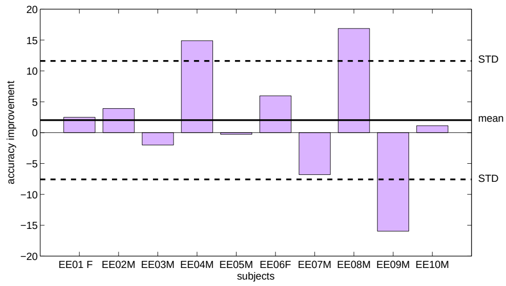
(b)



(c)



(d)



(e)

Figure 4.2.5.: The accuracies and their improvements for the Cognitiv Test. (a) Shows the accuracies for block 1 (blue) and block 2 (green) for each single subject. The mean accuracy, calculated over all subjects, is 64.53 (±14.86 STD). (b–d) Shows the mean lift energy (±STD) for all subjects over time. The means for the blue curve were calculated using all neutral trials and the red curve was calculated from all lift trials. The colored area represents the STD over time (blue = neutral, red = lift). (e) Represents the accuracy improvements for each subject between block 1 and block 2. A positive value indicates that the accuracy improved from block 1 to block 2. A negative value states that the accuracy is lower in block 2 than in block 1. The mean improvement for all subjects is +2.02% (±9.6 STD).

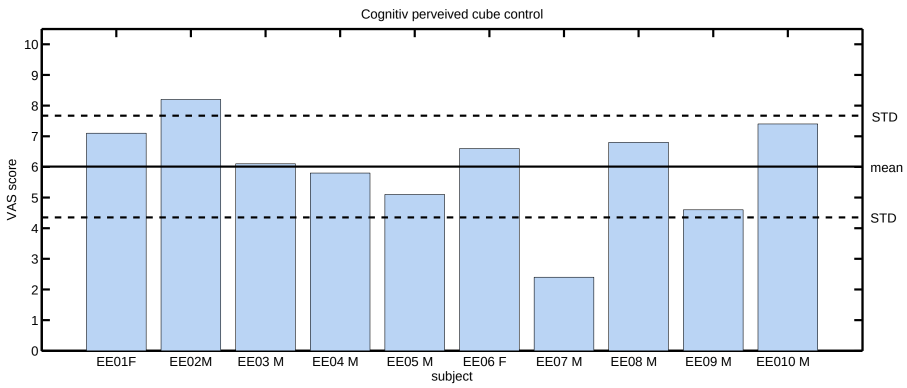


Figure 4.2.6.: The results from the VAS raising the perceived virtual cube control of each subject. A value of 10 on the vertical axis indicates that the subject felt like it had total control over the cube and a value of zero states that the subject felt like it had no control. The mean value of all VAS scores is 6,01 ($\pm$1,66 STD). The three subjects who marked the highest control VAS values also had the best accuracy (compare with Figure 4.2.5(a)).

4.2.3. Affectiv Test

The detailed results of the questionnaires are shown in Figure 4.2.7 and Table 4.2.2.

The values recorded by the Emotiv Affectiv Suite were re-sampled with 100 Hz to calculate the mean signal values (±STD) over all subjects. The signals for each emotion are shown in the following figures:

1. Frustration and instantaneous excitement, see Figure 4.2.8.
2. Engagement/boredom and meditation, see Figure 4.2.9.

The differences for the mean values during each phase of the trials are presented in Table 4.2.3.

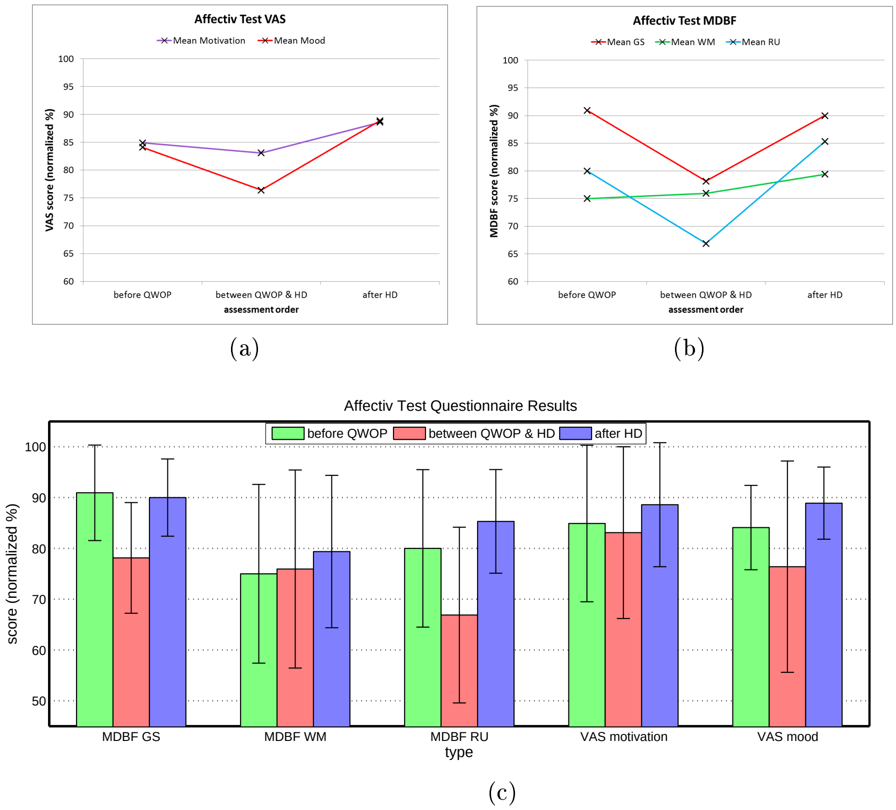
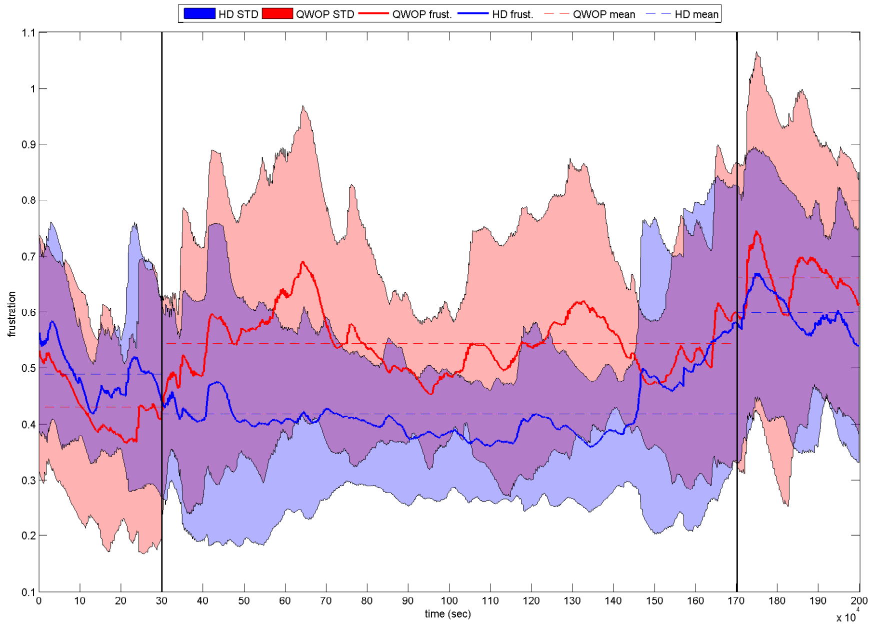
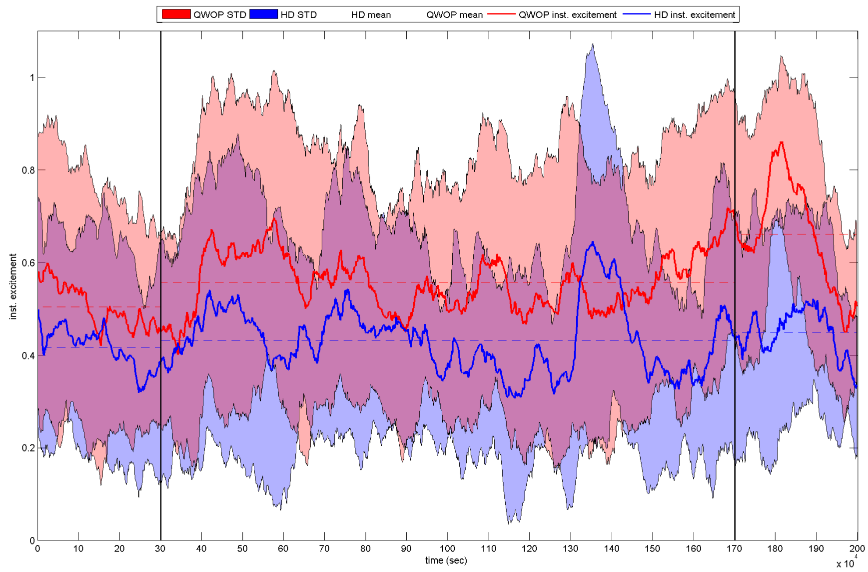


