



# Modellierung des stimulierenden Effekts von emotionalem Feedback bei fMRT-Neurofeedback

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieurin**

im Rahmen des Studiums

**Biomedical Engineering**

eingereicht von

**Victoria Caic, BSc**

Matrikelnummer 01204686

an der Fakultät für Informatik  
der Technischen Universität Wien

Betreuung: Univ.Prof. Dipl.-Ing. Dr.techn. Eduard Gröller  
Mitwirkung: Dipl.-Ing. Manfred Klöbl, BSc (Medizinische Universität Wien)

Wien, 27. Oktober 2021

---

Victoria Caic

---

Eduard Gröller





# Modelling the Effect of emotional Feedback as Stimulus in fMRI Neurofeedback

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

**Diplom-Ingenieurin**

in

**Biomedical Engineering**

by

**Victoria Caic, BSc**

Registration Number 01204686

to the Faculty of Informatics  
at the TU Wien

Advisor: Univ.Prof. Dipl.-Ing. Dr.techn. Eduard Gröller  
Assistance: Dipl.-Ing. Manfred Klöbl, BSc (Medizinische Universität Wien)

Vienna, 27<sup>th</sup> October, 2021

---

Victoria Caic

---

Eduard Gröller



# Erklärung zur Verfassung der Arbeit

Victoria Gaic, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 27. Oktober 2021

---

Victoria Gaic



# Danksagung

Für die Entstehung meiner Masterarbeit möchte ich zuallererst Manfred Klöbl für die weltbeste Betreuung auch unter widrigen Umständen danken. Ich hätte mir keinen besseren Betreuer wünschen können, und danke dir für die geduldige und auch humorvolle Einführung in die Welt der Forschung. Ebenso geht mein Dank an Rupert Lanzberger und das sehr hilfsbereite Team des Neuroimaging Labs. Meine Zeit, die ich in eurer Gruppe verbringen durfte, hat nicht nur diese Arbeit, sondern auch meinen weiteren Lebensweg geprägt.

Außerdem möchte ich mich bei Meister Eduard Gröller bedanken, der mir sowohl Vertrauen entgegengebracht als auch den nötigen Freiraum gelassen hat, um diese Arbeit nach meinen eigenen Wünschen und Vorstellungen gestalten zu können.

Danke außerdem an meine Familie und Freunde, vor allem an die beste Lerngruppe der Welt. Ohne euch hätte ich es gar nicht bis hierher geschafft!

Zum Schluss möchte ich ein generelles Danke an Patrick und Joy für eure bedingungslose Unterstützung in jeder Lebenslage aussprechen: Danke, dass ihr mein Rudel seid!



# Acknowledgements

For the creation of my master's thesis, first of all, I would like to thank Manfred Klöbl for the world's best support, even under adverse circumstances. I couldn't have asked for a better supervisor, and thank you for your patient and humorous introduction to the world of research. My gratitude also goes to Rupert Lanzenberger and the very helpful team of the Neuroimaging Lab. The time I was allowed to spend in your group not only shaped this work, but also my future path in life.

I would also like to thank Meister Eduard Gröller, who had trust in me and left me the necessary freedom to be able to design this work according to my own wishes and ideas.

Thanks also to my family and friends, especially the best study group in the world. Without you I wouldn't have made it this far!

Finally, I would like to say a general thank you to Patrick and Joy for your unconditional support in every situation: Thank you for being my pack!



# Kurzfassung

Neurofeedback (NF) auf der Basis von funktioneller Magnetresonanztomographie (fMRT) bietet vielversprechende Möglichkeiten für Therapieansätze bei neurologischen und psychiatrischen Erkrankungen. Mithilfe von Feedback über die aktuelle Aktivität in spezifischen Hirnregionen kann eine bewusste Kontrolle über diese erlernt werden, die in weiterer Folge der krankheitsspezifischen Symptomatik entgegenwirken kann. Als sehr intuitive Art des Feedback wird häufig soziales Feedback in Form eines Gesichts mit sich ändernder Mimik gewählt. Da im Zusammenhang mit psychiatrischen Erkrankungen in der Regel Hirnregionen betroffen sind, die an der Wahrnehmung und Verarbeitung von Emotionen beteiligt sind, besteht die Möglichkeit, dass diese Regionen durch emotionales Feedback zusätzlich aktiviert werden. In dieser Arbeit wird untersucht, ob eine solche zusätzliche Aktivität einen signifikanten Einfluss auf die gemessene Aktivität besitzt, da diese zu einem ungenauen Feedback und in Folge zu suboptimalen Lernergebnissen führen könnte. Dazu werden die Daten einer zuvor veröffentlichten Studie erneut analysiert, wobei in erster Linie der mögliche Einfluss des Feedbacksignals im Fokus steht. Mithilfe unterschiedlicher Modellansätze wird die genaue Natur dieses Einflusses untersucht, ebenso ob positives und negatives Feedback sich in ihrem Einfluss unterscheiden. Unter Beachtung der stark individuellen Natur von NF und dem Ziel, notwendige Korrekturen für das Training eines einzelnen Probanden in offen zugängliche NF-Software zu integrieren, wurden die Analysen sowohl auf einer individuellen als auch auf Gruppenebene durchgeführt, um die Generalisierbarkeit zu testen.

Auf der Ebene der Einzeldurchgänge konnte ein signifikanter Einfluss sowohl des Feedbacks als auch seiner zeitlichen Änderung festgestellt werden. Positives Feedback besaß häufiger einen signifikanten Einfluss auf die neuronale Aktivität als negatives Feedback. Bezüglich der zeitlichen Änderung konnten öfter signifikante Ergebnisse bei negativem Feedback gefunden werden. Auf der Gruppenebene zeigte nur die zeitliche Änderung des Feedbacks einen signifikanten Einfluss auf die Aktivität der Zielregion. In einer Kreuzvalidierung konnte für keines der verwendeten Modelle eine Generalisierungsfähigkeit über einen einzelnen Messdurchgang hinaus festgestellt werden.

Der untersuchte Effekt scheint sehr individuell für einzelne Probanden und einzelne Messungen zu sein und sollte daher von Fall zu Fall unterschiedlich behandelt werden. In NF-Studien, in denen emotionales Feedback genutzt wird, um eine Hirnregion zu trainieren, die an Emotionsverarbeitung beteiligt ist, könnte die Berücksichtigung des Feedbackinflusses die Genauigkeit des präsentierten Feedbacks verbessern und damit ebenso den Lern- und Therapieerfolg.



# Abstract

Neurofeedback (NF) based on functional magnetic resonance imaging (fMRI) offers promising possibilities for therapeutic approaches in neurological and psychiatric diseases. By providing information over the current activity in a target brain region, conscious control can be learned allowing for counteracting disease-specific symptoms. Social feedback in the form of a face with changing expressions is often chosen as a very intuitive type of feedback. Since the brain regions affected in psychiatric conditions are often involved in the perception and processing of emotions, it is possible that these regions are additionally activated with emotional feedback. In this thesis it is examined whether such an additional activity has a significant influence on the measured activity, as this could lead to inaccurate feedback and, as a result, to suboptimal learning outcomes.

For this purpose, the data of a previously published study is reanalysed while particularly taking the potential influence of the feedback signal into account. Using different model approaches, the exact nature of the influence is investigated, as well as whether positive and negative feedback differ in their influence. Given the highly individual aspects of NF and the goal to implement corrections for the training of a single subject in an openly available NF software, the analyses were conducted on an individual but also the group level allowing for tests of generalizability.

At the single run level, a significant influence of both the feedback and its change over time was found. Positive feedback more often had a significant impact on the neuronal activation than negative feedback. With regard to the change over time, significant results could more often be found with negative feedback. At the group level, only the change in feedback showed a significant influence on the activation of the target region. In a cross-validation, it was not possible to determine generalizability beyond a single run for any of the models under investigation.

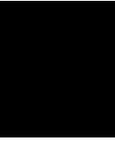
The examined effect seems to be very individual both for subjects and measurements and should therefore be treated separately from case to case. In NF studies in which emotional feedback is used while training a brain region involved in emotion processing, accounting for the influence of the feedback signal could improve the accuracy of the presented feedback and, hence, learning performance and therapeutic success.



# Contents

<b>Kurzfassung</b>	<b>xi</b>
<b>Abstract</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and problem definition . . . . .	1
1.2 Aim of the work . . . . .	2
1.3 Structure of the work . . . . .	2
<b>2 Neurofeedback in clinical neuroscience</b>	<b>5</b>
2.1 Introduction to neurofeedback . . . . .	6
2.2 fMRI Neurofeedback . . . . .	14
<b>3 Introduction to emotion processing</b>	<b>25</b>
3.1 Methods and limitations of localizing the processing of emotions . . . . .	25
3.2 Emotion processing in healthy people . . . . .	28
3.3 Emotion processing in affective disorders . . . . .	35
3.4 The subgenual anterior cingulate cortex in emotion processing . . . . .	36
3.5 Implications for the use of emotional feedback in neurofeedback studies	37
<b>4 Mathematical background</b>	<b>39</b>
4.1 General linear model in fMRI data analysis . . . . .	39
4.2 Linear mixed-effects model . . . . .	47
4.3 Cross-validation . . . . .	53
4.4 Test-retest reliability . . . . .	54
<b>5 Emotional feedback in neurofeedback studies and other related work</b>	<b>57</b>
5.1 Emotional feedback in neurofeedback studies . . . . .	57
5.2 Previous research in this project . . . . .	64
5.3 The sgACC in fMRI neurofeedback . . . . .	66
<b>6 Methods and data</b>	<b>69</b>
6.1 Data origin . . . . .	69
6.2 Linear approaches of feedback influence . . . . .	72

6.3	Design of the GLM . . . . .	77
6.4	Linear mixed-effects model design . . . . .	81
6.5	Cross-validation for single subjects . . . . .	85
6.6	Implementation in Matlab and SPM . . . . .	86
<b>7</b>	<b>Results</b>	<b>95</b>
7.1	Single run analysis . . . . .	95
7.2	Group level analysis . . . . .	100
7.3	Cross-validation . . . . .	103
7.4	Test-retest reliability . . . . .	105
<b>8</b>	<b>Discussion</b>	<b>107</b>
8.1	Single run analysis . . . . .	107
8.2	Group level analysis and cross-validation . . . . .	110
8.3	Reliability of the results . . . . .	111
8.4	Limitations . . . . .	112
8.5	Outlook . . . . .	113
<b>9</b>	<b>Conclusion</b>	<b>115</b>
	<b>List of Figures</b>	<b>115</b>
	<b>Bibliography</b>	<b>119</b>



# Introduction

## 1.1 Motivation and problem definition

The emergent field of neurofeedback (NF), especially the possibility to use it on the basis of functional magnetic resonance imaging (fMRI), has revolutionized the study of the human brain [Wei12]. NF is a special form of biofeedback, which informs subjects about the ongoing neuronal activity in predefined brain areas, networks, or the whole brain. The given feedback enables subjects to develop strategies to control this activity in a desired way. These strategies are highly individual and, for instance, cover specific thoughts and imagery, as well as the recall of autobiographic memories. Even subconsciously executed processes that cannot be described in words are possible to reach regulation ability [MJM<sup>+</sup>16].

The measurement of the neuronal activity can be carried out with various methods. Electroencephalography (EEG) has been used for this purpose for around 50 years and is thus the oldest method for NF training. Although there have already been considerable successes in self-regulation through EEG-based neurofeedback [MMM16], it also has decisive limitations: The choice of brain regions that can be used as a training target is limited to cortical areas. Additionally, the spatial resolution of the measured brain waves is comparatively low. These restrictions can be overcome by using fMRI instead, which utilizes the difference in the magnetic properties of oxygenated and deoxygenated blood. It can achieve a uniformly high spatial resolution even in deep brain areas and offers the possibility to define subject-specific localized targets for NF training. However, for a long time the required data processing in fMRI was too computationally too expensive to make it a viable alternative. Due to technical improvements in the past 20 years, especially more powerful hardware, it is now possible to carry out fMRI NF sessions with virtually ‘real-time’ data processing and feedback estimation [MBZ<sup>+</sup>15]. Nevertheless, a delay of around four to six seconds remains due to the physiological processes on which fMRI relies, which is high compared to the milliseconds in EEG measurements.

The success of NF training depends on various parameters. Careful selection of the feedback signal may have a decisive influence. While in most studies a moving bar or a raising thermometer indicates the change in neuronal activity, there is evidence that implicit social reward, like the picture of a smiling or sad face, can lead to increased learning outcomes in a neurofeedback context [MAK<sup>+</sup>15].

Similar to the EEG-based method, fMRI NF is also suitable as a promising therapeutic tool for various neurological or psychiatric disorders. The possibility to volitionally control the activity within specified brain areas, which show a pathological change in their activation pattern, could lead to a self-induced alleviation of symptoms or even long lasting improvements. Successes pointing in this direction have already been demonstrated for cases of schizophrenia, post-traumatic stress disorders (PTSD), Parkinson's disease, or major depressive disorder (MDD) [DA13].

A significant number of brain areas, which play a prominent role in neuro-psychiatric diseases, especially in affective disorders like depression or bipolar disorders, are also involved in emotion processing [TA12]. If such regions are chosen as target regions for NF therapy and emotional or social feedback is given, it can be assumed that these regions will show changes in their activation following the perception and processing of the emotional stimulus delivered by the feedback signal. In this case it cannot be distinguished between the activation change due to the perception of the feedback and the change due to regulation strategies of the subjects.

This additional influence of the feedback should be integrated into the analysis of NF studies to improve the accuracy of the estimated regulation success, to present more precise feedback and thus optimize therapy conditions.

### 1.2 Aim of the work

The aim of this work is to investigate the presence and magnitude of the influence of emotional feedback stimuli on the measured neuronal activation in fMRI NF settings. For this purpose, the presented feedback is integrated into the analysis of the primary data and the statistical significance of this effect is estimated.

In case of a significant variation of the signal due to the feedback perception, the results of this work could be used for more effective feedback presentation during the measurement, by taking this effect into account in common NF software and the reanalysis of the measurements where this was not done.

### 1.3 Structure of the work

In Chapter 2, an introduction to the principle of neurofeedback is given and existing and future medical applications are presented in order to discuss the medical background and clinical relevance of this work. In the second part of that chapter, the method of

functional MRI (fMRI) is explained in more detail. This section is mainly necessary to understand the special data processing steps that are necessary for fMRI measurements and have thus been carried out within the scope of this work.

Since the problems dealt with in this work primarily affect regions of the brain that are involved in the processing of emotions, Chapter 3 deals with emotion processing in health and disease. Chapter 4 summarizes the necessary mathematical fundamentals, which are necessary in order to be able to understand the analysis carried out, and the connection to fMRI settings.

Social or emotional feedback has been used in a number of studies, usually during neurofeedback training in a region that is involved in emotion processing and is therefore affected by the problem of additional activation through the feedback. These studies and how they deal with this problem are discussed in Chapter 5. In addition, the study that provides the data for this work is presented in detail.

Then the methods that are used to analyse a possible additional feedback influence are explained, as well as the necessary implementation steps in MATLAB and associated software. Different model approaches to investigate the problem are provided. Chapter 7 presents the results for all model approaches, both at the level of the individual subject and at the group level.

The interpretation of these results takes place in Chapter 8. The limitations of the methods used are also discussed there and an outlook on potential next steps is given. The conclusion in Chapter 9 finally summarizes the findings of the entire work.



# Neurofeedback in clinical neuroscience

Neurofeedback is a kind of biofeedback that enables people to receive direct information about the ongoing activity in their brain, which is usually inaccessible to the conscious mind. In this way, a deliberate manipulation of this activity can be practiced by trying different cognitive strategies and observing their effects on the corresponding brain activity. As a result, this consciously changed activity can cause changes in the perception and even the behaviour of the trained person.

This phenomenon can be used, for example, to improve the performance of athletes by training their psychomotor and self-regulating skills [ET11]. Improved performance after neurofeedback training was also found in both artists and surgeons [MSPJG08] [Ver05] [Bud09] and Escalano et al. observed increased memory performance in subjects who had previously increased the alpha activity in their electroencephalography (EEG) spectrum with neurofeedback [EAM11]. Furthermore, one of the most important aspects of using this training method is that it offers a therapeutic approach for diseases that are based on changed neuronal conditions compared to healthy people. A disease-related increase or decrease in the activity of a certain brain region or network can in this way be counteracted. Successes in this direction have already been demonstrated for attention-deficit/hyperactivity disorder (ADHD) [LS76] [AdRS<sup>+</sup>09], epilepsy [SE06] [TTH<sup>+</sup>09], schizophrenia [RLS<sup>+</sup>13], and Parkinson's disease [SHJ<sup>+</sup>11]. In this work, the focus will mainly be on the clinical application of neurofeedback, as this is where the greatest relevance for improved designs lies.

In principle, any imaging method, which is able to map the neuronal activity or subsequent metabolic reactions in a time that enables a useful connection to be established between cognitive strategy and neuronal reaction, can be used to carry out neurofeedback. The oldest method, which is still used in most cases today, is EEG [ADFS16]. But also other methods, such as functional near-infrared spectroscopy (fNIRS) or functional magnetic resonance tomography (fMRI) are nowadays often used for neurofeedback measurements.

The following sections explain how neurofeedback can work and in which areas there are already successful or promising therapeutic approaches.

After that fMRI neurofeedback (fMRI NF) is explicitly discussed. On the one hand, because this method was used in the study on which this work is based and, on the other hand, because it offers decisive advantages over other methods. These advantages are explained in more detail in the last section, as well as the disadvantages that must be accepted if choosing this method.