Figure 4.2.7.: The questionnaire results of the Affectiv Test. To get comparable results the values were normalized for the VAS and MDBF. (a) Shows the mean VAS scores of all subjects before playing QWOP, between playing QWOP & HD, and after playing HD. A higher score on the vertical axis states that the subjects had a better mood or more motivation. A lower value indicates that the subjects got more frustrated or less motivated. (b) Shows the mean MDBF dimensions GS, WM, and RU before playing QWOP, between playing QWOP & HD and after playing HD. A detailed description of these values can be found in Chapter 3.4.3. (c) Shows the mean values (±STD) for each assessed value of the questionnaires in the Affective Test.

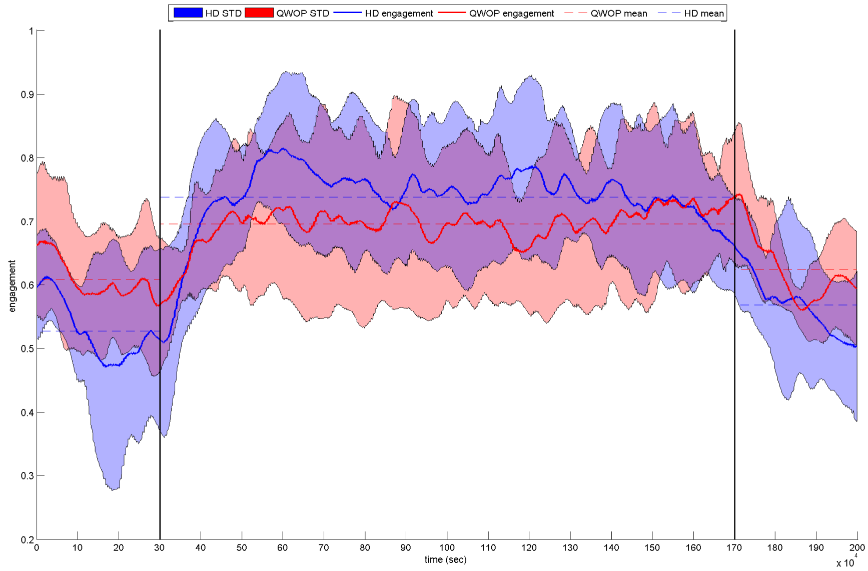


(a) Affectiv Suite frustration

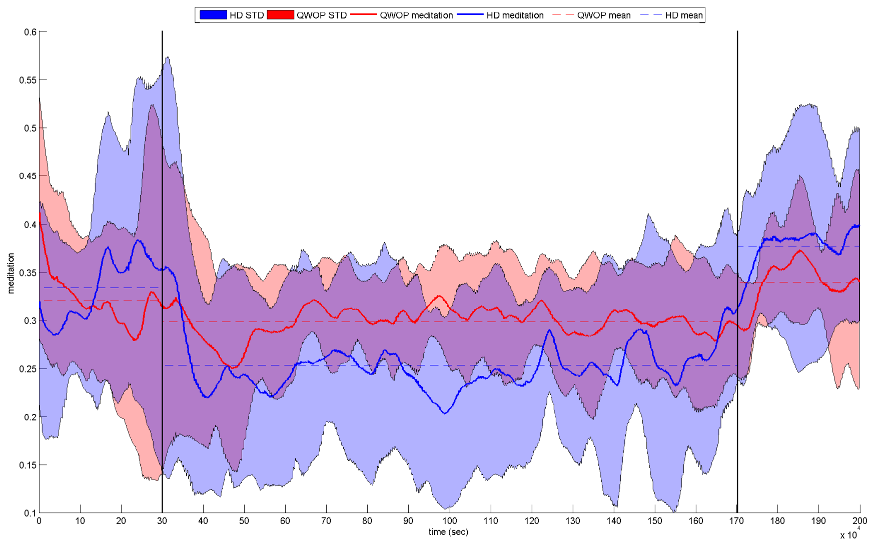


(b) Affective Suite instantaneous excitement

Figure 4.2.8.: Affectiv Suite measurements during the game tasks of QWOP (red) and High Delivery (blue). The average values per trial phase are marked by the dotted lines. (a) Shows the mean frustration of all subjects. (b) Shows the mean instantaneous excitement of all subjects.



(a) Affective Suite engagement



(b) Affective Suite meditation

Figure 4.2.9.: Affectiv Suite measurements during the game tasks of QWOP (red) and High Delivery (blue). The average values per trial phase are marked by the dotted lines. (a) Shows the mean engagement of all subjects. (b) Shows the mean meditation of all subjects.

Type	Score	Before QWOP	Between QWOP & HD	After HD
VAS	Mood	84.9 (±8.3)	83.1 (±20.8)	88.6 (±7.1)
VAS	Motivation	84.1 (±15.4)	76.4 (±16.9)	88.9 (±12.2)
MDBF	GS	90.9 (±9,4)	78.1 (±10.9)	90.0 (±7.6)
MDBF	WM	75.0 (±17.6)	75.9 (±19.5)	79.3 (±15.0)
MDBF	RU	80.0 (±15.5)	66.9 (±17.3)	85.3 (±10.2)

Table 4.2.2.: The normalized questionnaire results of the Affectiv Test in % (±STD).