## 2.1 Introduction to neurofeedback

The discovery of the principle of neurofeedback took place in the course of an EEG study with cats carried out by Stermann et al. [SWR69] in 1969. The purpose of the study was to investigate to what extent previously trained behavioural patterns can be suppressed in order to obtain a reward. It turned out that cats are able to regulate certain EEG waves if they are rewarded with food. In a follow-up study in which epileptic seizures were to be stimulated in these cats, the trained cats showed a significantly higher seizure threshold compared to untrained cats [SE06]. This hour of birth of neurofeedback was thus at the same time the hour of birth of its therapeutic effects. Since those early days, scientific interest in neurofeedback has grown rapidly. Under the keyword "Neurofeedback", currently around 25,500 entries can be found on Google Scholar, while this was only 1,500 at the turn of the millennium. The exponentially increasing number of studies on this topic is shown in Figure 2.1. These studies are concerned with the application of various imaging methods, the improvements of design, the illumination of the mathematical analysis, and the possible applications.

Neurofeedback training has meanwhile developed into a valuable therapeutic tool, so that the main focus of scientific research lies on its development and optimization, as well as its adaptation to new clinical fields [ADFS16]. Since the 1970s, subjects have been trained to deliberately control different components of the EEG spectrum [Bir06][BC07] [SLG<sup>+</sup>06], as different waves are known to be linked to different mental states or tasks (e.g. sleep, problem solving, recalling, learning, etc.) [MMM16]. In this way, the neuronal 'source' of a disease can be specifically targeted. The advantages of such a new therapy method are primarily that no persistent side effects have been found so far and it is non-invasive and non-medicinal [DA13]. It also supports the feeling of self-efficacy, which is an important part of the healing process for all diseases [Ban97]. Other forms of therapy that go in a similar direction, e.g. cognitive behavioural therapy, rely on inaccessible information about cognitive processes [BRSE87], while control over this processes can be achieved with neurofeedback.

In the following, the basic principles and functionality of this method will be explained, and some interesting and promising therapeutic approaches will be presented. Since the aim of this work is to optimize the given feedback, the last subsection gives a brief overview of the feedback schemes used to far.

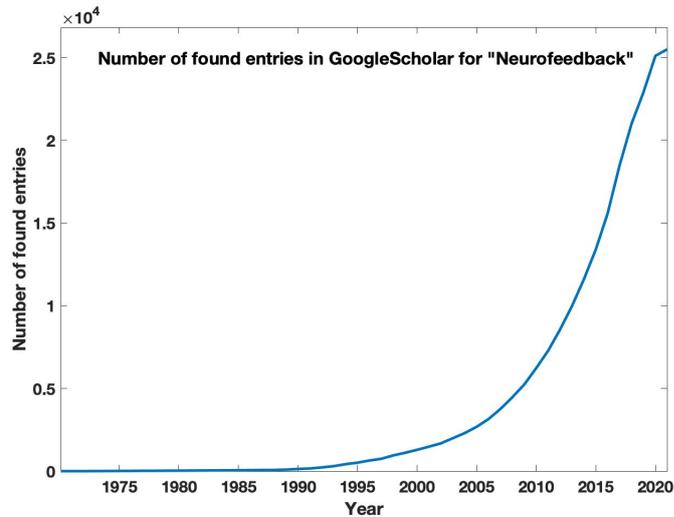


Figure 2.1: **Entries available on GoogleScholar up to the respective year for the keyword "Neurofeedback"**. Since the first attempt in 1969, the number of studies on this topic has grown exponentially. The studies are concerned with the application of various imaging methods, the improvement of the design, the illumination of the mathematical analysis and the possible applications, especially in a therapeutic environment.

### 2.1.1 Intention and basic principles

In neurofeedback training, the brain activities of the participants are measured in real time and they are instructed to either increase or decrease the measured activity to get a certain response [Wei12] [LaC11]. Either the level of the activity is returned directly as feedback, e.g. in the form of a graph that changes its level depending on the activity, or the performance when regulating this activity is fed back, for example in the form of a smiley that smiles when the task of decreasing the activity has been successfully fulfilled. In this way, subjects are enabled to self-regulate their brain activity [RLS<sup>+</sup>13].

A training session usually consists of several training runs, which are separated from each other by a longer resting period. One run is usually further divided into resting blocks and regulation blocks. The actual training takes places during these regulation blocks, i.e. the subjects should regulate the activity during this time [Wei12]. The feedback can be given continuously so that, for example, the information about the performance is updated every second, or as intermittent feedback, i.e. at the end of the block for the entire performance [Wei12]. While continuous feedback has the advantage that changes are visible very quickly and strategies can be adapted immediately [dMG<sup>+</sup>05], intermittent feedback is significantly less susceptible to noise, as it assesses the overall performance so that single peaks in the measurement do not play a major role [JHL<sup>+</sup>12]. In rare cases, the feedback is only given at the end of the entire run or even the session, for example if the performance is rewarded with money [BSO07].

In many studies, so-called ‘transfer runs’ are carried out before and after training. In these runs it is observed whether the participant is able to regulate the measured activity even without actual feedback. If the training was successful, the subjects should be able to implement the learned strategies without constant feedback. This generalization ability is necessary to show that neurofeedback also has effects beyond the training setting.

In order to test whether the measured successes or failures may have causes other than neurofeedback, e.g. physical rest, scanner noise, concentration on the task, or observing a variable signal, a control group is often designed in such a way that it undergoes a pseudo training in which incorrect feedback is shown, for example from another brain region or another participant [DA13].

While EEG neurofeedback usually trains frequency bands in the EEG spectrum or the differences in the potentials between the brain hemispheres, fMRI in most cases regulates clearly defined brain regions. In order to precisely identify these regions for later training, either a brain atlas is used or a functional localizer is carried out [DA13]. With the functional localizer, the needed brain region is identified by a suitable task. If, for example, a region that is involved in processing negative emotions is wanted, images of sad people can be shown and the area with the strongest measured activation can be selected for training. Figure 2.2 shows an exemplary functional localizer of various language tasks: Different brain areas are activated, depending on whether the subject is listening to a story, reading it silently, etc.

In order to train the defined target region, cognitive strategies can be used in a similar way. To regulate a region involved in emotion processing, it can be useful to imagine appropriate facial expressions [DLS<sup>+</sup>19]. If the activity of a motor area is to be increased, kinaesthetic imagery is often used, in which the subject imagines, for example, moving the right hand without actually doing it [ADFS16]. In this way, situations, in which certain tasks are voluntarily carried out, are simulated in the scanner and certain brain regions are activated in response. In real-life situations, the activation of the brain region is not a conscious process, but in neurofeedback training the (mental) execution of an action can be associated with an activation or deactivation of a brain region and thus brought into consciousness [dMG<sup>+</sup>05] [deC07].

In order to understand how neurofeedback works and to be able to use it in therapeutic settings, it is essential to understand the psychobiological process that enables the learning of self-regulation. This is one of the most important, still open, topics in the research field of neurofeedback [EKS<sup>+</sup>16] [RBR<sup>+</sup>14] [SHS<sup>+</sup>13]. On the one hand, the process of regulating brain activity resembles humans’ ability to consciously regulate their attention and commitment up or down [MG15] [RMR<sup>+</sup>10] [STH<sup>+</sup>12]. On the other hand, the neurofeedback scheme follows a classic operant conditioning setting: The participant receives a reward for adapting the behaviour (in form of brain activation) according to explicitly or implicitly made demands. Sepulveda et al. [SSR<sup>+</sup>16] were able to show that a learning effect also set in when the subjects were not aware of which strategies could be useful to activate the corresponding brain region or did not even know what kind of region should be trained at all. Even if it was not known that it was a training situation

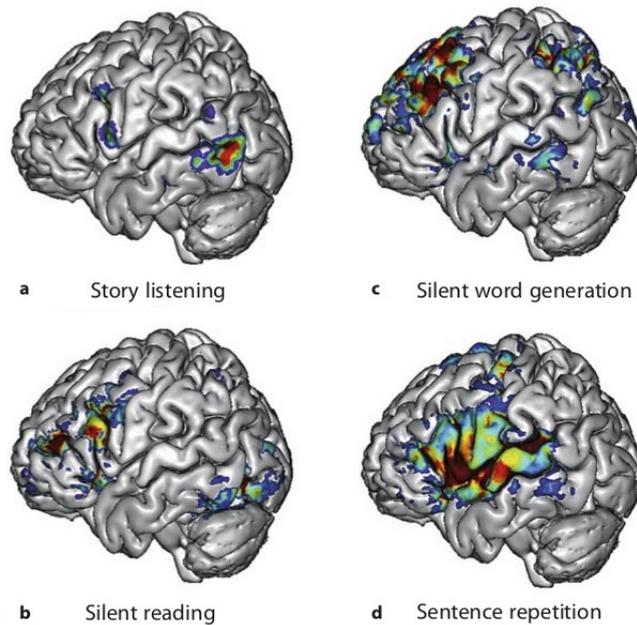


Figure 2.2: **Brain areas activated by different tasks concerning language [PMN11].** With the help of a functional localizer the specific brain regions involved in a specific task can be identified. These regions can further be used for neurofeedback training.

at all, in which the brain activity should be regulated, after a while an adjustment of the brain activity in the rewarded direction became apparent [KWN<sup>+</sup>13] [WKN<sup>+</sup>13]. These findings suggest that the participants used control strategies that occurred subconsciously without explicit cognitive effort. Marxen et al. and others [MJM<sup>+</sup>16] [SSR<sup>+</sup>16] therefore advise against proposing strategies, as it limits the subjects' freedom to develop the method that is best for them.

Successful neurofeedback training results in specific behavioural changes that can be traced back to altered brain function. For example, a change in the reaction time could be determined after regulation of motor areas [BSO07], or a change in pain sensation depending on the activity in the anterior cingulate cortex [dMG<sup>+</sup>05]. These changes in brain function and behaviour can last for weeks [SHJ<sup>+</sup>11] [YLO<sup>+</sup>07] [YLO<sup>+</sup>08], or even months to years [ASK<sup>+</sup>16] [MYKI15] [RKGC<sup>+</sup>17] [RMK<sup>+</sup>17]. Both the level of self-regulation and the learning effect are influenced by various factors, such as the type of feedback, the reward, the instructed mental strategies, and the duration of the session, amongst others [SA03] [SHS<sup>+</sup>13]. A precise design, which is adapted as fittingly as possible to the respective subject, can therefore contribute significantly to success. In this way, neurofeedback makes a decisive contribution to basic research: With its help, the knowledge gained so far, as to which behaviour activates which brain region, can be completed by non-invasively examining which behaviour is triggered by activated brain regions [DA13].

However, there are still some difficulties to overcome before neurofeedback can be used on a larger scale, e.g. as a therapeutic tool. Only about two-third of all subjects examined so far are able to gain control over their brain activity [FWSN14], half of them only after prolonged training. In addition, in many cases no control could be determined in the transfer runs (f.ex. [RLS<sup>+</sup>13] [DLS<sup>+</sup>19] [HGH<sup>+</sup>11]) without feedback, which prevents an effect beyond the training scenario. In order to find a solution to these problems, it is necessary to further improve the process. This thesis is intended to contribute to being able to give more precise feedback, one of the aspects that could have a decisive influence on learning success.

### 2.1.2 Medical application and techniques

Neurofeedback in therapeutic applications aims to train the control of physiological targets in order to induce changes in specific neural networks, which could be useful to bring pathologically altered cognition and behaviour back to a healthy level [SGG<sup>+</sup>14]. In the ideal case, in this way a skill is learned that enables the subject to gain control over affected brain regions even in everyday situations.

In order to test whether a certain region can be treated with neurofeedback, it is first necessary to find out whether this region can be regulated at all with the aid of neurofeedback. The regions that are tested for this controllability are usually already selected with regard to possible therapy options for neurological or psychiatric disorders: If, for example, the aim is a therapy for Autism Spectrum Disorder (ASD), an interesting object of investigation is the superior temporal sulcus, since it is involved in (dynamic) face recognition, and in the task of empathizing with the mental state of others. Also it shows a disturbed function with ASD [DLS<sup>+</sup>19].

EEG neurofeedback has been used successfully for a long time, primarily to treat neurological diseases. Here again, the advantage of neurofeedback compared to conventional treatment is that medication often has unpleasant side effects or does not work in all patients. Successful therapeutic approaches with EEG exist for epilepsy [SE06] [TTH<sup>+</sup>09], insomnia [CGT<sup>+</sup>10], migraine [DDOB12], obesity [FLP<sup>+</sup>12], or chronic pain [dMG<sup>+</sup>05]. ADHD is one of those examples in which classic medication such as Ritalin has no effect on some children or lead to severe side effects, such as anxiety, abdominal pain, insomnia, or headaches [MMM16]. It was shown that EEG neurofeedback has a similar therapeutic effect as stimulant drugs. The first training attempts took place in 1976 already [LS76] and major effects on impulsivity and inattentiveness were found, as well as medium effect sizes on hyperactivity [AdRS<sup>+</sup>09].

Neurofeedback can also be used for very severe neurological disorders, such as Parkinson's disease. Patients who underwent such fMRI NF training learned to increase the activity in the somatomotor cortex by using motor imagery and thereby improved the motor speed of finger tapping [SHJ<sup>+</sup>11].

The situation is a little more complex with psychiatric illnesses and mental disorders. Some mental diseases have heterogeneous causes, but if a causal relationship between the illness and dysfunctional activity can be shown in a particular brain region or network, the ability to regulate that region may have a positive effect on the patient's mental health

[DA13]. Until 2013, no successful application of neurofeedback in psychiatric populations had been presented [RLS<sup>+</sup>13]. Nowadays however, there exist some promising approaches for serious psychiatric disorders [LHJ<sup>+</sup>12] [MMM16].

In patients with schizophrenia, various aspects of cognitive functions are impaired, such as working memory, attention performance, verbal production, and face recognition [SJH<sup>+</sup>02]. So far, no drug has been found that could improve these social and cognitive limitations, but cognitive behavioural therapy has shown some success [PBK<sup>+</sup>02] and it can be assumed that neurofeedback could have similar effects. There seems to be a neural correlate, the anterior cingulate cortex (ACC), whose hypoactivity plays an important role in the development of these deficits [SJH<sup>+</sup>02]. In addition, volume reduction [TSH<sup>+</sup>04] and perfusion abnormalities [CFWA<sup>+</sup>01] often occur in this brain region in connection with schizophrenia. Cordes et al. [CMD<sup>+</sup>15] were able to show that control of the ACC is possible, and Ruiz et al. [RLS<sup>+</sup>13] examined the clinical effect of volitional regulation. The success in regulation actually correlated negatively with the severity of the symptoms. Additionally, faces expressing disgust could be recognized more easily and the duration of the illness was significantly shortened.

The picture is similar in major depressive disorder. About 30% of patients do not respond to standard pharmacological or psychological treatments [RTW<sup>+</sup>06]. The remaining 70% also experience problems during therapy due to serious side effects of the medication, such as sexual dysfunction, gastrointestinal problems, or weight gain [BFF09]. Apart from the fact that neurofeedback could provide a remedy in these cases, it would also make sense to use it in previously unproblematic cases: Depression is often accompanied by a general feeling of powerlessness and a low sense of agency [Ban97]. Linden et al. [LHJ<sup>+</sup>12] even make the daring thesis that successful control over one's brain activity could convey the feeling of control and the experience that the own thoughts can be constructed, so that the exact choice of the trained regions even plays a subordinate role. In general, however, NF training targets regions that are related to depression. In this field, some successes can be seen: Increased activation of amygdala [YZP<sup>+</sup>14] and other regions (f.ex. [LHJ<sup>+</sup>12] [HGB<sup>+</sup>16]) with emotional connotations, lead to reduced symptoms. After EEG neurofeedback training, the patients were able to achieve an increased processing speed of the working memory [ENGGC<sup>+</sup>14]. Similar to schizophrenia, however, it is still a long way before neurofeedback could be used as a comprehensive therapy for depression. A certain social relevance also plays a role, since EEG neurofeedback already showed initial success with criminal psychopaths, in that trained patients showed reduced aggression and impulsiveness and a greater sensitivity to error processing than untrained patients [KVE<sup>+</sup>15].

Further results in the direction of neurofeedback could therefore contribute to an increased mental and neurological health of the population as well as to new ways of dealing with mentally conspicuous offenders.

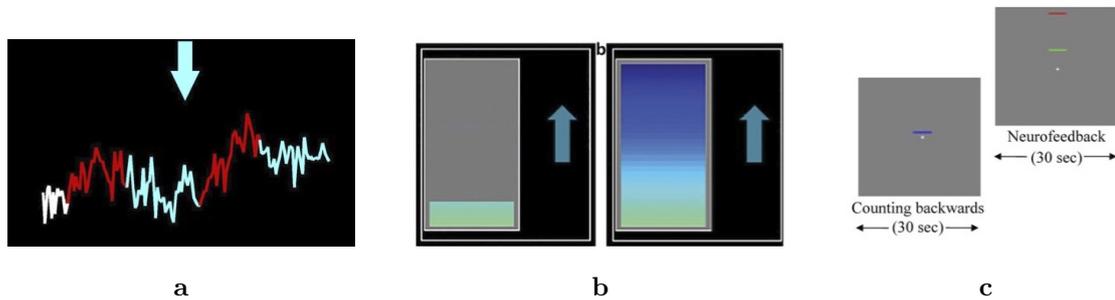


Figure 2.3: **Classic simple feedback schemes.** In order to show the level of current brain activity, in most cases either a continuously updated graph (a) [GFR<sup>+</sup>16], a thermometer with changing colours (b)[ZWB<sup>+</sup>19] or a bar (c) [MAK<sup>+</sup>15] that moves up and down is used. The advantage of such a simple feedback design is that no cognitive load is deducted from the actual task [KBL<sup>+</sup>17].

### 2.1.3 Different kinds of feedback

The question of which feedback is best suited to achieve optimal learning success has already attracted a lot of interest in the history of neurofeedback. Very different approaches were taken, with different reasons and foci. Some of these ideas are briefly presented here.

First of all, it has to be decided which sensory organ should be used to convey the feedback. In principle, all of them are possible, but due to practical reasons, tactile, olfactory, or gustatory feedback is usually renounced. Auditory feedback can, for example, be given in the form of a pleasant or unpleasant tone in an MRI scanner environment, but mostly visual feedback is used, which is presented to the subject on a screen in the scanner. This preference is due, on the one hand, to the ease with which it can be implemented, and, on the other hand, to the fact that it has been shown that visual feedback delivers better learning results than auditory feedback [HNP<sup>+</sup>04]. This finding fits in with the fact that vision is the most dominant human sense [DA13].

The classical design of such visual feedback consists of a graph that is continuously updated to show the current level of the measured activity (Figure 2.3 a). Similar simple forms of representation consist of a thermometer with varying height (Figure 2.3 b) or a bar that moves up and down (Figure 2.3 c). In contrast to more complex approaches, the advantage of such simple feedback schemes is that only a low level of cognitive performance is required for the perception and interpretation of the feedback [KBL<sup>+</sup>17].

Other authors presented feedback to the subjects that was in terms of content related to the region to be trained. The idea behind this approach is that such a feedback mimics an answer that would also occur in real life. DeCharms et al. [dMG<sup>+</sup>05] presented a video of a larger or smaller burning fire (Figure 2.4 a) to demonstrate activity in a region involved in pain processing. If the subject manages to regulate the pain, only a small fire is displayed to match the 'small pain'. In another example by Sokundi et al. [SLI<sup>+</sup>14], the feeling of hunger was supposed to be regulated; If this was successful, the



Figure 2.4: **Feedback design that has a content-related connection to the trained brain region.** The idea behind such a representation is to mimic the answer that would occur in real-life regulatory situations, such as resisting a tempting looking but unhealthy food. In **a**, the size of the fire symbolizes the activity of a region that is responsible for the sensation of pain [dMG<sup>+</sup>05], in **b**, the size of a picture with a food motif is adapted to the activity of an area that is responsible for hunger [SLI<sup>+</sup>14].

displayed image of an appetizing food was reduced in size (Figure 2.4 **b**), the 'temptation' therefore became smaller as a result of the regulation. This design should simulate real-life situations in which it is, for example, necessary to refrain from eating due to a diet.

One approach that also focuses on simulating real-life situations relies on the use of social feedback. The first steps in this direction by Mathiak et al. [MAK<sup>+</sup>15] showed improved results in the training runs if the subjects were presented with a human avatar whose facial expression changed depending on the subject's performance. While a neutral facial expression was shown in the baseline or rest period, the avatar began to smile when the activity of the target region was regulated in the desired direction (Figure 2.5). The main benefit of such emotionally connoted feedback is that it offers a very intuitive reward and simulates real-life regulatory measures, which also attempt to adjust behaviour in order to receive positive social feedback. Chapter 5 goes into more detail on the benefits of social feedback.

In order to increase the motivation of the participants and to alleviate negative aspects of the training situation, e.g. severe fatigue, another possibility is to give the feedback by integrating it into a game. Following this approach, which is particularly suitable for children and young people, several designs have been tried. Often the point is to move a virtual character, e.g. to guide a Lego robot through a landscape [MSS13] or to let a fish swim towards food (Figure 2.6) [SCV<sup>+</sup>07]. If the technical infrastructure is available, it is even possible to have two subjects play with each other or against each other, as in the brain-pong game by Goebel et al. [GSK04].

In addition to receiving continuous feedback via various sensory organs, there are also experiences with giving a monetary reward [BSO07]. The participant is informed during

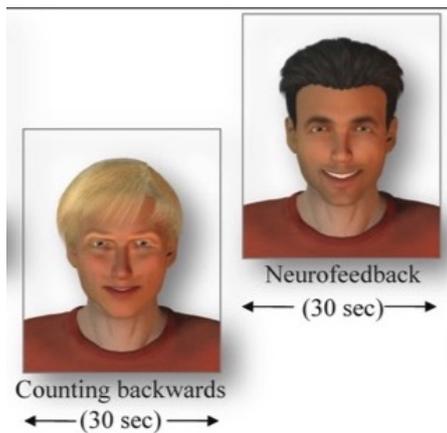


Figure 2.5: **Social feedback using changing facial expressions of a human avatar** [MAK<sup>+</sup>15]. The main benefit of such a feedback is the intuitive reward and simulated real-life social behaviour.

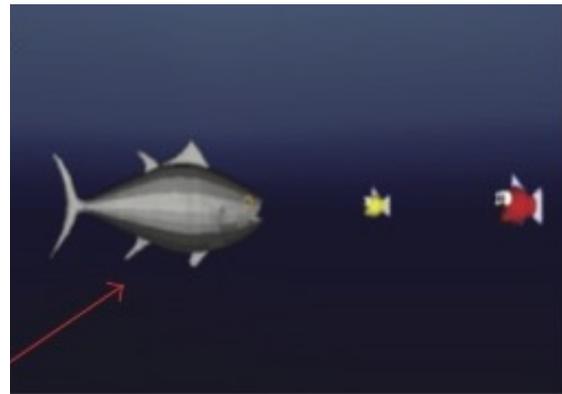


Figure 2.6: **Regulation of a game character as feedback.** In order to increase the motivation and fun of training, some authors rely on the control of computer games with the help of regulated brain activity. In this picture, a virtual fish is controlled with the help of fMRI neurofeedback to swim in the direction of smaller fish to eat them [SCV<sup>+</sup>07].

the training if a certain threshold of activation is reached and thus, for example, one dollar has been added to the profit. The entire profit will be paid out at the end of the training.

The idea behind this feedback is to implement instrumental conditioning adapted to humans, as it is also successfully used with animals.

The multitude of possibilities to give feedback, all of which lead to the successful regulation of the corresponding brain activity, opens up perspectives for an individual adaptation to the respective subject, the selected brain region and the technical possibilities that shape the special neurofeedback training.

## 2.2 fMRI Neurofeedback

Neurofeedback that is based on fMRI is a method for learning how to gain control of almost any region in the brain, a decisive advantage over approaches based on EEG or near-infrared spectroscopy (NIRS), which are mostly restricted to the measurement of cortical brain activity. The study of emotional processing and affective disorders is difficult and limited with EEG since important subcortical areas, such as the amygdala, cannot be reached. Technical advances in fMRI, such as multi-echo planar imaging (EPI) [PWG<sup>+</sup>99][SH98], and the development of real time fMRI (rtfMRI) by Cox et al. [CJH95], reducing the data processing time to a tolerable low rate, have made this method a useful tool in a variety of basic and clinical applications, including diagnosis, disease monitoring, and therapeutic uses based on neurofeedback [Wei12].

The following sections provide a brief overview of the basic principles of fMRI and outline the advantages and disadvantages of this method compared to other imaging and therapy methods.

### 2.2.1 Short overview: MRI and fMRI

In order to highlight the advantages and disadvantages of using fMRI in neurofeedback applications, it makes sense to briefly discuss the physical and physiological principles of this imaging method. A more in-depth explanation can be found in many standard medical works; the following overview and the images are based on A.L. Beart's book 'Clinical functional MRI' [Goe07].

#### 2.2.1.1 Physical principles of MRI

This section briefly outlines the physical principles of magnetic resonance imaging (MRI), which makes it possible to generate both anatomically and functionally resolved images of the human brain. The recording of an MRI image essentially consists of two steps: At first, the electromagnetic signal, which reflects the material-specific properties, is generated and measured. In a second step, a three-dimensional image is reconstructed based on these signals.

The underlying principle of MRI is based on the magnetic stimulation of body tissue and the measurement of the signal returned by the body. Such a magnetic excitation is possible for all atoms that have an odd number of protons. The atom of choice in this case is the hydrogen isotope  $^1\text{H}$ , which consists of a single proton. It has well-known magnetic properties and is abundant in body tissues.

As with all elementary particles, protons have a quantum mechanical property called 'spin'. As a model, this spin can be assumed to be a rotation around its own axis, which creates a very small directed magnetic field. Usually, the spins of the hydrogen atoms in the body are randomly oriented and cancel each other out, so that the body as a whole has a magnetic field strength of approximate zero (Figure 2.7 a). If the body is placed in a strong external magnetic field, called  $B_0$ , the spins of all protons align according to the field lines of this external field, either parallel or antiparallel (Figure 2.7 b). It is energetically more favourable for the single atom to align the spin parallel to  $B_0$  since it requires more energy to flip it to an antiparallel orientation. Therefore, a larger part of the atoms is aligned parallel to the external magnetic field, and the magnetic fields of the spins no longer completely nullify each other. Instead, there is a small additional magnetic field that is oriented in the same direction as  $B_0$ , called  $M_0$ , that consists of the sum of the surplus atoms with a spin parallel to  $B_0$ . The strength of  $M_0$  depends on the strength of  $B_0$ : The stronger the external field, the greater the energy required for a single atom to reverse its spin direction. Just as a spin top in the earth's gravity field begins to revolve its axis of rotation around the axis of gravity (a phenomenon called 'precession'), the spin of a proton precesses around an axis parallel to the external field

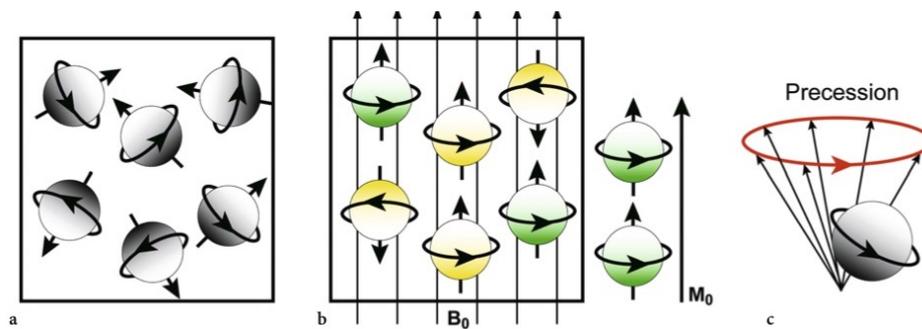


Figure 2.7: **Protons act like little magnets because of their spin property [Goe07]. a:** Without an external magnetic field, the directions of the spins are randomly distributed and in sum cancel each other out. **b:** Under the influence of an external magnetic field  $B_0$ , the spins orient either parallel or antiparallel to the field lines. Since it takes more energy to align antiparallel, a slight excess of spins are oriented parallel, resulting in a magnetization of the body called  $M_0$ . **c:** Just as a spin top in the earth's gravity field, the spin start to rotate their spinning axis around an axis parallel to the external field. This phenomenon is called 'precession'.

(Figure 2.7 c). The frequency of this precession, called Larmor frequency  $\omega_L$ , is directly proportional to the strength of the external magnetic field:

$$\omega_L = \gamma B_0 \quad (2.1)$$

The factor  $\gamma$  is called 'gyromagnetic ratio' and is specific to the type of atom that is being examined. If an energy pulse  $B_1$  with the corresponding Larmor frequency is sent out, this energy can be completely absorbed by the protons, resulting in a 'resonance', which means that the protons are set into a strongly excited state in which more atoms have the energy to change the direction of the spin to 'antiparallel'. The resulting  $M_0$  vector no longer points exclusively in the direction of the  $B_0$  field, but is flipped by a certain angle  $\alpha$  depending on the number of rotated spins. If there were the same number of spins in the 'antiparallel' as in the 'parallel' orientation,  $M_0$  would be rotated by  $90^\circ$  (as in Figure 2.8 a). In addition, this energy pulse means that all spins are brought 'in phase', that is, they all point at the same time to the same point in the precession circle. At the time of excitation, the resulting vector consists of the sum of the amounts of the single spins and rotates with  $\omega_L$  around the  $B_0$  axis (Figure 2.8 b). The horizontal component of  $M_0$ ,  $M_{xy}$ , can be measured with a receiver coil as it induces a small but measurable current in the coil.

The protons do not stay in this excited state for long. Interactions with the surrounding tissue lead to a loss of energy, and an increasing number of spins fall back into the position parallel to  $B_0$ . The time until the (almost) complete restoration of the original  $M_0$  vector is referred to as  $T_1$  and is strongly dependant on the material. For an fMRI measurement, however, two other parameters are decisive. Since the magnetic fields of the single spins interact with each other, they quickly dephase, which means that each spin precesses with a different frequency. The greater the difference in phase between the

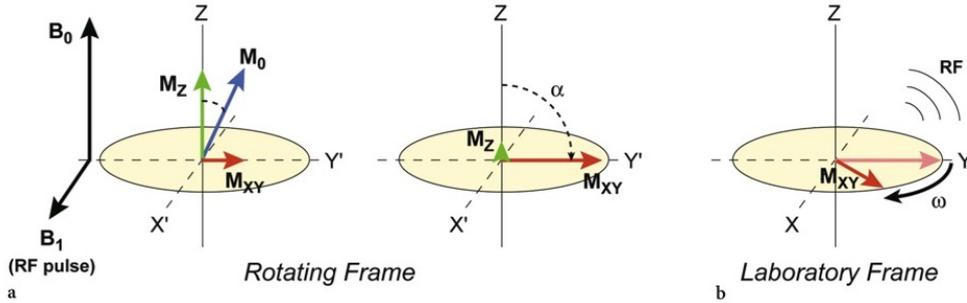


Figure 2.8: **An electromagnetic pulse with the according resonance frequency leads to an excitement of the proton spins [Goe07].** **a:** Following this excitement a higher number of spins turn into the antiparallel direction and the resulting magnetic field vector  $M_0$  is flipped by an angle  $\alpha$ . As can be seen on the graphic in the middle the horizontal component  $M_{xy}$  increases, while the transverse component  $M_z$  decreases. **b:** As an additional effect the electromagnetic pulse leads to an in-phase precession of all spins. The resulting vector is the sum of the total amounts of all single spins and rotates with the Larmor frequency around the external field direction.

single spins, the smaller the resulting vector  $M_{xy}$ . The corresponding decay process is shown in Figure 2.9 as an example for three spins. It follows the law

$$M_{xy} = M_0 \cdot e^{-t/T_2^*} \quad (2.2)$$

whereby the decisive time constant  $T_2^*$  is the time until the horizontal component is nearly equal to zero and no more signals can be measured. The speed of the dephasing and thus the decrease of  $M_{xy}$  is an important observation for fMRI, since it depends on the physiological conditions of the tissue, for example the local blood oxygenation. The state of completed dephasing, with  $M_{xy} \approx 0$ , can be reversed with an  $180^\circ$  pulse. The single spins are ‘flipped’ around an axis so that they run completely in phase again after twice the decay time has elapsed and the process can be restarted. As a model, this process corresponds to a racetrack with runners at different speeds who are asked to run back after a certain time. If they run as fast as before, they should all be back at the start at the same time and be able to start over. The time until the restart of the dephasing is called echo time (TE). However, the dephasing is not completely reversible, as part of the energy is irretrievably lost through random interactions. The whole process over several echo times is shown in Figure 2.10. As a result of the processes described above, the maximum amplitude of  $M_{xy}$  becomes smaller over time; this decay also follows an exponential law with the time constant  $T_2$ . After the maximum signal can no longer be measured, the spins have to be excited again with a high-frequency pulse. The time between two excitations is called repetition time (TR).

In the second step, an image must be reconstructed from the measured magnetic signals. To do this, it must first be determined from which part of the body a corresponding signal originated. This is where additional gradient coils come into play, which

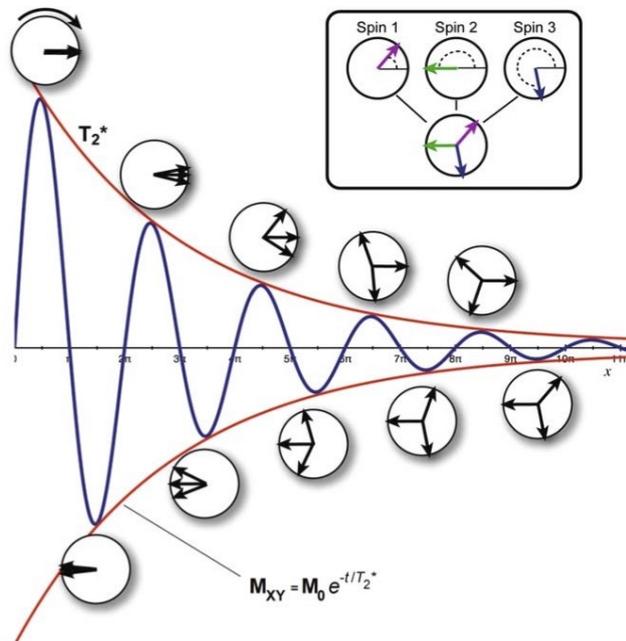


Figure 2.9: **The signal amplitude of the measured MRI signal decays exponentially with the time constant  $T_2^*$  [Goe07].** The raw signal oscillates with the resonance frequency. The signal is lost with de-phasing, as indicated for three single spins. When the spins are completely out of phase, the signal is completely lost, which correspond to a zero length of  $M_{xy}$ .

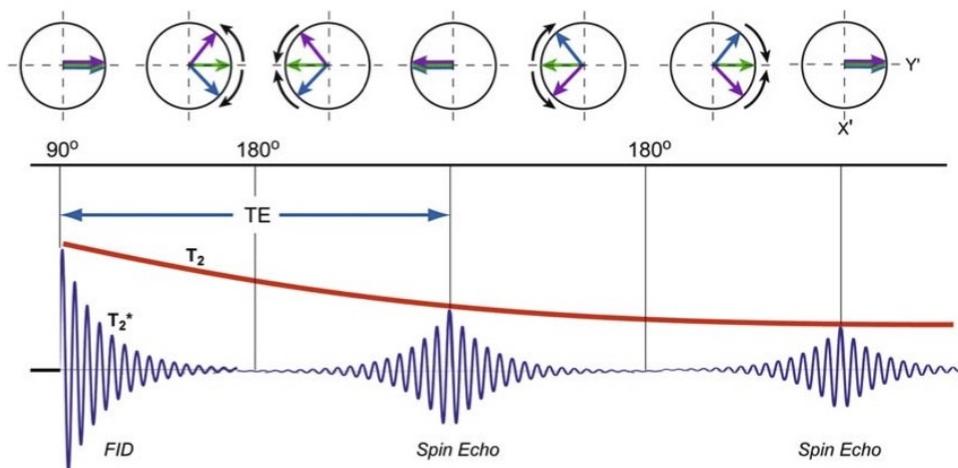


Figure 2.10: **The process of de-phasing can be reversed by the application of a  $180^\circ$  pulse, which flips the vectors around an axis, so that they are in phase again after the echo time (TE).** The maximum amplitude of the echoes gets smaller over time due to an inevitable loss of energy caused by random interactions [Goe07].