	Relax phase 1	Play time	Relax phase 2
Frustration	-6%	+12%	+6%
Engagement	+8%	-4%	5%
Inst. excitement	+9%	+13%	21%
Meditation	-1%	+5%	-4%

Table 4.2.3.: Affectiv Suite measurement differences for the gaming tasks. The differences for the different phases of the trials are calculated by subtracting the mean value of the QWOP trial from the mean value of the HD trial. A positive value indicates that the mean value in the certain phase is higher in QWOP and vice versa.

4.2.4. Expressiv Test

The mean RTs for each expression are shown in the order of their occurrence in Figure 4.2.10. Table 4.2.4 shows the mean RTs (±STD) for all successful trials of the training, block 1 and block 2.

Accuracy

A trial is either considered successful or not depending on whether the subject was able to perform the right signal during the five-second trial time. All other classifications, apart from the one being requested by the cue, are ignored. As a result, there are only TP and FN samples. The accuracies for each facial expression are therefore determined by the TPR. The accuracies for each expression, calculated from a total of 150 samples (training + block1 + block2) per expression, are shown in Figure 4.2.11. The mean accuracy for all facial expressions is 89.6% (±11.5 STD).

	Training	Block 1	Block 2
Mean	1694 ms	1501 ms	1479 ms
STD	886 ms	825 ms	801 ms

Table 4.2.4.: The mean RTs and STD calculated out of all successful trials per block.

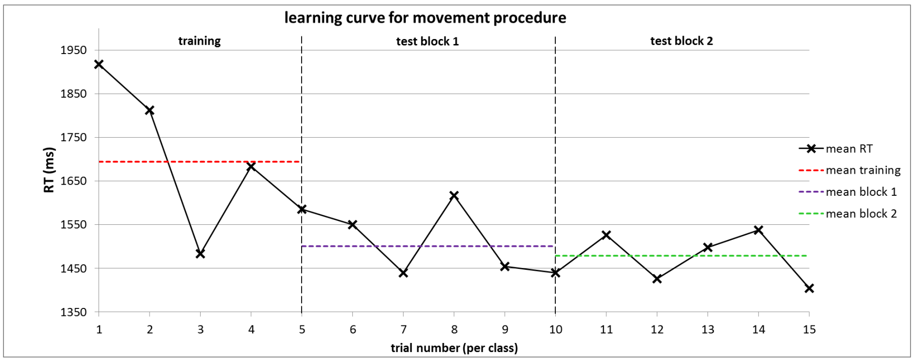


Figure 4.2.10.: The mean RTs in relation to the repetition count. Each expression was performed five times per block (15 in total). The means were calculated over all successfully performed expressions with the same repetition count. The graph shows that the subjects get faster as the repetition count rises. Also included are the mean RTs of all successful trials in the training, block 1, and block 2.

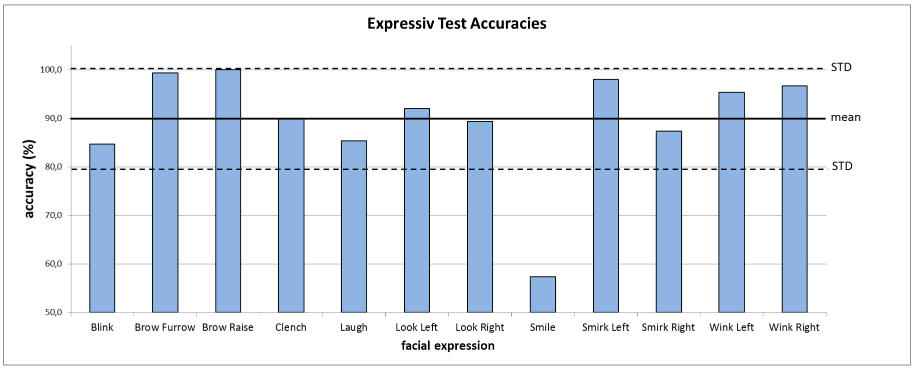


Figure 4.2.11.: The accuracies for each facial expression. A trial is considered successful if the subject was able to give the right signal during the five-second trial time. All other signals apart from the one being requested are ignored. In cases where the subject could not recreate the facial expression or the Emotiv EPOC was not able to detect it, the trial was unsuccessful. The mean accuracy for all facial expressions is 89.6% (±11.5 STD).

Expression	mu (μ)	sigma (δ)	tau (τ)	max. like. (ms)	max. power
blink	557,48	296,77	840,34	910	0,00066
brow furrow	721,78	432,75	519,36	1059	0,00068
brow raise	1032,59	144,51	267,35	1174	0,00170
clench	1321,98	253,83	7,64	1330	0,00160
laugh	1169,66	429,42	683,54	1560	0,00061
look left	812,86	136,13	272,51	951	0,00170
look right	833,15	15,02	438,59	867	0,00210
smile	877,13	415,89	1125,34	1362	0,00049
smirk left	1080,42	453,57	558,91	1439	0,00064
smirk right	1070,03	379,61	870,16	1482	0,00058
wink left	774,00	0,00	756,87	774	0,00130
wink right	837,05	45,67	671,89	927	0,00130
all	830,18	300,88	659,40	1150	0,00075

Table 4.2.5.: The estimated parameters (μ, δ, and τ) for the ex-Gaussian distributions from the DISTRIB toolbox for each single expression and all expressions. Additionally, the point of the maximum likelihood (maximum power) and the corresponding RT (maximum likelihood) was computed.

Premotor Times

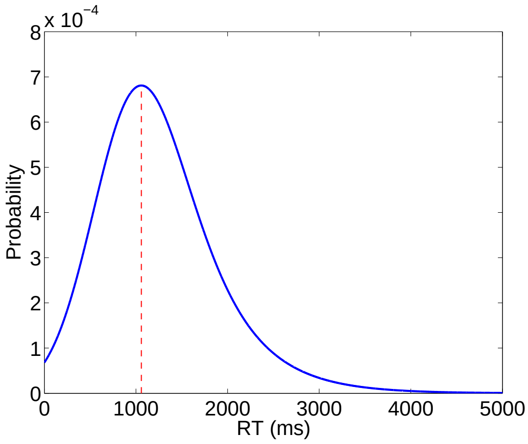
To evaluate the premotor times of each facial expression the successfully completed trials from the Expressiv Test block 1 and block 2 were used to estimate probability distributions. Because of the findings of [92], the ex-Gaussian distribution was selected for evaluating the RTs. The curves were calculated from the samples using the DISTRIB matlab toolbox for characterizing RT (V2.3) distributions provided by [46]. The resulting estimated distributions for each single expression are shown in Figure 4.2.12 and Figure 4.2.13. The estimated ex-Gaussian distribution for all successfully detected facial expressions is shown in Figure 4.2.14. The detailed parameters for each distribution are listed in Table 4.2.5.

Interference Matrix

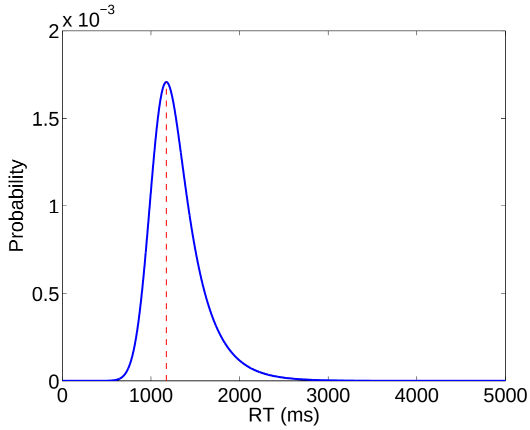
To investigate the influences between the detections of facial expressions, the interference matrices of each block and each subject were summed up (see Table 4.2.6). The structure of the interference matrix is explained in detail in Chapter 3.4.10.