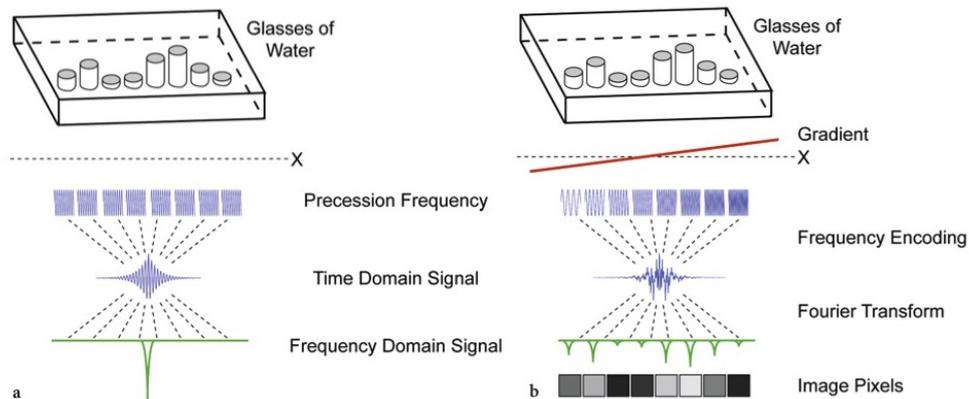


Figure 2.11: **The principle of frequency encoding using glasses of water with different heights as example [Goe07].** The glasses are placed in a MRI scanner along the x-axis and a single slice containing all glasses has been excited. **a:** In the absence of any gradients, all spins from all glasses precess with the same frequency. The measured signal oscillates with the same frequency and its amplitude reflects the sum of spins. The spatial positions of the glasses are not estimable. **b:** If a gradient is applied in the x-direction, the spins precess at frequencies that depend on their position along the gradient. Spatial information is now frequency-encoded; the obtained signal is the sum of these frequencies. With an inverse Fourier transformation the amplitude of each frequency can be estimated, and an image for each of the glasses can be formed that reflect the respective height of water.

generate spatially variable magnetic fields. Due to the different magnetic field strength at each location, the spins of the protons precess with different Larmor frequencies and thus emit their electromagnetic signal with different frequencies (see Figure 2.11).

First, the measured volume is divided into slices. A slice is selected by an energy pulse that exactly covers the frequencies, with which the spins in the corresponding slice precess, and excite only them. Thanks to the gradient field, all other spins have a too high or too low Larmor frequency so that there is no resonance effect.

A second gradient field orthogonal to the first one is used for the spatial resolution of the first spatial axis of the slice. In this way, spins with different Larmor frequencies are created along this spatial axis while the signal is measured. With the help of an inverse Fourier transformation, the signal strength of each frequency can be determined and since this frequency depends on the distance in the gradient field, the starting position of this signal can be estimated along this spatial direction. This process is called ‘frequency encoding’ and is shown in more detail in Figure 2.11 using an example with water tanks of different heights.

Unfortunately, it is not possible to analyse the second spatial direction of the slice with frequency encoding as this would lead to ambiguous results. Instead, a procedure called ‘phase encoding’ is used here. For this purpose, an additional gradient field is applied along the last spatial direction, which does not permanently influence the spins, but only for a short moment before the echo is measured. This results in a frequency

difference for a brief moment during precession. After switching off the gradient all spins rotate again with the same Larmor frequency, but now have a phase difference, that is, each spin vector has a different direction. With a clever combination of frequency encoding and phase encoding, a two-dimensional image of the slice can be generated in the desired resolution. The phase encoding procedure must be carried out several times for each layer with different phase shifts in order to achieve complete encoding. The idea behind the repeated phase shifts is that they depend on the distance in the gradient field. For example, the spin of a proton close to the source experiences a smaller phase shift than the one of a proton located further away. The summed phase shifts after a few applications of the gradient ‘oscillate’ more slowly for a nearby proton than for a more distant one, since it takes longer for the phase shift to correspond to a complete precession period. By using this detour, an inverse Fourier transformation can then be carried out again, which leads to a spatial encoding in the last dimension.

A process called echo-planar imaging (EPI) is usually used to record fMRI images. All phase-encoding steps of a slice are measured during a single  $T_2^*$ -decay. In this way, the speed of the recorded images can be increased drastically, which is ideal for capturing neural processes, but as a downside a high performance of the scanner is required, since the gradients have to be switched very quickly. In addition, image distortions can easily occur due to inhomogeneities in the magnetic field, called ‘susceptibility artifacts’. These inhomogeneities are caused by transitions between different materials, for example brain tissue and air. For a precise analysis, these distortions must be taken into account and corrected in the preprocessing of the reconstructed images.

### 2.2.1.2 Physiological principles of fMRI

Neural activity consumes energy, which is made available in the form of glucose and oxygen. The supply takes place via the vascular system, which transports oxygen-rich blood from the lungs to the brain via the arteries. Oxygen is delivered to the corresponding nerve cells in a fine-grained vessel network, the capillaries, while the oxygen-poor blood is transported away via veins.

A local increase in neuronal activity thus leads to an increase in local cerebral blood flow (CBF) and cerebral blood volume (CBV) within a time span of five to six seconds in order to be able to provide sufficient oxygen. This response of the vascular system to neuronal activity is called ‘hemodynamic response’. The increase in CBF leads to a local oversupply of oxygenated blood and the deoxygenated blood is removed more quickly, as shown in Figure 2.12.

The difference in blood oxygenation can be used to record neuronal activity through an MRI signal. Oxygenated blood has diamagnetic capabilities, which means that it hardly influences the external magnetic field, while deoxygenated blood is paramagnetic, leading to a local change in the magnetic field, which is therefore less homogenous. As explained in the previous section, the less homogenous the field, the faster the de-phasing of the spins followed by a faster decrease of magnetic signal. With increased neuronal activity, the deoxygenated blood is transported away more quickly, so it can interfere

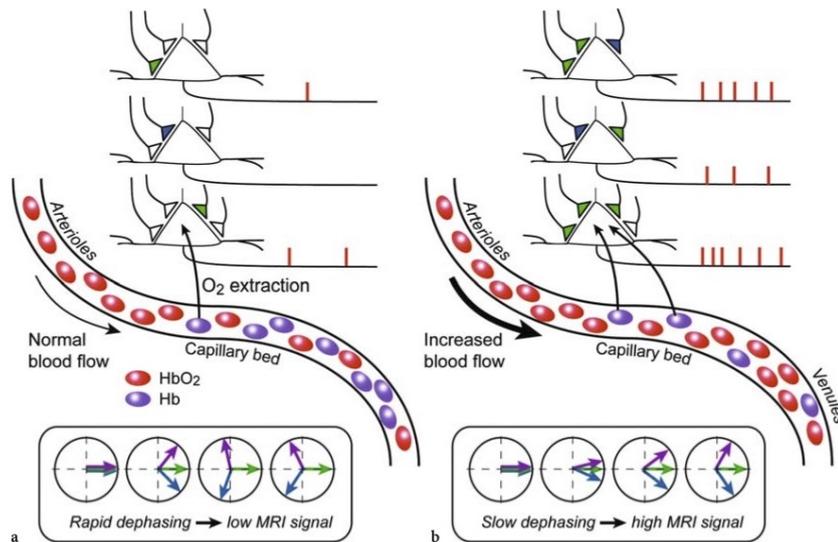


Figure 2.12: **From neuronal activity to BOLD signal [Goe07]. a:** At rest, neuronal activity and CBF are low. A constant oxygen extraction rate fuelling neural activity leads to a fixed ratio of deoxygenated (violet dots) to oxygenated (red dots) blood. Since deoxygenated blood is paramagnetic, it distorts the magnetic field, leading to a rapid de-phasing of excited spins and thus a low MRI signal. **b:** If the cortical region is in an activated state, the oxygen extraction rate increases. The high CBF flushes deoxygenated blood away from the capillary bed. Oxygenated blood does not substantially distort the homogeneity of the local magnetic field, excited spins dephase slower than in the baseline state resulting in an enhanced MRI signal (BOLD effect).

less strongly with the external field than in the resting state and an increased magnetic signal is generated (Figure 2.12). This signal is called blood-oxygenation level dependent (BOLD) signal.

Since the enlarged blood vessels need longer to return to their normal size than the CBF, there is an accumulation of deoxygenated blood and a short ‘undershot’ in the measurement of the signal, i.e. a signal that is lower than in the resting state. The resulting hemodynamic response is shown in Figure 2.13 a.

With the fMRI measurement, the neuronal activity is measured only indirectly via its metabolic processes in the form of oxygen supply, which leads to a significant delay in the signal as well as inaccuracies in rapidly succeeding stimuli, as the vascular system can no longer react to each individual stimulus. Thus, only a combined response is measured, as can be seen in Figure 2.13 b.

### 2.2.2 Advantages and limitations

The use of functional MRI as an imaging method for neurofeedback offers decisive advantages over other imaging methods. The most important advantage is the relatively high spatiotemporal resolution, which makes it possible to map the entire brain within a couple of seconds with millimeter accuracy [Log08]. In contrast to EEG and functional

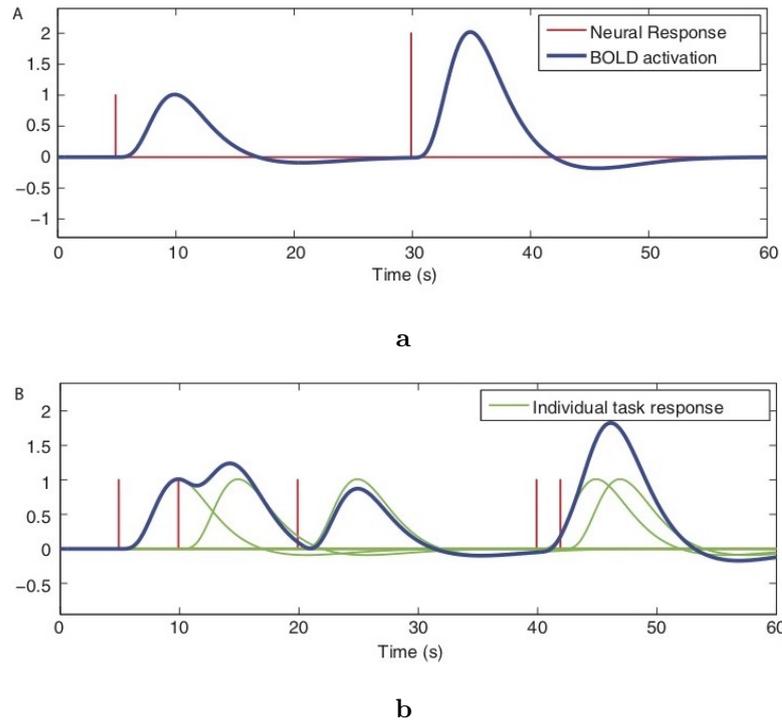


Figure 2.13: **The hemodynamic response to neuronal activation [PMN11]. a:** After a spike of a neuron (red line) the vascular system reacts with an overshoot of oxygenated blood. This excess can also be seen in the magnetic signal (blue line) with a delay of five to six seconds. It is assumed that the hemodynamic response is linear, that is, a neuronal signal that is twice as strong also generates a magnetic signal that is twice as strong. After the CBF has fallen back to a normal level, there is a brief excess of deoxygenated blood, which leads to an undershot in the magnetic signal. **b:** Since the hemodynamic response is assumed to be linear, the responses to neural spikes, which take place in quick succession, add up. However, if the two spikes are too close together, as shown on the right, the resulting BOLD signal can no longer be distinguished from that of a single, but stronger, spike.

NIRS (fNIRS), it is possible to reach deep cortical regions as well as to select a brain region or just individual voxels whose activation is measured. Furthermore, complex networks that are involved in a task can also be examined. With the help of functional localizers, it is therefore possible to define an individual area for each subject that is to be trained and in this way to make an important contribution to personalized medicine [LHJ<sup>+</sup>12]. Similar to most other imaging methods used for neurofeedback, fMRI is free of known long-term side effects, non-invasive, and drug-free, so that longitudinal studies can be performed that examine a single subject over time [DA13].

In order to be able to use these advantages of fMRI, however, some disadvantages have to be accepted. Since the magnetic resonance of the hemodynamic response is measured instead of the electrical activity of the nerves themselves, it is an indirect measure that has a relatively long time delay. Although this is sufficient for the measurement

of long-term neuromodulatory effects, such as arousal, memories, attention, etc., very short-term neuronal impulses are only inadequately recorded [Log08]. In addition, it has not been conclusively clarified whether fMRI NF actually modulates neuronal activity or just the blood flow through the target region [Wei12]. In general, the signal-to-noise ratio in fMRI is rather small, it can be improved by higher field strengths, but these also lead to increased tissue heating [GDK<sup>+</sup>15], which causes an uncomfortable training atmosphere. Both the spatial specificity and the temporal response are sensitive to motion and physiological changes, e.g. in breathing or heartbeat, leading to a reduced continuity between behavioural changes (e.g. changing the cognitive strategy) and the feedback signal [DA13]. Compared to an EEG setting, fMRI NF is a relatively time-consuming procedure and since movement should be avoided as completely as possible, the number of possible tasks is quite limited [LHJ<sup>+</sup>12].

A large part of these disadvantages could be mitigated either by combining fMRI with other methods or by more complex processing of the data [LHJ<sup>+</sup>12]. It remains subject to further studies if these possibilities can also be implemented in a neurofeedback context. Despite the limitations mentioned, neurofeedback based on fMRI is 'currently the best tool we have' [Log08].



# Introduction to emotion processing

Emotion processing is a primal function of the brain that is to be classified by tribal history prior to the emergence of language. Feelings are not limited to humans; also animals have feelings, process emotions, and align their behaviour to them. Almost all psychological variables are influenced by emotions; they have an impact on the perception of the environment, the self-image, memory, learning performance, and social behaviour [HS07].

In order to determine the influence of the perception and processing of the emotional or social feedback on the activation of certain regions of the brain, it is important to consider how emotions and feelings are processed. This includes the information which brain areas are involved, whether there are differences in the processing of different emotions, and what influence boundary conditions, such as the type of task, have.

Since neurofeedback may be used as a therapy method for various psychiatric diseases, it is necessary to take into account how emotion processing differs between patients with affective disorders and healthy subjects and to what extent this difference could affect the processing of social feedback.

In addition, the subgenual anterior cingulate cortex (sgACC) is presented as a sample region, since the data underlying this work describes this region as a target of neurofeedback training.

## 3.1 Methods and limitations of localizing the processing of emotions

Before the actual processing of emotions will be discussed, it must first be clarified what an emotion actually is and in which way it differs from a feeling or mood. Above all, this distinction is important because the three terms are often used synonymously in

common language. Dolan [Dol02] defines the difference as follows: An emotion is a short-term, fierce reaction with automatic reaction patterns, which is nevertheless felt specifically and deliberately. On the other hand, a mood or feeling is a longer-term reaction tendency that can promote or inhibit the occurrence of certain emotions. A mood is based on physiological changes due to the processing of emotion-releasing states or objects whose mental representation the current mood is. Therefore, it can also be assumed that emotions and feelings are mediated by different neuronal systems.

The focus of research is clearly on the short-term emotions, as a general mood has significantly more complex relationships and is unsuitable for a measurement setting due to its long-term nature.

Before the dissemination of imaging methods, brain research concerning emotions and their impact on behaviour was based mainly on animal and lesion studies. Since only long-term consequences of lesions, such as general behavioural changes, can be considered in such studies, emotion research was a long-neglected branch of neuroscience. The research interest significantly increased with the possibility to map and localize the emotion-specific correlates in the brain through imaging methods such as fMRI or PET that are able to produce whole-brain volumetric images [HS07]. Despite these newly created possibilities of tracking experienced emotions, the problem remains that emotions are very difficult to objectify and thereby to be comparable between different subjects. In addition, in the tomograph, no natural emotional conditions can be produced in general. Nevertheless, the following methods of experimental emotion induction are applied to simulate corresponding situations [HS07]:

- The subject is instructed to feel into a certain emotional state based on materials presented (texts, films, music, etc.).
- In contrast, implicit emotion induction experiments are conducted to elicit behaviour that the subject is not aware of.
- The subject is asked to reminisce about experiences he or she made in the past.
- By giving a task and providing feedback on success or failure, satisfaction or frustration are triggered in the subject.
- Physiological changes are triggered, for example, by drug delivery.

This variety of methods leads to significantly different results and lack of comparability between studies. Additional complexity is generated by the consideration of various emotions, the type of stimulus (e.g. visually assigning or verbally designating an emotion), or the use of emotion discrimination, which is the recognition of emotions usually via facial expressions. The latter can take place either explicitly, i.e. the emotions are named or assigned, or implicitly, in this case the emotions are only perceived, mostly while distracting the subject with other tasks.

In order to illustrate to how diverse, partially contradictory results this abundance of measurement methods can lead, some results of studies are presented that have dealt with the different processing of implicit and explicit emotion discrimination:

- In a study by Critchley et al. [CDP<sup>+</sup>00] it has been shown that the temporal cortex was active in explicit discrimination but the amygdala in implicit discrimination.
- Both Gur et al. [GST<sup>+</sup>02] as well as Habel et al. [HWD<sup>+</sup>07] could show that the amygdala and the hippocampus were activated in explicit but not in implicit tasks.
- Another study by Lange et al. [LWY<sup>+</sup>03], which included only implicit tasks, revealed an involvement of the hippocampus and inferior occipital cortex.
- The amygdala, orbitofrontal cortex, and nucleus caudate were activated in Gorno-Tempini et al. [GTPS<sup>+</sup>01] by explicit emotion discrimination, but only for certain emotions.

In order to handle this diversity of results, Schneider et al. [SGGM94] designed a standardized method for emotion induction: from a reservoir of 40 pictures with male and female actors of different ages, a selection is presented to the subjects with the instruction to feel sad or cheerful using the picture. On these black and white pictures visual expressions with the corresponding emotions are shown. In Figure 3.1 an example of such mood-inducing studies used by Wilhelm et al. [WHM<sup>+</sup>14] can be seen. This or comparable methods are actually used in many studies investigating emotion processing, resulting in a higher comparability of the results of these studies. However, in this way, only a very limited selection of information can be obtained. The only information gathered this way is which brain regions are activated when viewing emotional images and the induced emotion, which is only a very small part of the overall process. In order to be able to examine other aspects of emotion processing, it is necessary to deviate from this standardized method.

Another complexity factor that cannot be reduced with the choice of the correct emotion induction method is the strong individual variability of the brain regions involved in emotion processing. Personality factors such as extraversion or neuroticism can act as moderator variables here [CZD<sup>+</sup>01]. In addition, the brain is in constant transformation, due to learning processes or as a result of injury ('neuroplasticity'). In the course of these transformations, it may happen that a certain task is taken over by another region or that it comes to significant performance changes, which can then lead to a change in the measurement results, although the same subject's brain is investigated.

Many authors also mention gender differences in emotion discrimination. Thayer and Johnson [TJ00] found a general performance difference in their study, in which women showed significantly better results. Other authors observed differences in the brain regions used for various emotions. Women showed, e.g., a high amygdala activity with annoyed and neutral faces, while in men this region was more activated with scared faces [SSS<sup>+</sup>05]. In addition, men showed a stronger asymmetry in the used brain hemisphere [KWAS<sup>+</sup>01]. General behavioural differences could not be observed, however.



Figure 3.1: **Mood-inducing pictures with different facial expressions.** In this example used by Wilhelm et al. [WHM<sup>+</sup>14] composite emotions are used to induce the corresponding mood to the subjects with the instruction to feel the corresponding emotion.

In the following, a brief summary will be given from the abundance of emotion processing investigations, how they are carried out in the case of healthy subjects and which changes are observed in psychiatric diseases.

### 3.2 Emotion processing in healthy people

Smooth emotion processing is a basis for an individual's health in almost all areas of daily life. There is nearly no cognitive task, whose execution is not modulated by emotions. Emotional content, for example, is more easily learned, retained longer, and emotions affect the access to this content [Bow81].

The way from an emotional stimulus, for example listening to a sad song, to (emotional) reaction (maybe the flow of tears) consists, according to Ochsner et al. [OSB14], of the following four steps:

- In the first step, the emotional stimulus is perceived in its environmental context. This can be either an external stimulus like a facial expression of another person or a witnessed event, but also an internal thought. The further processing of the stimulus depends on this context, i.e. if it is generated internally or externally.
- In a second step, attention to some of these stimuli is generated. Not all stimuli are considered to be important and by paying attention to only some of them, a first appraisal of the importance of the stimuli is carried out by lower order areas.
- This selection of stimuli is then appraised on the basis of their significance by higher-order cognitive areas. Aspects of this evaluation are, for example, the relevance for one's goals, the current wishes or needs.
- In the last step, the experience of the stimulus will be modulated according to this evaluation. Emotion-expressive behaviour and autonomic physiologic changes, such as raising the heart rate in fearful situations, are generated.

This emotion processing is not a one-way procedure. Each of the steps mentioned is constantly controlled by higher cognitive areas and, if necessary, adapted.

Although this process is approximately equal for all emotions, so far no region in the brain has been found that is activated for all emotions without exception [PWTL02]. Nevertheless, some brain regions seem to play key roles in the processing of emotions. These regions are briefly presented in the following subsection.

### 3.2.1 Brain areas involved in emotion processing

The regions playing a role in a variety of emotions and tasks consist, according to a multitude of studies [HS07][Dol02], of the amygdala, the anterior cingulate cortex (ACC), the posterior cingulate cortex (PCC), large parts of the prefrontal cortex (PFC), especially the orbitofrontal cortex (OFC) and ventromedial PFC, and parts of the anterior temporal cortex, especially the insula. These areas are highlighted in Figure 3.2. Together, they form a network that extracts emotional and socially relevant information from faces and enables the experience of emotion as well as emotional behaviour [HS07].

For each of the steps of emotion processing described by Ochsner et al. [OSB14], different regions are actively involved. The activation by the stimulus in the first step depends on the context of its generation [RLA<sup>+</sup>97]. Internally generated emotions via thoughts, memories, or the feeling into certain emotional conditions activate the insula [LRB<sup>+</sup>97]. External stimuli rather activate the amygdala, where a presentation below the conscious perception threshold is already sufficient [WRE<sup>+</sup>98]. A special focus of the amygdala is on stimuli of potential danger, so it responds, e.g., particularly strong to anxious faces. In addition, in the perception of externally generated stimuli further areas of the anterior temporal cortex are involved, such as the lateral fusiform gyrus and the superior temporal sulcus, which reacts exclusively to visual stimuli, mainly faces [SSD11].

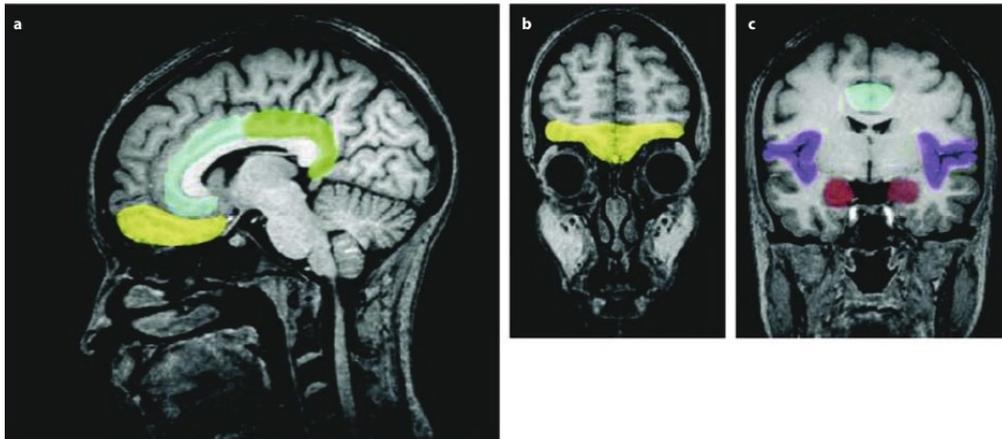


Figure 3.2: **Regions playing a role in emotions, in lateral, anterior and posterior view.** Orbitofrontal cortex (yellow), insula (violett), anterior (blue), and posterior (green) cingulate cortex as well as amygdala (red) [Dol02].

Furthermore, schematic emotional facial expressions, as often used for social feedback, lead to an activation of these regions [WMS<sup>+</sup>02].

The control of attention to certain stimuli in the second step is also modulated by the amygdala [OSB14]. A first assessment is made to determine to what extent this stimulus is relevant to current or lasting goals. In contrast to pure perception, this task is carried out mainly by the part of the amygdala that is located in the left hemisphere. While the right hemisphere part of the amygdala habituates fast to recurring stimuli, the left one is responsible for a more detailed and specific analysis and shows more endurance [SCA08]. In the third step follows a more accurate assessment of the stimulus regarding its context. Historical/autobiographical aspects, the current motivation, and (behavioural) goals play a role in this step. Since this is a higher cognitive function, it is fulfilled almost exclusively by the PFC [OSB14], which is also involved in higher cognitive tasks outside emotion processing.

After an emotional stimulus has undergone the first three steps and was valued as important, an emotional reaction usually follows. This can be expressed in the change of mood and behaviour. In addition, various physiological parameters are modulated according to this reaction, e.g., pulse and breathing can adapt to a situation classified as dangerous. These changes are mainly controlled by the insula [OSB14], which in the brain forms a representation of the sensations of the body and influences the awareness of them.

Although this base network of emotion processing is well documented in the references, most areas in the brain are activated only for certain emotions or tasks and are responsible for the unique shade of each emotion. The most important emotions and their neuronal correlates are presented in the next subsection, followed by a more accurate view of the areas that are activated for cognitive aspects in the field of emotion processing, as required in the analysis of emotional feedback.

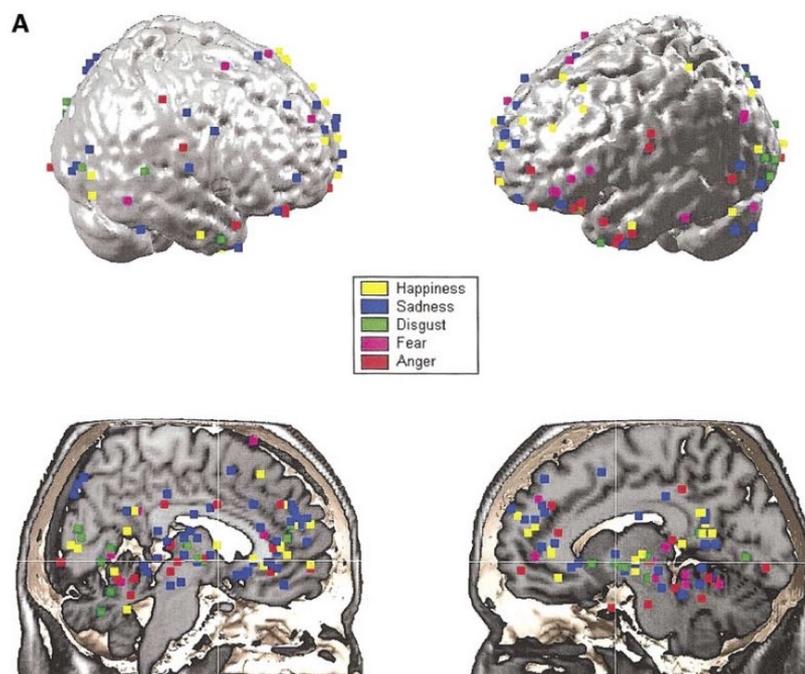


Figure 3.3: Neuronal correlates of the single emotions *happiness*, *sadness*, *disgust*, *fear*, and *anger* in a meta-analysis of Phan et al. [PWTL02]. None of the emotions is represented by a single region, whole networks of brain areas are responsible for the unique shade of each emotion.

### 3.2.2 Dependence on emotional valence

The vast majority of studies that deal with emotion processing have a certain emotion as focus of their investigations or compare two or more emotions with each other. In this context, it is often spoken of the ‘valence’ of emotions, which describes the extent to which an emotion is pleasant or unpleasant [PWTL02]. The emotions fear, sadness, disgust, and anger are usually described as unpleasant, while happiness has a positive valence. For some of these emotions, e.g., disgust, the found involved regions are homogenous across studies, for other emotions, such as happiness, every investigation seems to provide a separate result. In Figure 3.3, the results regarding the neural correlates of the individual emotions of a large-scale meta-analysis of Phan et al. [PWTL02] are presented.

For the processing of sadness-inducing stimuli, the involvement of the amygdala has been very well documented (e.g., [SHK<sup>+</sup>00], [BMF<sup>+</sup>99], [LEJ<sup>+</sup>03]), and activation of the entire ACC, the ventrolateral PFC and the temporal cortex has been shown [HKK<sup>+</sup>05]. A decisive role of the sgACC has also been described particularly often (e.g., [PWTL02], [LEJ<sup>+</sup>03], [LWK<sup>+</sup>12]), Lindquist et al. [LWK<sup>+</sup>12] also specifically mention the posterior part of the ACC. Apparently the neural correlates of sadness consist primarily of the basic network of emotion processing.

The situation looks a little different in the investigation of happiness. Here some brain regions are mentioned that apparently are exclusively involved in this pleasant emotion. However, very different framework conditions and triggers were examined in these investigations. In the happiness studies, reactions to addictive substances, stimulation of the reward system, recall of enjoyable activities, happy faces, and pleasant images are mixed up. The results are accordingly diverse: Habel et al. [HKK<sup>+</sup>05] found involvement of the basal ganglia, amygdala, dorsolateral PFC, inferior temporal cortex, dorsal PCC, and cerebellum. Other authors name the anterior parahippocampal gyrus [EKG<sup>+</sup>03], supramarginal gyrus [PBH<sup>+</sup>98], ventral striatum, and putamen [PWTL02].

The authors of studies describing the emotions fear and disgust are very unanimous. This could be due to the fact that both emotions are relatively primitive, phylogenetically old mechanisms that primarily serve the survival of the individual and hardly have any higher cognitive aspects. Disgust can probably be traced back to a primitive food rejection reflex or a physical aversion to disease-threat [CYR<sup>+</sup>96]. Fear protects the individual from dangers and unnecessary risks; fear-inducing stimuli are therefore evaluated by the amygdala in an accelerated processing step (e.g. [PWTL02], [CYR<sup>+</sup>96], [Agg92]). This finding is also confirmed by lesion studies. Patients with a missing or damaged amygdala showed difficulties in recognizing instances of fear in voices [BMSD04] and correctly assigning fearful faces [CYR<sup>+</sup>96], additionally they did not describe any fear experienced in close contact with snakes, spiders, or when frightened [FADT11]. The amygdala receives help in processing fear-inducing stimuli from the fusiform gyrus and the PFC [SREP98]. The number of brain regions involved in processing of disgust is also very limited. Only the basal ganglia [PWTL02] and the insula ([PBH<sup>+</sup>98], [Cal03]) are mentioned here. In addition to imaging studies, difficulties in recognizing disgust in faces and voices were found in lesion studies in these two regions [ATD03]. In neurodegenerative diseases, such as Parkinson's disease, which can lead to a reduction in nerve tissue in the basal ganglia and insulae, a reduced experience of disgust as a reaction to foul smelling odours has been demonstrated [MHN05].

Little attention has been paid to the emotion anger. The few studies that exist in this area refer to the cingulate and medial temporal areas [PBH<sup>+</sup>98] and the OFC [LWK<sup>+</sup>12]. There exist also counter-tendencies to this small-scale division of emotions, which leads to an abundance of partially competing results. Davidson's valence hypothesis [DH95] assumes that the two hemispheres of the brain are responsible for different groups of emotions each. According to this hypothesis, the left hemisphere processes approach behaviour and positive affects, while the right hemisphere is responsible for withdrawal behaviour and negative affects. Many studies have been able to support these assumptions with their results [DEYH05].

Some authors even go one step further. In their 'psychological constructionist approach' it is assumed that neither the concepts 'emotions', 'perception', or 'cognition' are represented in the brain, nor do special emotions have their own correlates [GB09]. Instead, small networks are activated depending on the actual situation, which then in combination lead to the feeling of the respective emotion or mood. However, since this approach does

not contribute to the understanding of the relationship between emotion processing and social feedback in neurofeedback, in which individual regions are to be trained, this idea is not further explained here.

### 3.2.3 Cognitive tasks related to emotional stimuli

In this section on cognitive aspects of emotion processing, two different fields are highlighted. In the first part, the control of the single steps of emotion processing described in Section 3.2.1. by higher cognitive areas is presented. In the second part, cognitive functions are considered that do not belong to the processing of emotions per se, but can be influenced by the interaction with them, such as memory performance, learning effects, or the assessment of experienced situations.

The steps of emotion processing described so far largely rely on simple stimulus-outcome associations and fast but stereotypical responses. Although this procedure is sufficient for primitive escape behaviour, an additional evaluation of these primary processes is necessary for long-term sensible behaviour in a complex environment. This higher cognitive control takes place among others through the prefrontal cortex, whereby several dimensions of appraisal come into play [DTTC17]:

- Evaluation of the environment based on its implication for one's current goals and needs, depending on their current hierarchy
- The reflection about the intention of others and the consequences of others' actions to one's goals. This aspect involves 'mentalizing', which means the understanding of others' mental states, goals and needs
- Self-evaluation, mostly based on the alignment between one's action and social norms supported by the knowledge about norms and feedback from others
- Appraisal of the costs of certain actions, physiological processes in the body, one's mental state etc.

In order to cope with this complex task, the PFC has developed a functional specialization of the single subregions, each of which is also subject to a different connectivity to other brain regions and thus has access to different information [DTTC17]. These subregions and their respective specializations are described in great detail by Dixon et al. [DTTC17] based on the result of a large-scale meta-analysis. Five of these regions are briefly presented here as examples, which could be involved in processing emotional feedback in neurofeedback training.

The lateral OFC is responsible for evaluating the goal relevance of objects or events that are perceived from the external environment. This includes the assessment of visually presented images, auditory stimuli, smells, tastes, and touches with regard to their usefulness and also danger (survival is a long-term goal of most people, too). This area is particularly sensitive to changes in other people's facial expressions that signal reward

or punishment.

The evaluation of internally generated events, such as remembering past experiences, hoping, or imagining a future event, is taken over by the medial OFC, which is also involved in dreaming. The more pleasant the internally generated event is valued, the higher its activity.

The dorsomedial PFC is responsible for the above-mentioned mentalizing. In contrast to the lateral OFC, its task is to recognize and assess the internal thoughts and intentions of other people that are not directly observable. Thus it is necessary to put oneself in another head or body and to imagine how one would behave with the wealth of experience and the intentions of the other person. A precise knowledge of the characteristics of the other is essential. For this purpose, a further specialization of the dorsomedial PFC takes place: In imaging studies different personality types activate different voxel patterns in this region.

The appraisal of self-related information is located in the rostromedial PFC, this part assigns positive or negative values to the self-image. This assessment is done on the basis of social feedback, subjective feelings, and autobiographical narratives. In studies in which traits should be assessed as self-descriptive, the activity of the rostromedial PFC positively correlated with the degree of conformity of the trait with the self-image. This area seems to play a special role in explicit self-reflection. The self-image created in this way can be modulated by the valenced feedback from others about one's own personality, whereby, interestingly, positive feedback has a stronger influence.

Last but not least, the lateral PFC should be named, which is responsible for evaluating the appraisals and regulatory strategies of the other areas. The participation of this region makes it possible, for example, to value future rewards higher than immediate satisfaction, to make exploratory choices or to create an internal model of a possible but not explicitly named task. Overall, the lateral PFC is responsible for metacognitive awareness, i.e. the ability to reflect, describe, and evaluate mental content.