Motor Times

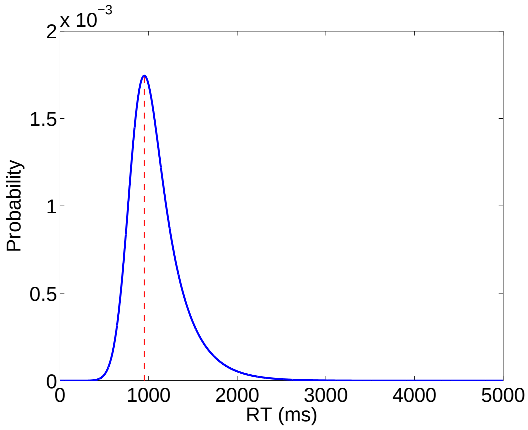
Only the completed trials from the Expressiv Test were used to calculate the mean motor times for each facial expression. The results are shown in Figure 4.2.15.



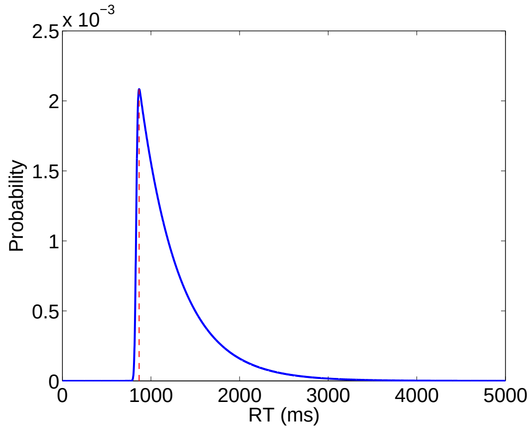
(a) Brow Furrow



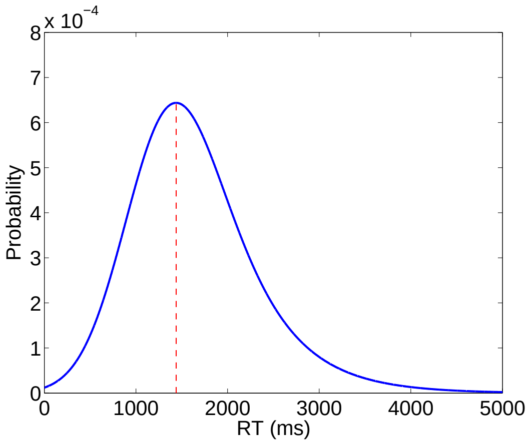
(b) Brow Raise



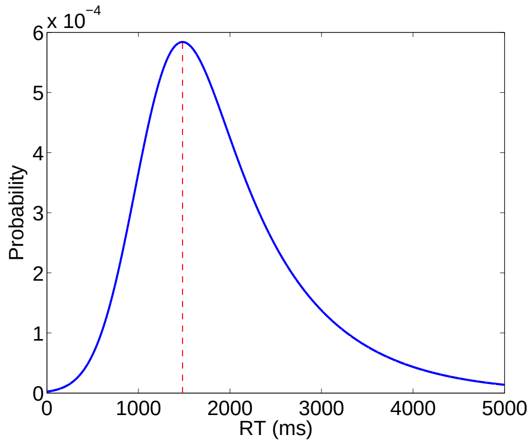
(c) Look Left



(d) Look Right

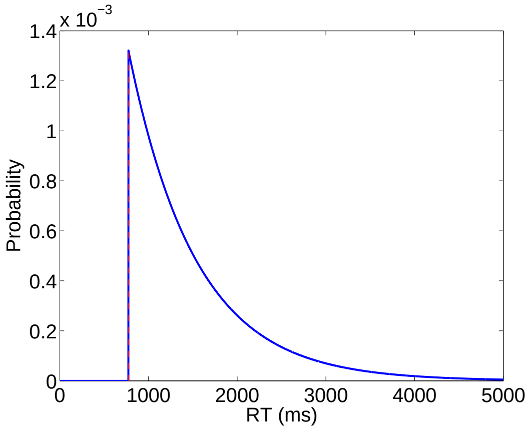


(e) Smirk Left

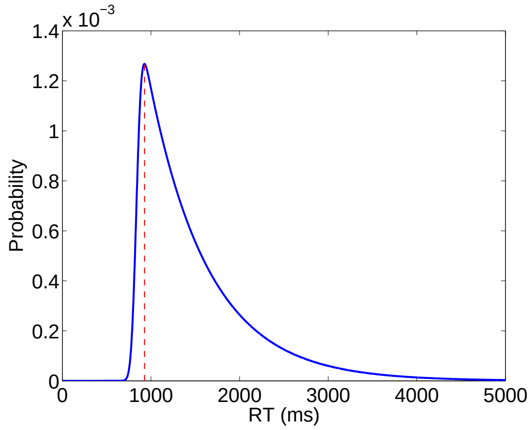


(f) Smirk Right

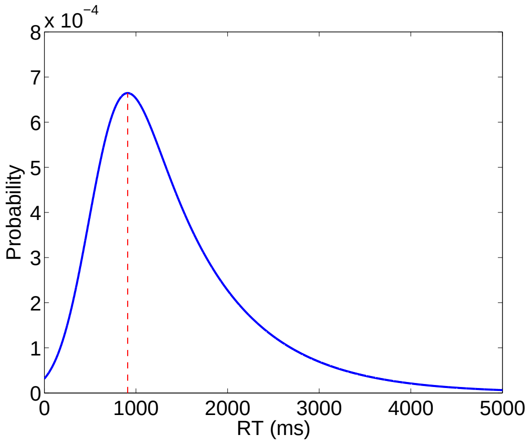
Figure 4.2.12.: The ex-Gaussian distributions (a–f) for each facial expression calculated from the successfully completed RT trials of the Expressiv Test (block 1 + block 2).