In natural situations it will almost never happen that only one of these regions is activated. Instead, in the processing of every emotion, there is interplay of subcortical and evaluating cortical areas at every point, which together contribute to the experienced feeling.

In addition to the cognitive evaluation in emotion processing, there are a number of cognitive tasks that are closely related to the experience of emotions. As mentioned before, emotions have a decisive influence on memory performance, but a memory can also trigger an emotion. Many tasks traditionally used in studies examining emotion processing require some degree of cognitive involvement. Examples are focusing on a certain feeling, evaluating one's feelings, evaluating the value of an emotional stimulus, estimating genders in presented images, or recalling emotional memories [PWTL02].

Almost all studies that deal with this link between cognitive and emotional tasks name in this context a participation of the medial PFC and the ACC [PWTL02], where the latter is only activated by the cognitive task and hardly by passive emotional conditions, such as the sole viewing of emotional images. Activation was also found in 60% of recall induction studies for the insula, which is particularly involved in the processing of emotions from

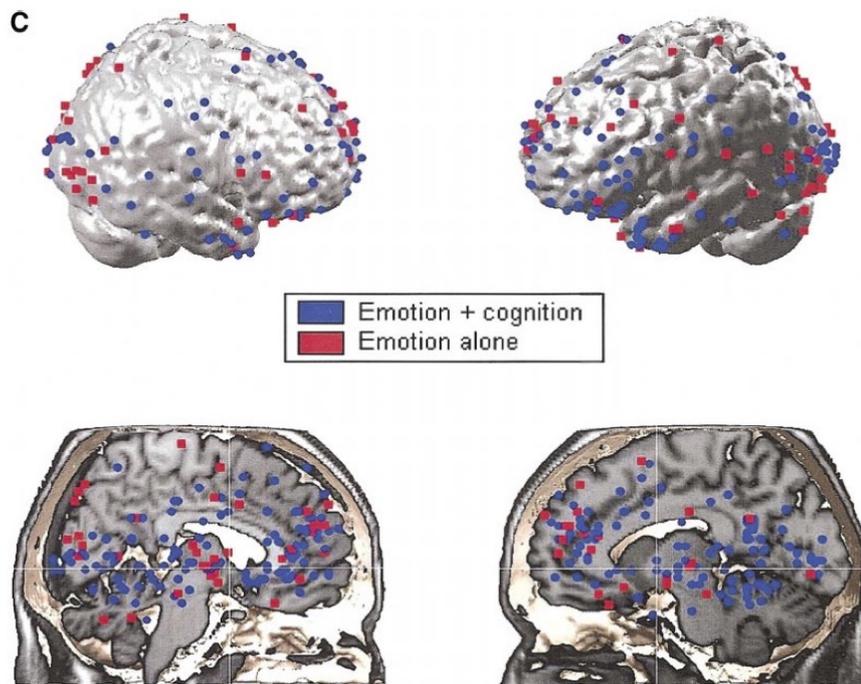


Figure 3.4: **Results of the meta-analysis of Phan et al. [PWTL02] concerning cognitive participation in emotion processing.** Especially PFC, insula and ACC are activated, where the latter is only involved in 'Emotion+cognition', but not in passive emotional tasks, such as sole viewing of emotional images.

internal stimuli, as mentioned before [DGB<sup>+</sup>00]. Figure 3.4 shows the results of the meta-analysis by Phan et al. [PWTL02] with regard to cognitive participation: All three regions show an accumulation of activation across studies.

### 3.3 Emotion processing in affective disorders

Affective impairments are characteristic for many psychiatric and neurological diseases. Emotional abnormalities can be seen in addiction disorders, schizophrenia, affective disorders (bipolar disorder and unipolar depression), neuroses, anxiety disorders, autism, and other personality disorders such as borderline syndrome or sociopathy [HS07]. Depressed patients, for example, seem to be less sensitive in identifying emotional facial expressions; they tend to interpret neutral faces as sad and happy faces as neutral [BDP10] [EZDA11]. There are also changes in the processing speed compared to healthy subjects: negative faces seem to be processed faster and more deeply, while processing of positive faces is slowed down or even completely inhibited [GKYJ04][LRSK07][SDLM<sup>+</sup>04]. These abnormalities in facial recognition lead to an increased suffering of the patients and possibly to a worsening of the symptoms, since they feel social rejection and a restriction of their non-verbal expression [SD05].

There is well-established evidence that there are specific neural substrates that are involved in these dysfunctions. The clinically relevant research to determine these correlates is driven by the hope for a clear diagnosis and long-term therapy [HS07]. Interestingly, the research results on the brain basis of psychiatric disorders are significantly more uniform than those on the processing of emotions in healthy subjects. However, a connection is clearly evident: Almost all regions that are involved in healthy emotion processing show pathological changes in their structure or functionality in affective disorders.

The amygdala shows changes in almost all psychiatric disorders, including the neurological conditions of Alzheimer's disease, Parkinson's disease, and epilepsy [HS07]. Yang and Raine [YR09] found structural and functional changes in the medial OFC that seem to be directly related to the increased aggression in psychopathy and antisocial disorders. Almost all regions that generate emotions from facial expressions are also involved in clinical depression. Mayberg [May97] and Phillips et al. [PLD08][PDRL03] described increased activity in the amygdala and OFC, regions that are responsible for the identification and production of emotions. In contrast, decreased activity was found in the regions that regulate emotions (dorsolateral PFC and ACC). In more recent studies, e.g., by Suslow et al. [SKK<sup>+</sup>10], hyperactivity of all regions involved in emotion processing was reported for negative stimuli; to positive stimuli they reacted with reduced activity. The subgenual part of the ACC (sgACC) seems to play a prominent role in the clinical picture of major depressive disorder and has thus been explicitly investigated in many studies. This region is treated separately in the next section.

### 3.4 The subgenual anterior cingulate cortex in emotion processing

The backward facing subgenual ACC (sgACC) is the lowest part of the ACC arc (marked in red in Figure 3.5). It is involved in the processing of emotions in a variety of ways, but mainly in tasks with a certain cognitive load [LWK<sup>+</sup>12]. Many authors describe the sgACC as a kind of gatekeeper between cognitive and emotional networks, responsible for the transfer of emotional information between regions of the limbic system, e.g., the amygdala, and higher order cognitive structures such as the PFC [SNP<sup>+</sup>20]. Due to this mediator function, it plays a role in emotion reappraisal, positive emotion upregulation, monitoring internal states, sadness induction, social interactions and decision-making, as well as processing of reward and of negative and positive emotional stimuli [GDK<sup>+</sup>15][GKP<sup>+</sup>95][MLB<sup>+</sup>99] [RD08]. In addition, the sgACC modulates the autonomic/neuroendocrine responses and the release of neurotransmitters during the neuronal processing of reward, fear, and stress [DST08]. Lesion studies have shown that patients with a damaged sgACC exhibited an inability to extract information concerning the likelihood of reward or punishment from the social behaviour of others [BDDA94].

Despite its importance in the healthy processing of emotions, the sgACC has mainly been studied because of its crucial role in mood disorders, especially major depressions. In response to emotional stimuli, the sgACC shows increased activity in depressed patients compared to healthy subjects [SSD11], with the activity correlating positively with the



Figure 3.5: **The subgenual part of the ACC.** In this cross-section of a human brain, the sgACC, the lowest, backward-facing part of the ACC arc, is marked in red [DMRB13].

severity of symptoms [OKK<sup>+</sup>00]. This hyperactivity could be determined for both positive and negative stimuli, but spatially separated, so the sgACC seems to have a functional specialization for different valences [GSG<sup>+</sup>05]. Only while in a resting state, as well as in response to neutral stimuli, the measured activity in depressed subjects did decrease [DPNP02][GKP<sup>+</sup>97]. In addition to this increased activity, severe depression leads to an increased cerebral blood flow and glucose metabolism, as well as to a significant reduction in grey matter volume of the sgACC [DPJ<sup>+</sup>97] [DP05]. This volumetric reduction occurs early in the course of the disease and can also be observed in people with a high familial risk [BRD<sup>+</sup>02] [HSS<sup>+</sup>99] [DRB<sup>+</sup>07] [BMCN08].

For some of these pathological changes in the sgACC, good therapeutic success has already been achieved. For example, Mayberg et al. [MBT<sup>+</sup>00] were able to demonstrate a normalization of activity after treatment with antidepressants. Deep-brain stimulation of the region can also lead to an alleviation of the symptoms [HGH<sup>+</sup>11] [RM07]. Only the reduction in volume has so far not been reversed by therapy [DST08]. Due to the therapeutic successes to date, the sgACC is also suitable as target region for neurofeedback-based therapy for major depressive disorder.

### 3.5 Implications for the use of emotional feedback in neurofeedback studies

Although the field of emotion processing is a very complex and sometimes controversial aspect of neuroscience, brain regions can be worked out that are likely to be activated in certain tasks. In the perception and processing of emotional feedback in a neurofeedback context, clearly a cognitive task, the regions that are responsible for the cognitive aspects of emotion processing are likely to be involved. The feedback must be properly classified and interpreted as a reward or punishment. In order to train the subject according to

the desired goal by changing the given feedback, the intentions behind the perceived feedback must be correctly interpreted. It can be assumed that abstract schematic facial expressions also activate the corresponding areas; such a connection has already been shown [WMS<sup>+</sup>02]. If areas of the brain are targets of the training, that are also involved in processing of the feedback, it is not possible to clearly distinguish whether a measured activity is caused by the regulatory strategies of the subject or the interpretation of the feedback. It is therefore possible that both aspects interact and that the measured activation must be reduced by the influence of feedback processing in order to be able to provide more meaningful feedback on the success of the regulatory strategies.

If the neurofeedback training is used as a therapy for neuropsychiatric illnesses, additional attention should be paid to the possibly changed activation caused by the illness.

The sgACC is a useful target region for investigating the influence of feedback on activation in neurofeedback training, as it is involved both in processing tasks with a cognitive load and is clearly present in the clinical picture of major depression, as was shown before.

# Mathematical background

## 4.1 General linear model in fMRI data analysis

In the following, the general linear model (GLM) is presented, which is often used in functional brain research and is also the chosen model in this work. Most of the theoretical background is taken from Moodbrugger's book [Moo11]; the last two subsections refer to Poldrack et al. [PMN11].

### 4.1.1 Basic assumptions

In the GLM it is assumed that one of the observed quantities ('dependent variables') depends on one or more other observable quantities ('predictors', 'regressors', 'independent variables'). The overall goal is to determine this relationship between independent and dependent variables in order to be able to identify the magnitude of the influence of the former on the latter ones. The resulting model can be used to predict the value of the dependent variables if the independent ones are known. This way, the effect of a drug depending on the dosage or behavioural changes as a result of certain therapeutic measures can be estimated, to name just two medical examples.

An important clarification, which should be made right at the beginning, is that the GLM cannot find any causal justification for the expression of the dependant variable, but rather a correlation between dependant and independent variables. To illustrate the principle of correlation, Figure 4.1 shows two-dimensional value pairs with a different correlation. The higher the correlation coefficient  $r$ , the narrower and more aligned to a line the point cloud is and the better the dependant variable can be estimated linearly from the independent one.

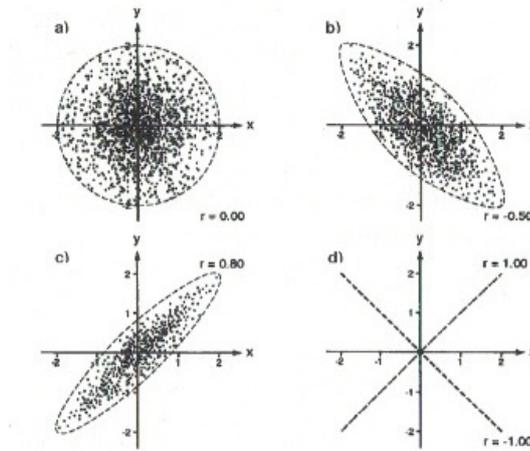


Figure 4.1: **Correlation between two-dimensional value pairs.** The higher the correlation coefficient  $r$ , the narrower and more aligned to a line the point cloud is. With a higher correlation the linear predictability of dependant values of the independent values is higher [Moo11].

The mathematical approach in the GLM for determining the relationship between predictors and dependent variables  $y$  consists of breaking down the latter into a linear combination of predictors  $x$  and an error term  $\varepsilon$ :

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \varepsilon_i \quad (4.1)$$

The values of measurement  $i$ , e.g. various properties of the  $i$ -th subject, are connected by a weighted sum. These weights are denoted as  $\beta_j$ . To each of the  $m$  predictors a  $\beta$  weight is assigned.  $\beta_0$  is the regression constant; it describes the average value of  $y_i$  that is independent of the predictor variables. The variation in  $y_i$  that cannot be explained by the regression constant or the weighted predictors is described by the error term  $\varepsilon_i$ . This term includes both systematic errors that have not been taken into account, such as the variability of the measuring devices, as well as random fluctuations. The value  $\hat{y}_i$ , which will be predicted by the model as an estimation of  $y$ , can thus be described by the following relationship:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} \quad (4.2)$$

This gives the error term as:

$$\varepsilon_i = y_i - \hat{y}_i \quad (4.3)$$

#### 4.1.2 Regression analysis

The linear  $\beta$  weights are determined to best represent the data und resulting in the smallest possible model error.

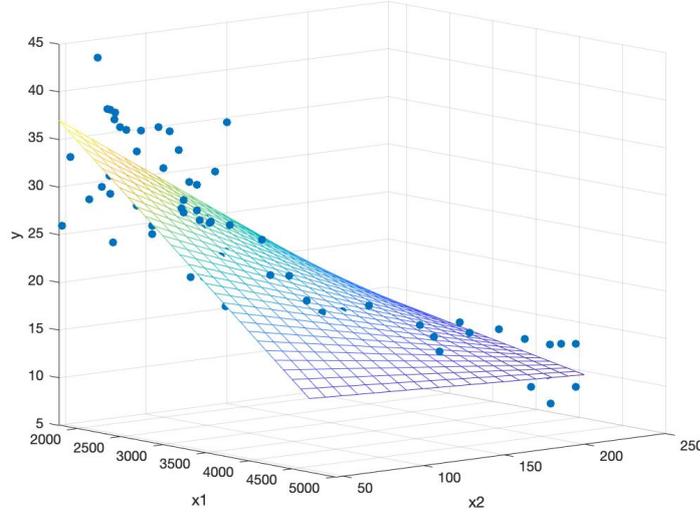


Figure 4.2: **Example of a multidimensional regression.** The 3-dimensional data is modelled by a 2-dimensional plane.

The data is located in an  $(m + 1)$ -dimensional space ( $m$  regressors and 1 dependent variable) and is modelled by an  $m$ -dimensional hyperplane, which is characterized by the  $\beta$  weights (Figure 4.2). Usually, the model error for each subject  $i$  results from the distance between the  $i$ -th data point and the hyperplane parallel to the axis of the dependent variable (Figure 4.3). The GLM allows the regression to be determined for  $i$  measurements simultaneously. In the GLM it is assumed that the predictors and the dependent variables are subject-specific, but the  $\beta$  values are the same for all measurements. Their true population values are representing the general linear relationship between the measured values and are therefore neither dependent on the specific characteristics of the single measurements nor the time point of measurement. The dependent variable  $y$  of all measurements can be summarized in the vector  $\vec{y}$ , the predictors of the individual measurements in the predictor vectors  $\vec{x}_1, \dots, \vec{x}_m$ . The individual error values, also called 'residuals', are combined to form the error vector  $\vec{\varepsilon}$ . This results in the regression model

$$\begin{aligned}
 \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} &= \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix} + \dots + \beta_m \begin{pmatrix} x_{m1} \\ \vdots \\ x_{mn} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \\
 &= \begin{pmatrix} 1 & \dots & x_{m1} \\ \vdots & \ddots & \vdots \\ 1 & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}
 \end{aligned} \tag{4.4}$$

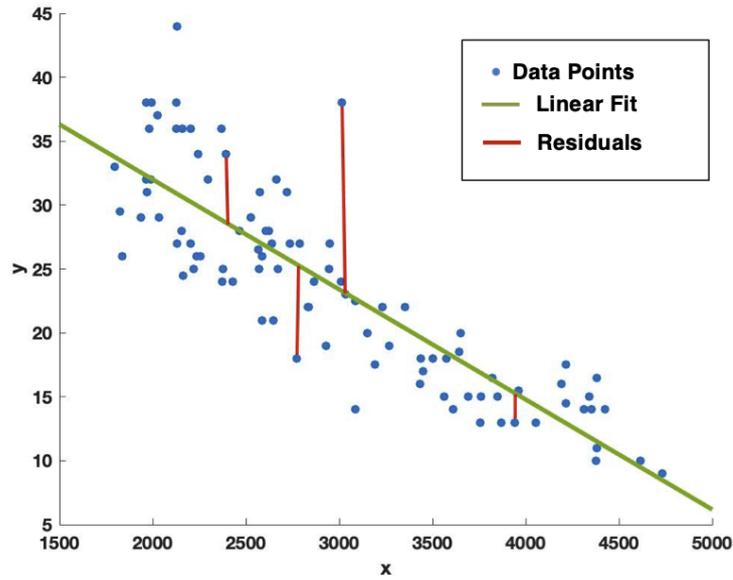


Figure 4.3: **Model errors called 'residuals' for a linear regression in two dimensions.** The error is the distance between the linear fit and the measured data point.

The linear combination of the predictor vectors can be formulated as a matrix multiplication between the predictor matrix  $\mathbf{X}$  and the weight vector  $\vec{\beta}$ . The columns of  $\mathbf{X}$  consist of the predictor vectors, the weight vector  $\vec{\beta}$  of the single weights  $\beta_j$ .