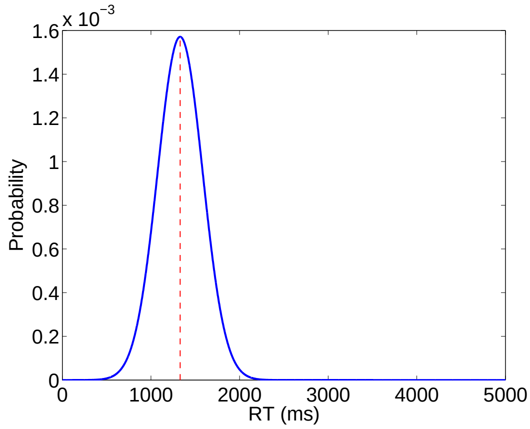
(a) Wink Left



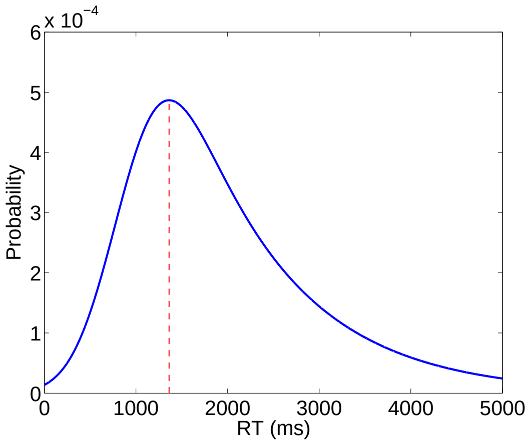
(b) Wink Right



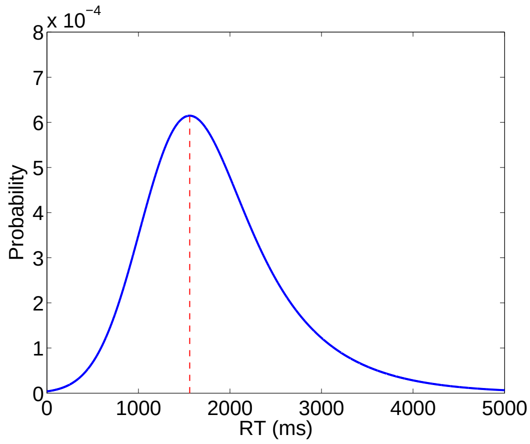
(c) Blink



(d) Clench



(e) Smile



(f) Laugh

Figure 4.2.13.: The ex-Gaussian distributions (a–f) for each facial expression calculated from the successfully completed RT trials of the Expressiv Test (block 1 + block 2).

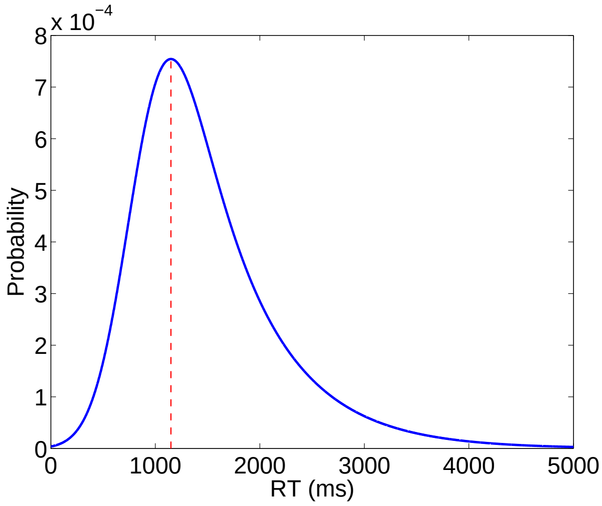


Figure 4.2.14.: The ex-Gaussian distribution calculated from all successful trials from the Expressiv Test (block 1 + block 2).

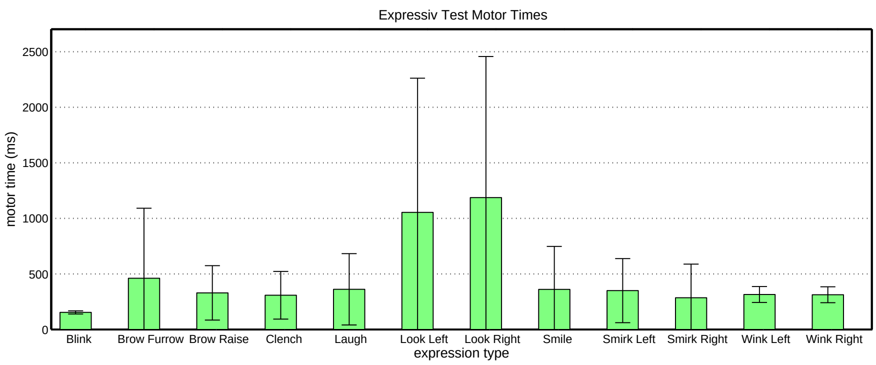


Figure 4.2.15.: The mean motor times (±STD) for each single facial expression. All successful trials from the Expressiv Test were considered in the calculation.

	blink	wink left	wink right	look left	look right	raise brow	fur. brow	smile	clench	laugh	smirk left	smirk right	∑
blink	-	46	32	2	3	2	68	2	0	5	3	3	166
wink l.	22	-	1	6	2	5	34	2	0	7	17	4	100
wink r.	33	2	-	2	0	13	33	3	0	0	10	2	98
look l.	17	2	4	-	21	6	29	8	1	0	8	4	100
look r.	30	7	20	27	-	2	48	12	0	3	4	2	155
r. brow	11	4	4	4	1	-	16	2	0	2	9	0	53
f. brow	7	5	4	2	2	9	-	1	2	4	5	2	43
smile	40	7	7	4	10	4	15	-	1	66	96	15	265
clench	15	8	6	4	5	1	19	17	-	10	10	15	110
laugh	26	10	16	6	7	33	32	34	5	-	19	17	205
smirk l.	12	8	3	12	0	0	15	12	1	8	-	1	72
smirk r.	15	2	14	4	6	0	24	9	0	13	30	-	117
∑	228	101	111	73	57	75	333	102	10	118	211	65	

Table 4.2.6.: The interference matrix for all subjects and all test trials (block 1 + block 2), showing the reliability for each facial expression and how they are influenced by other expressions. The structure of the interference matrix is explained in detail in Chapter 3.4.10.

4.2.5. Experience Test

Figure 4.2.16 shows the results of the VASs capturing the user experience and the compared performance of group A and B.

The mean score (kills + loots) of group A was 383 (± 76 STD) and group B was able to score 303 (± 70 STD) points on average. Only two out of the ten subjects had some experience (less than 15 days play time) with the game WoW (one in group A, one in group B).

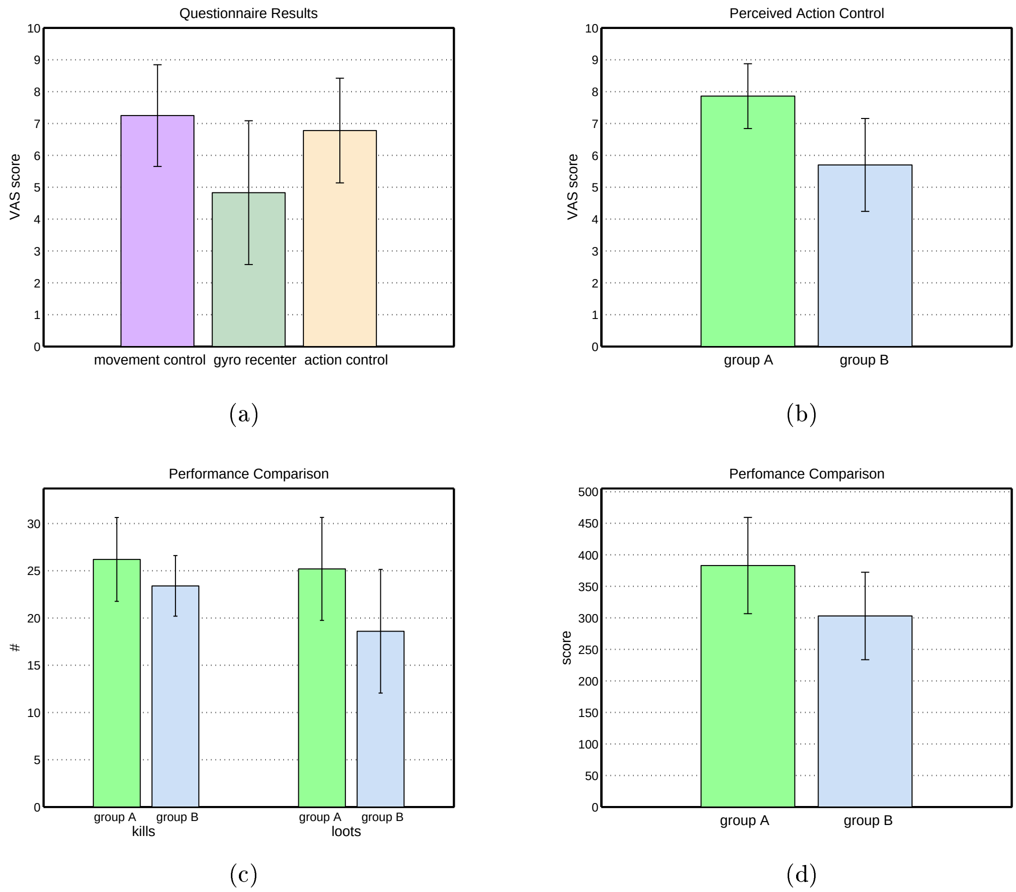
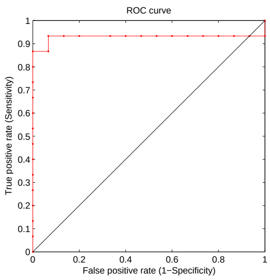
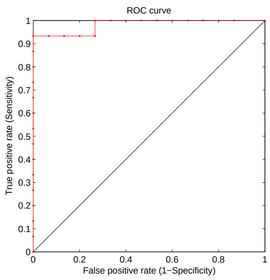


Figure 4.2.16.: (a) shows the mean scores of all subjects (group A + group B) for the three VASs given after the WoW experience test. The left bar represents how the participants perceived the movement control using gyro sensors (0 = no control, 10 = total control). The middle bar represents the mean score for the VAS which depicts how hard it is for the subject to re-center the gyro sensors using the GBGC bordered mode (0 = very hard, 10 = very easy). The right bar is the mean score for the control over the three in-game actions using facial expressions (0 = no control, 10 = total control). The Figures (b–d) compare various scores of the two control groups. Group A was given the three most reliable expressions to trigger in-game actions (target, attack, loot) and group B was given the least reliable, but performable, expressions. (b) Compares the mean VAS scores raising the perceived action control of group A and B. (c) Shows the kill and loot counts for group A and B. (d) Shows the comparison between the control groups using the introduced activity score (kill = 5 points and loot = 10 points).

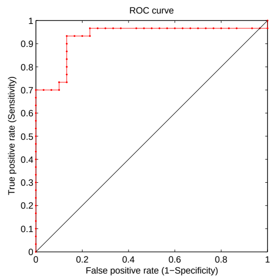
D. Cognitiv Test Results ROC Curves



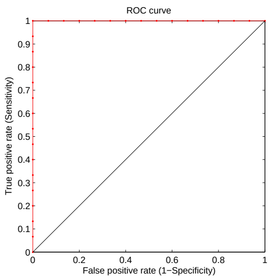
(a) EE01F block1



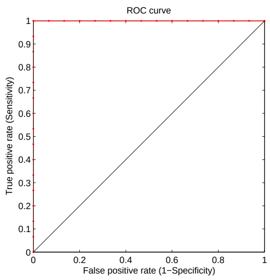
(b) EE01F block2



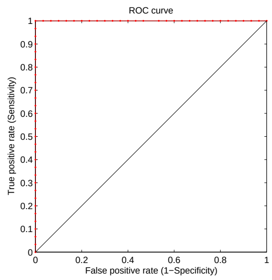
(c) EE01F all



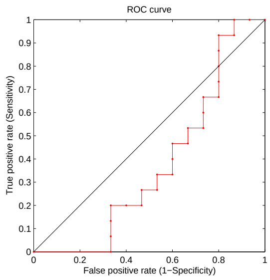
(d) EE02M block1



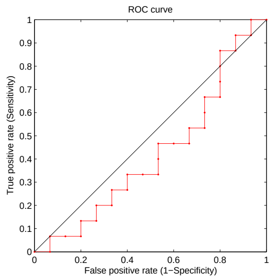
(e) EE02M block2



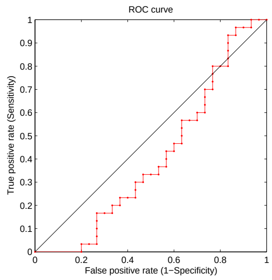
(f) EE02M all



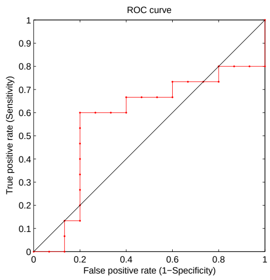