The  $\beta$  weights should now be determined in a way that the error  $\varepsilon$  is minimized. One way is to use the sum of squared errors (*SSE*): Since it should not matter in this model in which direction the deviations take place and to take into account the totality of the errors in all measurements, the residuals are squared and then added up. The *SSE* can be calculated as follows:

$$\sum_{i=1}^n \varepsilon_i^2 = \vec{\varepsilon}^T \vec{\varepsilon} = (\vec{y} - \mathbf{X}\vec{\beta})^T (\vec{y} - \mathbf{X}\vec{\beta}) = \vec{y}^T \vec{y} - 2\vec{\beta}^T \mathbf{X}^T \vec{y} + \vec{\beta}^T \mathbf{X}^T \mathbf{X} \vec{\beta} \quad (4.5)$$

The *SSE* thus has a quadratic dependence on the weight vector  $\vec{\beta}$ . The resulting parabolic shape must therefore always have a single extreme point. In this case, it is clearly a minimum, since error values can naturally continue to grow the more unsuitable the model is.

This minimum can be determined by taking the partial derivatives of the *SSE* after all  $\beta_i$  and determining their zero-argument:

$$\frac{\partial SSE}{\partial \vec{\beta}} = -2\mathbf{X}^T \vec{y} + 2\mathbf{X}^T \mathbf{X} \vec{\beta} = 0 \quad (4.6)$$

In case that  $\mathbf{X}^T \mathbf{X}$  is not singular, i.e. that no column is a linear combination of another, the estimation of  $\vec{\beta}$ ,  $\vec{\hat{\beta}}$  can be calculated analytically:

$$\vec{\hat{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y} \quad (4.7)$$

Otherwise,  $\vec{\hat{\beta}}$  is determined using approximate numerical methods. The computational ease of this ordinary least-squares approach makes it the prevailing method for linear model fitting. However, the quadratic dependence on the error terms results in increased susceptibility to outliers compared to linear forms.

### 4.1.3 Confidence interval and significance of the estimates

The data of the sample are distributed with a certain probability around the ‘true’ value, which is to be determined with the help of the model. By using the least-squares method, this distribution of the data is given by a normal distribution around the estimated population value  $\vec{y}$  with the standard deviation  $s_e$  of the error distribution of the sample:

$$s_e = \sqrt{s_e^2} = \sqrt{\frac{\vec{\varepsilon}^T \vec{\varepsilon}}{n}} = \sqrt{\frac{\vec{y}^T \vec{y} - \vec{\hat{\beta}}^T \mathbf{X}^T \vec{y}}{n}} \quad (4.8)$$

The distribution is depicted in Figure 4.4. This results in the standard estimation error for the prognosis of a new subject/data point, which is an unbiased estimator for the standard deviation of the error in the entire population  $\sigma_\varepsilon$ :

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{n \cdot s_e^2}{df_e}} \quad (4.9)$$

The number of degrees of freedom of the error  $df_e$  results from the sample size minus the number of auxiliary quantities that are used to calculate the sum of squares, in this case the number of predictors:

$$df_e = n - m \quad (4.10)$$

As can be seen from the probability distribution of the data, there is also only a certain probability of assuming a specific value for the modelled  $\hat{y}$ , or its determining parameters  $\hat{\beta}_j$ . Therefore it is common to specify a confidence interval for the  $\beta$ -estimate. Under the assumption that a given experiment would be repeated infinitely often,  $1 - \alpha$  percent of the confidence intervals cover the ‘true’  $\beta$ -value.  $\alpha$  is referred to as the level of significance. The true value  $\beta_j$  is covered by the interval

$$[\hat{\beta}_j - t_{\frac{\alpha}{2}; df_e} \cdot \hat{\sigma}_{\beta_j}, \hat{\beta}_j + t_{\frac{\alpha}{2}; df_e} \cdot \hat{\sigma}_{\beta_j}] \quad (4.11)$$

with a probability of  $1 - \alpha$  (in most cases  $\alpha$  is set to be 0.05). The term  $\sigma_{\beta_j}$  denotes the (unknown) standard error of the parameter  $\beta_j$ , which can be estimated using  $\hat{\sigma}_\varepsilon$  and  $a_{jj}$ , the  $j$ -th diagonal element of  $(\mathbf{X}^T \mathbf{X})^{-1}$ :

$$\hat{\sigma}_{\beta_j} = \frac{\hat{\sigma}_\varepsilon}{\sqrt{a_{jj}}} \quad (4.12)$$

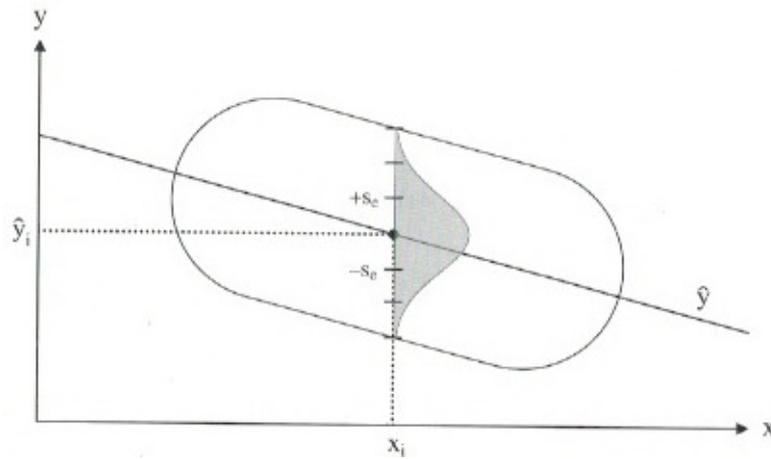


Figure 4.4: **Error distribution of the sample.** By using the least-square method, the modelled value  $\hat{y}$  is estimated in such a way, that the distribution of the residuals follows a normal distribution with standard deviation  $s_e$  [Moo11].

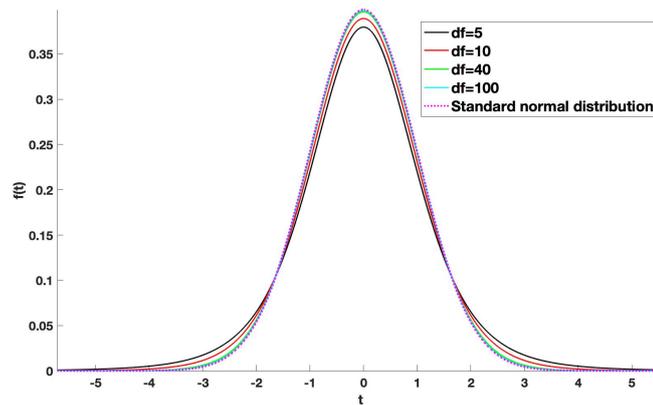


Figure 4.5: **T-distributions for different degrees of freedom.** The higher the number of degrees of freedom, the more the t-distribution approaches the standard normal distribution.

The parameter  $t_{\frac{\alpha}{2};df_e}$  in the confidence interval is the critical value  $t_{crit.}$  of a t-distribution with  $df_e$  degrees of freedom. The t-distribution is shown in Figure 4.5 the only parameter of this distribution is the number of degrees of freedom  $df$ . For large  $df$  the t-distribution approaches the standard normal distribution. The critical t-value describes the value at which the probability that  $t < t_{crit.}$  equals  $1 - \alpha$  (compare Figure 4.6).

Another crucial task is to determine the significance of the found  $\beta$ -estimates. In this context, a significant result means that the sample results allow conclusions to be drawn about the entire population. For this purpose, hypothesis tests are carried out in inferential statistics. To enable these hypothesis tests in regression, a sample must meet a number of requirements:

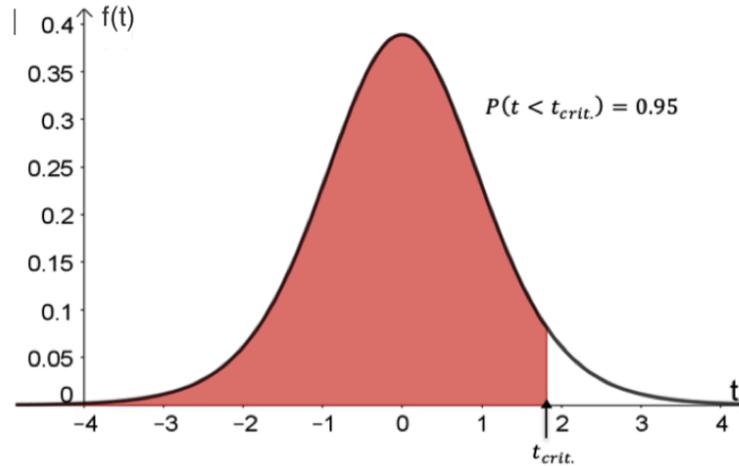


Figure 4.6: **T-distribution with critical t-value.** At the value  $t_{crit.}$ , the probability that  $t < t_{crit.}$  equals  $1 - \alpha$ , here depicted for  $\alpha = 0.05$  (filled area).

- The sample must be chosen randomly from the population
- $n$  must be much larger than  $m + 1$
- The predictor matrix  $\mathbf{X}$  must not have any linear dependencies, since  $\mathbf{X}^T \mathbf{X}$  becomes singular otherwise
- All measured predictor variables must be assumed to be error-free
- The residuals are normally distributed around zero, homoscedastic and uncorrelated

To carry out a hypothesis test, two hypotheses are compared to each other. The null hypothesis  $H_0$  formulates restrictions for an influencing parameter  $\hat{\beta}_j$ , such as  $\hat{\beta}_j = 0$ . In contrast, the alternative hypothesis  $H_1$  assumes that the value is different from what is claimed in the  $H_0$ , e.g.  $\hat{\beta}_j \neq 0$ . The aim of a hypothesis test is now to determine whether the sample data are compatible to the assertions of  $H_0$ , then  $H_0$  is retained, or whether they are incompatible, which is the case when the occurrence of the restrictions of the  $H_0$  for the given sample is very unlikely. In this case the  $H_0$  is rejected and the alternative hypothesis is accepted. The rejection of the null hypothesis cannot, however, be seen as a deductive proof that it is wrong. Furthermore,  $H_0$  is not accepted, if it cannot be rejected, i.e., if a parameter cannot be shown to be significantly different from 0, this does not imply it is 0.

The GLM offers a very flexible variant of formulating assumptions about the true characteristics of all  $\beta$ -estimates. The  $H_0$  is here defined as follows:

$$H_0 : \mathbf{L}\vec{\hat{\beta}} = \vec{\delta} \quad (4.13)$$

$\mathbf{L}$  is called the 'contrast matrix' and can take different forms, depending on which relations between the parameters are to be tested. If  $\mathbf{L}$  is chosen as the identity matrix  $\mathbf{1}$ , each parameter  $\hat{\beta}_j$  is tested individually for its equivalence to the assumed value  $\delta_j$ . Moreover, the parameters can be compared with one another. To test the difference of for example two parameters for equality with  $\delta_j$ ,  $\mathbf{L}$  should be defined as follows:

$$\mathbf{L} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad (4.14)$$

The so-called 'global null hypothesis', which will be used to test the significance of the  $\beta$ -estimates, states that all parameters  $\hat{\beta}_j$  have the value 0, i.e. that there is no parametric relationship between the predictors and the dependant variable. There are several possibilities to test this hypothesis. In the simplest case, the previously determined confidence interval can be used. If this confidence interval includes the value 0, the corresponding parameter estimate is not significantly different from 0.

Alternatively, the general linear hypothesis can be determined with the values

$$\mathbf{L} = \mathbf{1}, \vec{\delta} = \vec{0} \quad (4.15)$$

and tested by calculating the so-called 'p-values'.

The p-value for the null hypothesis

$$H_0 : c_j \hat{\beta}_j = 0, \forall j, \quad (4.16)$$

with  $c_j$  as the  $j$ -th row of  $\mathbf{L}$ , is defined as the probability  $P$  of a random variable  $T_{df_e}$ , following a t-distribution with  $df_e$  degrees of freedom, to be smaller than the absolute value of a critical value  $t$ :

$$p(\hat{\beta}_j) = P(T_{df_e} < |t_j|) \quad (4.17)$$

This critical value  $t$  is in this case calculated as

$$t_j = \frac{c_j \hat{\beta}_j}{\sqrt{c_j (\mathbf{X}^T \mathbf{X})^{-1} c_j^T \hat{\sigma}_\varepsilon}} \quad (4.18)$$

The null hypothesis is rejected, if  $p < \alpha$ , since the probability to reject it mistakenly is smaller than  $\alpha$  in this case, and the  $\beta$ -estimate is considered significant

#### 4.1.4 Transfer to fMRI settings

In principle, fMRI-measured values are modelled according to the methods described so far. In most cases, the entire volume is not modelled at once, but voxel-per-voxel; this method is called 'mass univariate data analysis'.

The time series of the measured activities here functions as the dependent variable  $\vec{y}$ . An exemplary time series is shown in Figure 4.7 (blue line). The predictor matrix  $\mathbf{X}$  is usually called the 'design matrix' and is made up of different types of regressors. The first regressor  $\beta_0$  describes the task-independent baseline activation. In addition, one

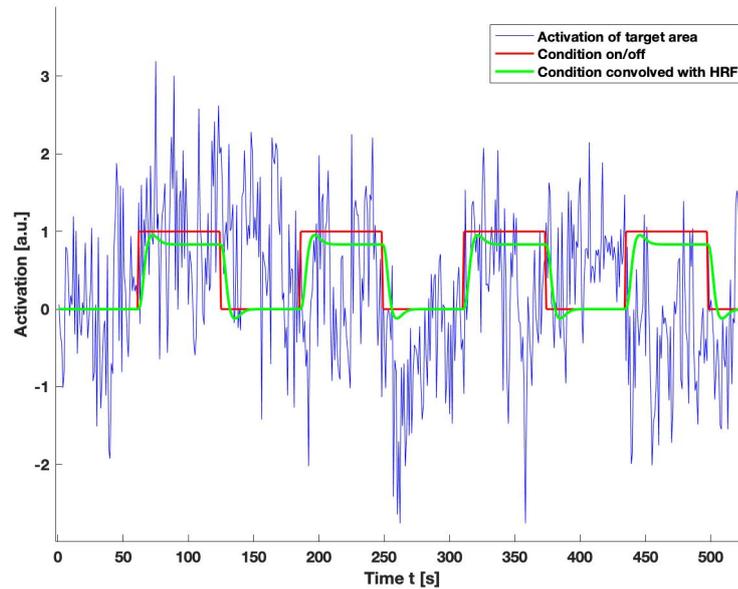


Figure 4.7: **Exemplary time series of the activation of a voxel measured with fMRI and the condition regressor.** The activation will be integrated into the GLM as dependant variable  $y$ , while the condition (convolved with the HRF) will be inserted as independent variable.

or more regressors are inserted that reflect the conditions of the measurement, i.e. the experimental design. These regressors initially only consist of the information whether a certain condition, e.g. a specific task, is currently performed (1) or not (0) at the measured time point (Figure 4.7, red line). Since several tasks can alternate or be performed at the same time, a separate regressor is inserted into the design matrix for every possible condition that can be either active or inactive. This way it can be determined whether a certain condition leads to a significant change in the activation in a voxel. Before model fitting, all these regressors are convolved with the HRF introduced in Chapter 2.2.1.2 in order to take into account the delay caused by the physiological nature of the measurement (Figure 4.7, green line).

Apart from these ‘interesting’ regressors, further regressors of ‘no interest’ can be inserted, which can also have an influence on the level of activation. These include, for example, head movement parameters, breathing or heartbeat, or technical artefacts such as magnetic drifts in the scanner that alter the fMRI signal.

## 4.2 Linear mixed-effects model

In the previous section, the modelling of the measurement data from a single run was described. However, generalizable statements for all runs of all subjects are desired. Therefore, a method is needed that combines these individual results into an overall

outcome, taking into account the variation of the data. A linear mixed-effects model was chosen for this purpose, which is outlined below. The description mainly follows the work of Gelman and Hill [GH07].

### 4.2.1 Basic principles and motivation

Comparing measurements that investigate the same relationships between dependent and independent variables, but originate from different core groups, incorrect conclusions can be drawn if this grouping is not taken into account. A fictitious example is shown in Figure 4.8a. The dependant variable appears to be negatively correlated with the independent variable. However, the data were collected from three different groups (see Figure 4.8b), in each of which there is a positive correlation between the variables, but which show strong group-dependant differences, so that a common consideration of all data (so-called ‘complete pooling’) leads to an opposite result. Although the error caused by neglecting the group membership does not always have to be as strong as in the example shown, it illustrates the problem quite well. The problem arises mainly from the fact that measured values from one group are more similar to each other than to values from different groups. As a result, the residuals are no longer independent and a simple regression cannot explain the complete variance in the data.

The total variance  $\sigma_{tot}^2$  can be divided into several effects: the variance  $\sigma_{w_i}^2$  within a group  $i$ , which describes the distribution around the group mean, and the variance  $\sigma_b^2$  between the groups, which describes the different levels of the groups. The division into different substructures with varying mean values results in a hierarchical structure in the data. There can be several groupings in a data set, which are formed according to different criteria; the groups can also consist of subgroups and their distribution can depend on additional parameters.