(g) EE03M block1



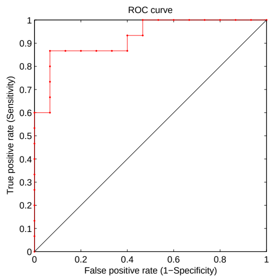
(h) EE03M block2



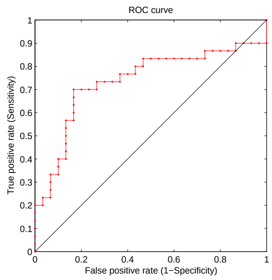
(i) EE03M all



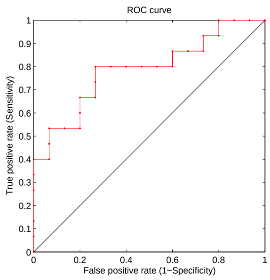
(j) EE04M block1



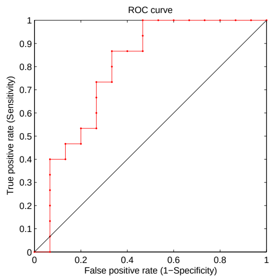
(k) EE04M block2



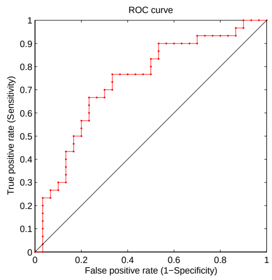
(l) EE04M all



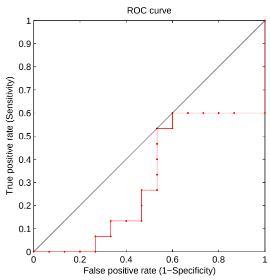
(m) EE05M block1



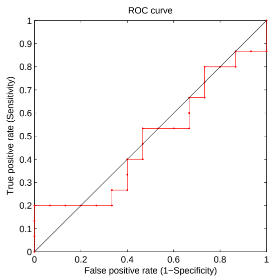
(n) EE05M block2



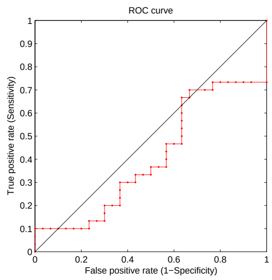
(o) EE05M all



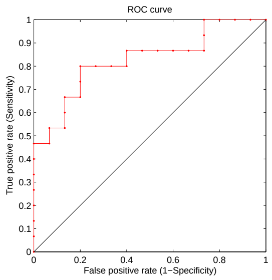
(p) EE06F block1



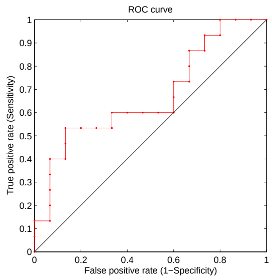
(q) EE06F block2



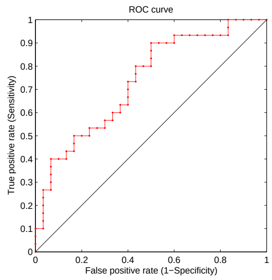
(r) EE06F all



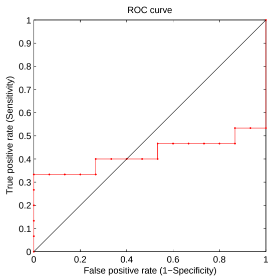
(a) EE07M block1



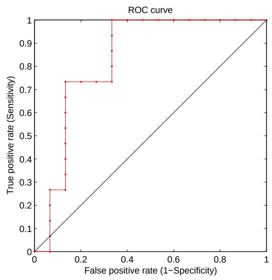
(b) EE07M block2



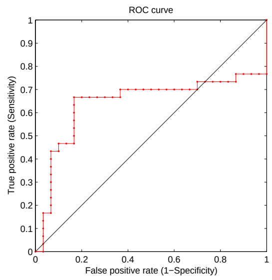
(c) EE07M all



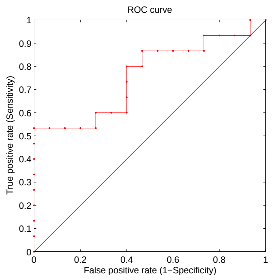
(d) EE08M block1



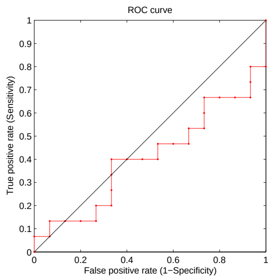
(e) EE08M block2



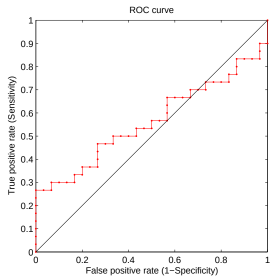
(f) EE08M all



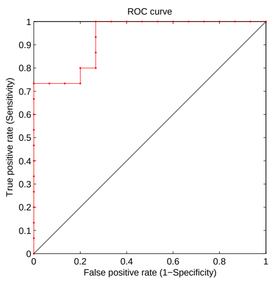
(g) EE09M block1



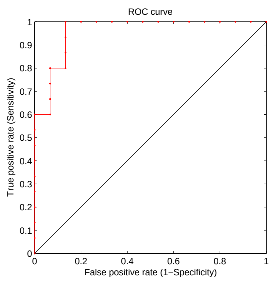
(h) EE09M block2



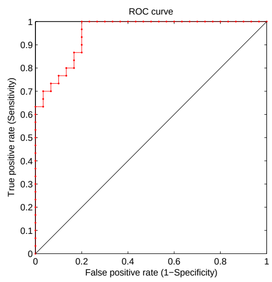
(i) EE09M all



(j) EE10M block1

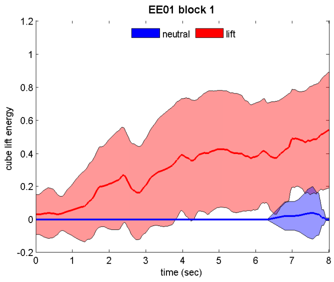


(k) EE10M block2

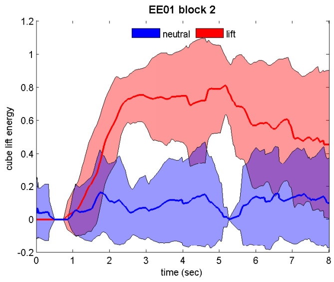


(l) EE10M all

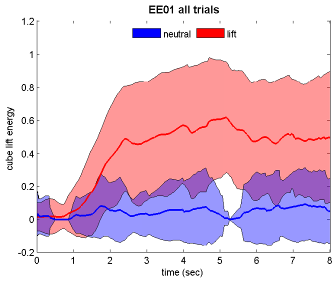
E. Cognitiv Test Results Lift over Time



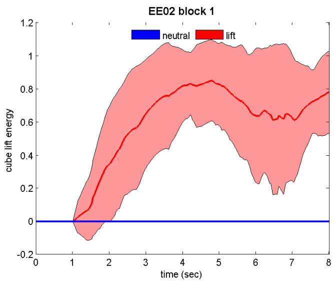
(a) EE01F block 1



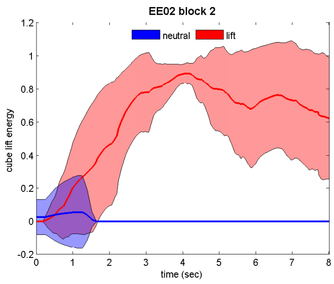
(b) EE01F block 2



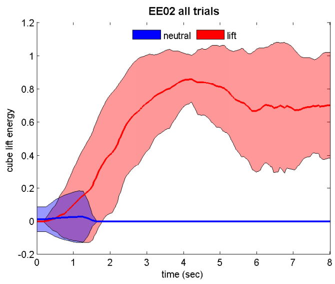
(c) EE01F all



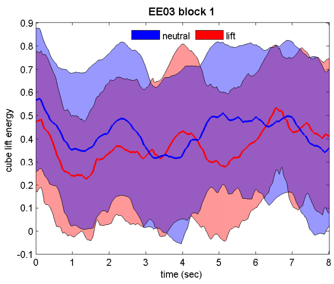
(d) EE02M block 1



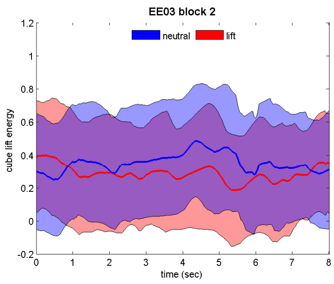
(e) EE02M block 2



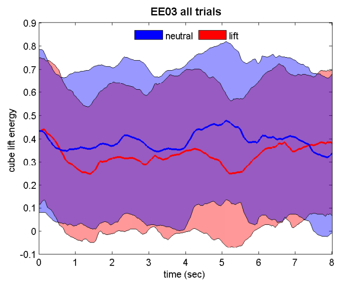
(f) EE02M all



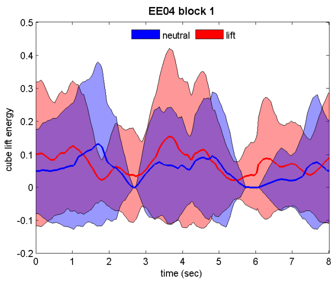
(g) EE03M block 1



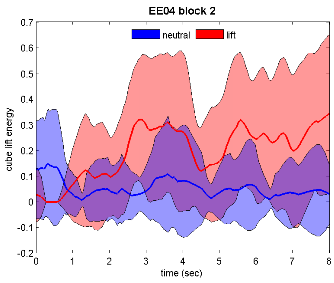
(h) EE03M block 2



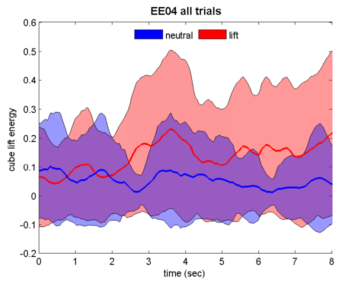
(i) EE03M all



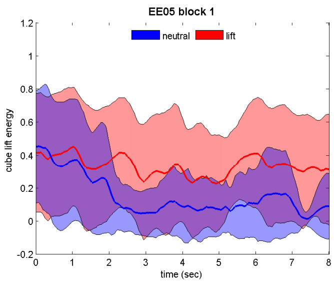
(j) EE04M block 1



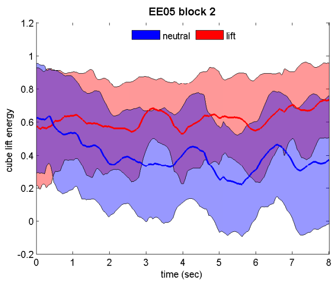
(k) EE04M block 2



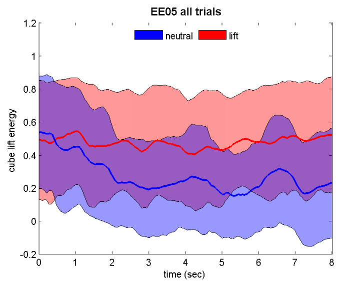
(l) EE04M all



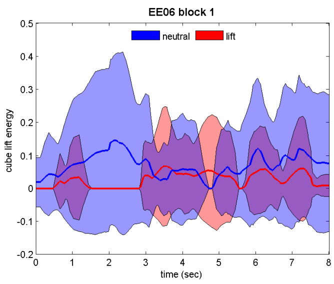
(m) EE05M block 1



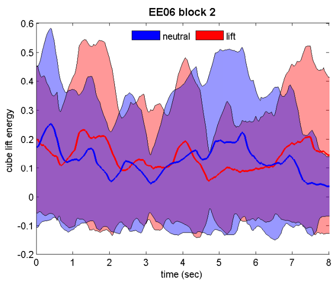
(n) EE05M block 2



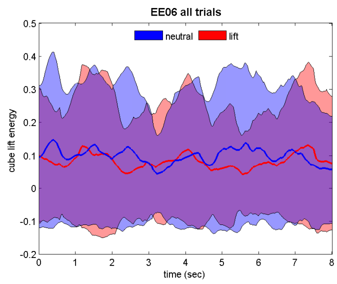
(o) EE05M all



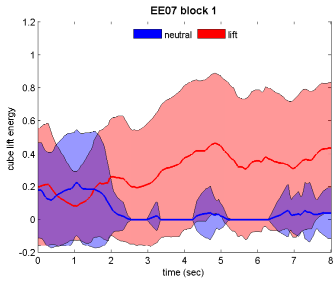
(p) EE06F block 1



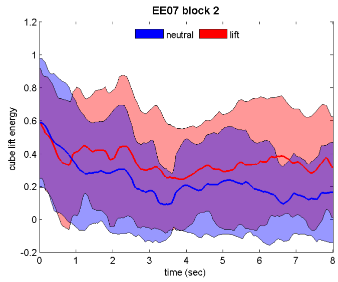
(q) EE06F block 2



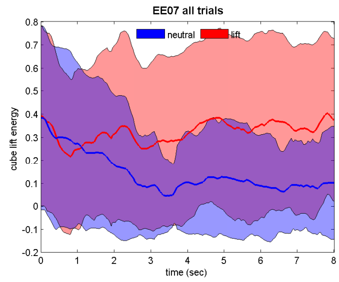
(r) EE06F all



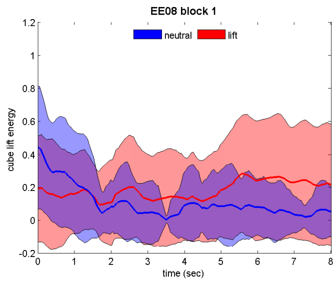
(a) EE07M block 1



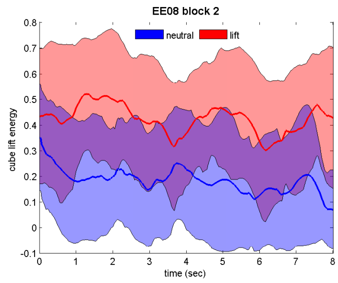
(b) EE07M block 2



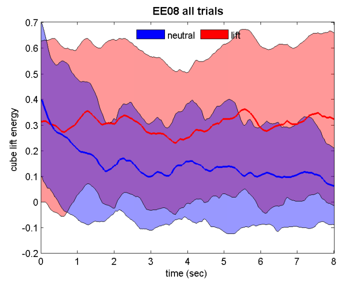
(c) EE07M all



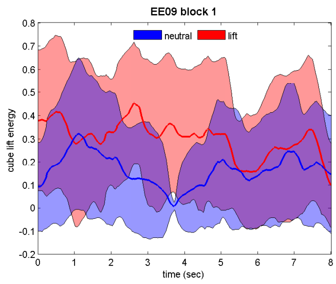
(d) EE08M block 1



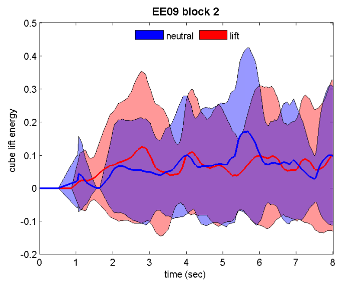
(e) EE08M block 2



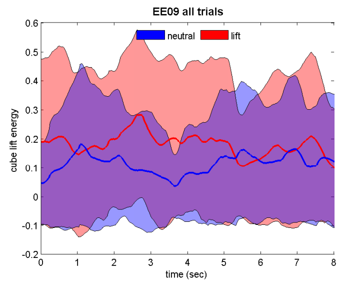
(f) EE08M all



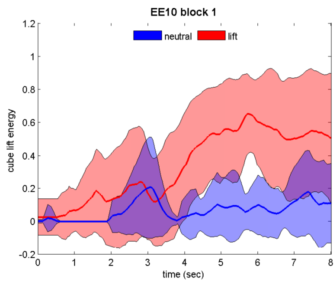
(g) EE09M block 1



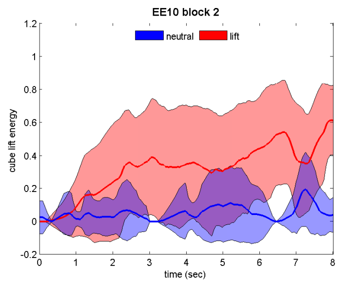
(h) EE09M block 2



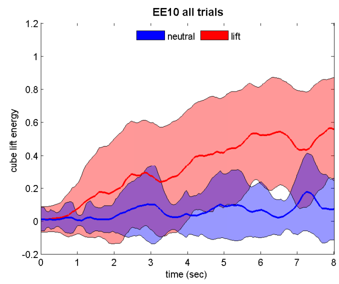
(i) EE09M all



(j) EE10M block 1



(k) EE10M block 2



(l) EE10M all