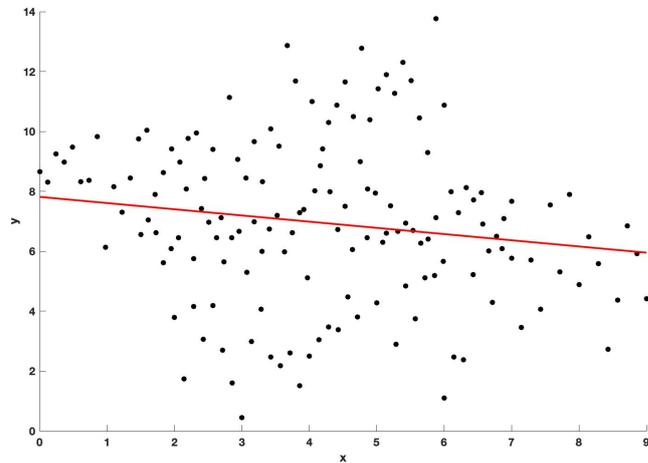
In order to test whether there is a hierarchical structure at all, the proportion of the variance between the groups in the total variance can be determined. For this purpose the intra-class correlation (*ICC*) is calculated:

$$ICC = \frac{\sigma_b^2}{\sigma_{tot}^2} = \frac{\sigma_b^2}{\sigma_b^2 + \sum_{i=1}^n \sigma_{w_i}^2} \quad (4.19)$$

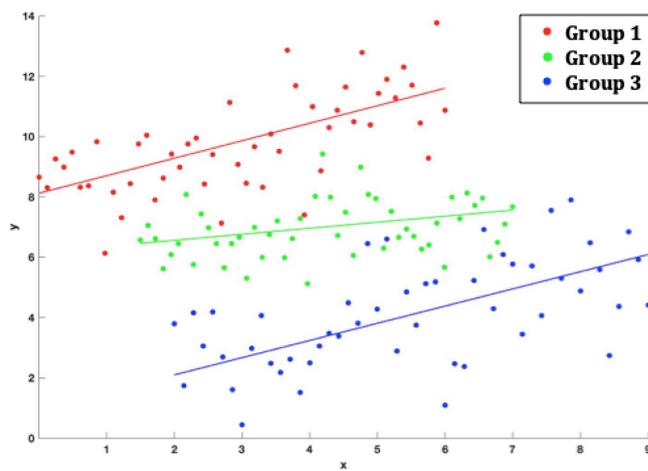
If the *ICC* is high, it can be assumed that a hierarchical structure is present that should be considered in the model.

One possibility to deal with the group structure would be to model all groups individually (‘no pooling’). This way, only a fraction of the total data could be used for the regression, which would lead to a severe restriction of the degrees of freedom and to a smaller ratio between the number of regressors and the number of measurement data. The validity of the found results would also be limited, as they would primarily relate to the group from which they were determined.

The linear mixed-effects model offers a good compromise, which allows all data to be used, but still keep the information that results from the group membership. The more detailed implementation of this model will be explained below.



a



b

Figure 4.8: **Complete pooling versus no pooling.** If the data originate from different groups, the pooling of all data can lead to a changed correlation between independent and dependant variable due to group-dependant differences in the data.

### 4.2.2 Definition of fixed effects and random effects

The mixed-effects model owes its name to the fact that it combines different effects: fixed effects and random effects. There are various definitions for these mentioned effects. A definition that intuitively conveys how these effects can be determined can be found in Searle et al. [SCM06]:

“*Fixed effects* are the effects attributable to a finite set of levels of a factor that occur in the data and which are there because we are interested in them. [...] The second kind of effects are *random effects*. These are attributable to an (usually) infinite set of levels of a factor, of which only a random sample are deemed to occur in the data.”

Put simply, fixed effects are the variables whose actual expression is of interest during the measurement, since the relationship between the expression and its potential impact on the expression of the dependant variable is to be determined. A medical example would be the dose of medication that results in a particular symptom relief.

Random effects, on the other hand, are variables that can have an influence on the expression of the dependant variable, but whose own expression is of no further interest, since they are only randomly taken from a population that is only of interest as a whole. An example would be the single subject chosen at random to represent the population. Although this subject may produce a different result than another would have done, this subjective expression is only coincidental and is not seen as decisive for determining the relationship between the dependant variable and the fixed effect.

### 4.2.3 Fixed effects and random effects in a mixed model

Considering only the fixed effects, the already known regression equation for the  $i$ -th measured value is obtained (exemplary for two dimensions), which is here also part of the  $j$ -th group:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij} \quad (4.20)$$

As before, the error term  $\varepsilon_{ij}$  follows a normal distribution around zero:

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad (4.21)$$

The regression weights  $\beta_0$  and  $\beta_1$  are the fixed intercept and slope, which represent the correlation between the fixed effect  $x_{ij}$  and the dependent variable  $y_{ij}$ .

As can be seen from Figure 4.8, it is possible that different groups are at different levels of the dependent variable  $y$ . So each group has an individual intercept. Therefore it makes sense to add an additional intercept  $u_{0j}$  to the regression equation, which is dependent on group  $j$  and is called ‘random intercept’:

$$y_{ij} = \beta_0 + u_{0j} + \beta_1 x_{ij} + \varepsilon_{ij} \quad (4.22)$$

This additional intercept can also be viewed as a second error term, which causes additional variance in the data, depending on the group. It also follows a normal distribution with 0

mean and variance  $\sigma_{u_0}$ , which indicates the variability between (all possible) groups and is estimated by generating a sample with random size and random selection of groups:

$$u_{0j} \sim N(0, \sigma_{u_0}^2) \quad (4.23)$$

The distribution around 0 describes the fact that all measured values have the fixed intercept  $\beta_0$  on average, but a certain amount must be added or subtracted for each group.

In addition to the intercept, there may also be a group-related dependency on the regressor  $x$ . For example, a member of a group, who already has a high value of  $y$ , no longer reacts as strongly to an increase in  $x$  as a member from another group. To take this factor into account, a ‘random slope’  $u_{1j}$  can also be added, i.e. a group-dependent parameter that describes the dependency on the regressor:

$$y_{ij} = \beta_0 + u_{0j} + (\beta_1 + u_{1j})x_{ij} + \varepsilon_{ij} \quad (4.24)$$

Analogous to the random intercept, the random slope follows the distribution:

$$u_{1j} \sim N(0, \sigma_{u_1}^2) \quad (4.25)$$

The described extensions can be applied to any number of dimensions by inserting a random slope  $u_{\{1, \dots, m\}j}$  for each of the  $m$  regressors  $x_{\{1, \dots, m\}ij}$ . Such random parameters can be inserted for any number of (independent) groupings; for example, a data set can be grouped into different classes of the subjects as well as into different measurement days, assuming that on each day subjects of different classes are randomly chosen for the measurement and the specific day is of no further interest. In this case, the random parameters  $u_{\{0, \dots, m\}j}$  would be inserted for the classes and  $v_{\{0, \dots, m\}j}$  for the days.

The group parameters can also be divided into further subgroups if necessary, for each of these subgroups a separate random parameter would be estimated.

Due to the complex nesting of the parameters, it is no longer possible to calculate them using an ordinary least square method. Instead, the maximum likelihood (ML) method is used.

In this method, the (iterative) search is made for the parameters that have the highest likelihood of having produced the measured values. According to Sir Ronald Fisher [Fis12], the likelihood  $L$  is defined as the probability of the occurrence of the data as a function of the model parameters, assuming that the model applies:

$$\begin{aligned} L(\text{Parameters}|\text{Sample}) &= P(\text{Sample}|\text{Parameters}) \\ &= \prod_{i=1}^n P(\text{Measurement } i|\text{Parameters}) \end{aligned} \quad (4.26)$$

The likelihood is calculated for every possible combination of parameters and the parameter set with the highest likelihood is selected.

It can be shown [VV05] that the ML method underestimates the value of the variance compared to the true variance of the  $\beta$ -estimate due to the fact that the already estimated mean value for  $\beta$  is used to estimate the variance. This problem can be solved by using the restricted maximum likelihood (ReML) instead of ML. Here, after the mean has been estimated, the ML function is integrated over it in order to remove its information from the likelihood equation. This adjusted function is then maximized to obtain the now unbiased estimate of the variance.

#### 4.2.4 Evaluation of the obtained information

In the linear mixed-effects model there is the possibility to model the data in different levels of complexity and hierarchical levels. However, not all of these potentially possible additional parameters are advantageous for the model. The factual context should be described as precisely as possible, but too many parameters can lead to overfitting, in which the model adapts too much to the specific sample and hardly generalizes. Different models with different numbers of parameters can be compared with the Akaike information criterion ( $AIC$ ) with regard to their information content in relation to the number of parameters:

$$AIC(i) = -2 \cdot \log(L_i) + 2 \cdot K_i \quad (4.27)$$

The  $AIC$  of the  $i$ -th model consists of the sum of the logarithmic likelihood  $L_i$  of the model, that was maximized in advance, and a penalty term depending on the number of parameters  $K_i$ . The larger the ML, the closer the first term approaches zero. Since a higher ML usually goes hand in hand with a higher number of parameters and these can easily lead to overfitting, the second term penalizes a higher  $K_i$  by increasing the value of the  $AIC$ . Accordingly, the number of parameters must be selected for a model that minimizes the  $AIC$ .

#### 4.2.5 Variance and Covariance of the random effects

As already mentioned, the random coefficients  $u_{\{0,\dots,m\}j}$  follow a multivariate normal distribution with the following covariance matrix:

$$u_{\{0,\dots,m\}j} \sim N \left( 0, \begin{pmatrix} \sigma_{u_0}^2 & \rho(u_1, u_0)\sigma_{u_1}\sigma_{u_0} & \cdots & \rho(u_m, u_0)\sigma_{u_m}\sigma_{u_0} \\ \rho(u_0, u_1)\sigma_{u_0}\sigma_{u_1} & \sigma_{u_1}^2 & \cdots & \rho(u_m, u_1)\sigma_{u_m}\sigma_{u_1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho(u_0, u_m)\sigma_{u_0}\sigma_{u_m} & \rho(u_1, u_m)\sigma_{u_1}\sigma_{u_m} & \cdots & \sigma_{u_m}^2 \end{pmatrix} \right) \quad (4.28)$$

The entries in this covariance matrix provide information about the distribution of the groups (and no longer about the distribution of the individual measured values). On the diagonal are the variances of the individual coefficients. They provide information about how much the groups differ with regard to this coefficient. A high variance thus indicates a strong difference between the groups and supports the need for a mixed-effects model. The entries off the diagonal reflect the covariance between two coefficients. This gives

insight on how strongly the individual coefficients correlate with each other, i.e. whether, for example, the response to a regressor depends on the intercept of the group.

In this way, not only information about the fixed effects can be obtained, but statements can also be made about the distributions and dependencies of group affiliations, which can prove useful in a factual context.

### 4.3 Cross-validation

In order to test the generalization capability of the model determined from the training data, the error in the prediction of a second data set, the test data set, is usually estimated. In many cases, however, the amount of available data is not large enough to be able to split it into two different data sets. In this case, cross-validation can be used, where any number of test sets can be formed from the existing data through subdivisions. The procedure of a  $k$ -fold cross-validation consists of the steps described below, following the notation of Hastie et al. [HTF17].

The data set is divided into  $k$  equally sized parts. Then one of these parts is left out as a test set, while the model is determined from the remaining parts, which is then tested on the small test set. The error of the model in predicting the data of the  $i$ -th test set is determined as the loss  $\mathcal{L}(y_i, \hat{y}^{-T_i}(x_i))$ . Here,  $\hat{y}^{-T_i}$  denotes the model function that was determined without the  $i$ -th test set,  $y_i$  and  $x_i$  the dependant and independent variables of the  $i$ -th test set.

This procedure is repeated for all  $k$  test sets, so that the mean prediction error of the model results as:

$$err_p = \frac{1}{n} \sum_{i=1}^k \mathcal{L}(y_i, \hat{y}^{-T_i}(x_i)) \quad (4.29)$$

The larger this mean error, the smaller is the ability of the model to generalize to an overall population. Similar to the *AIC*, cross-validation can be used for model selection by choosing the model with the smallest mean prediction error.

The calculation of the loss  $\mathcal{L}$  depends on the type of model. For a regression model, this can be determined using the correlation between the measured data  $y_i$  and the data  $\hat{y}(x_i)$  predicted by the model. The higher the correlation, the smaller the loss:

$$\mathcal{L}(y_i, \hat{y}^{-T_i}(x_i)) = 1 - \rho_c(y_i, \hat{y}^{-T_i}(x_i)) \quad (4.30)$$

Instead of the Pearson correlation coefficient  $\rho$ , the concordance correlation coefficient  $\rho_c$  is used in this work due to its ability to compare two measures of the same variable and to indicate how far the correlation between the two models deviates from another [Lin89]. Compared to the Pearson coefficient, concordance correlation is not invariant with respect to scaling and offset. This way not only the relationship of the true and predicted values, but also the values themselves can be compared.

The degree of agreement (concordance) between the two measures can be characterized by the expected value of their squared difference:

$$E \left[ (Y - \hat{Y})^2 \right] = (\mu_Y - \mu_{\hat{Y}})^2 + \sigma_Y^2 + \sigma_{\hat{Y}}^2 - 2\rho(y, \hat{y})\sigma_y\sigma_{\hat{y}} \quad (4.31)$$

Here, the variables  $\mu_Y, \mu_{\hat{Y}}$  denote the respective means,  $\sigma_Y^2, \sigma_{\hat{Y}}^2$  the respective variances and  $\rho(y, \hat{y})$  the Pearson correlation coefficient between the two models. If the two measures would be in exact agreement, i.e. if both means and variances were the same and  $\rho = 1$ , the above expectation would be 0. To obtain only concordance values between 1 and -1, this expectation is related to the completely uncorrelated case and subtracted from 1:

$$\begin{aligned} \rho_c &= 1 - \frac{E[(Y - \hat{Y})^2]}{E[(Y - \hat{Y})^2 | \rho = 0]} = 1 - \frac{(\mu_Y - \mu_{\hat{Y}})^2 + \sigma_Y^2 + \sigma_{\hat{Y}}^2 - 2\rho(y, \hat{y})\sigma_y\sigma_{\hat{y}}}{(\mu_Y\mu_{\hat{Y}})^2 + \sigma_Y^2 + \sigma_{\hat{Y}}^2} \\ &= \frac{2\rho(y, \hat{y})\sigma_Y\sigma_{\hat{Y}}}{(\mu_Y\mu_{\hat{Y}})^2 + \sigma_Y^2 + \sigma_{\hat{Y}}^2} \end{aligned} \quad (4.32)$$

Empirically,  $\rho_c$  can be estimated from a sample of  $N$  data points as follows:

$$\hat{\rho}_c = \frac{\frac{2}{N} \sum_{n=1}^N (y_n - \bar{y})(\hat{y}_n - \bar{\hat{y}})}{\frac{1}{N} (\sum_{n=1}^N (y_n - \bar{y})^2 + \sum_{n=1}^N (\hat{y}_n - \bar{\hat{y}})^2) + (\bar{y} - \bar{\hat{y}})^2} \quad (4.33)$$

with

$$\bar{y} = \frac{1}{N} \sum_{n=1}^N y_n, \quad \bar{\hat{y}} = \frac{1}{N} \sum_{n=1}^N \hat{y}_n \quad (4.34)$$

#### 4.4 Test-retest reliability

An important factor used to describe the quality of both a model and an experimental design is the so-called ‘test-retest reliability’. It indicates the ability of a measurement method to give consistent results under the same conditions. So, when performing an experiment multiple times in the same way, it is a requirement for good interpretability of the results that each subsequent experiment produces results similar to the first one. The probability that a certain design (for fMRI NF, e.g., including the selection of the target area, tasks, length of the measurement, number of runs, technical parameters) yields such reliable data can be estimated using various statistical methods [CHZ<sup>+</sup>09]. One measure that is often used to estimate test-retest reliability is the intra-class correlation coefficient (ICC), which was introduced in Section 4.2.1. in the context of hierarchical models. Since the ICC describes the proportion of the between-subject variability in the total variability, it is particularly high if the total variance mainly consists of fluctuations between the single subjects and at the same time the variance within the individuals is small. The ICC can take values between 0 and 1, where a value of 1 describes perfect reliability, since there is no within-subject variability at all, i.e. every newly measured value of a test person is exactly the same as all previous ones. An ICC of around 0, on the other hand, can occur, if the within-subject variability is very high, which would point towards a low reliability. However, if there is close to no variation between the subjects, ICC values will also be low. Hence, the ICC is actually measuring distinguishability rather than reliability, which often are just comparable

concepts. So, in order to compare or evaluate different experimental designs with regard to their variability, the ICC is still often employed. Further arguments brought up against the ICC are its sensitivity to outliers and that it only provides purely statistical and disregards, for example, clinical information [LLC<sup>+</sup>12].



# Emotional feedback in neurofeedback studies and other related work

## 5.1 Emotional feedback in neurofeedback studies

The use of emotional feedback in NF settings has some readily apparent advantages. Mathiak et al. [MAK<sup>+</sup>15] pointed out that it is an intuitive way for subjects to understand the amount and polarity of the feedback. In contrast to other forms of feedback, such as moving bars, it is not necessary to understand the exact connection between the feedback and the desired activity in the target region or even the entire training context. As mentioned in Mathiak et al., the perception of a moving bar in a feature film, for example, would not produce a rewarding effect, but a smiling face would [MAK<sup>+</sup>15]. This is especially important since there is evidence that explicit instructions about possible strategies to gain control over a specific target area have no beneficial [MJM<sup>+</sup>16] or even inhibiting effects. Instead it is proposed that a form of operant conditioning could take place also in NF settings [SSR<sup>+</sup>16]. Emotional reward, which is universally understandable, would be particularly useful for this approach.

Another form of implicit and context-independent feedback would be the addition of a monetary reward depending on the training success, as investigated in Sepulveda et al. [SSR<sup>+</sup>16]. Despite the fact that such a reward cannot be offered online during the training sessions, there are some other unwanted side effects: Following the differences in subjects' response towards reward, it increases the inter-individual variability of NF training outcomes. Monetary reward shifts the focus from intrinsic to extrinsic motivation. This can interfere with the training goals because, on the one hand, intrinsic motivation leads to more sustained long-term success and, on the other hand, extrinsic motivation is only effective as long as it is present. When the reward is discontinued, the training effect

disappears [SSR<sup>+</sup>16]. Emotional feedback, on the other hand, is well suited to reinforce intrinsic motivation. It simulates situations in daily life in which behaviour is adjusted in order to receive social reward and activates natural reward-related systems [MAK<sup>+</sup>15]. In the following, some studies are presented that concern with the possibilities and benefits of emotional feedback in a NF context. Some studies that use emotional feedback are cited as examples.

The first study discussed here that explicitly deals with the use of emotional facial expressions as a form of social feedback is the study by Mathiak et al. [MAK<sup>+</sup>15].

The aim of the investigation was to test if social reward leads to improved learning outcomes compared to other feedback designs, in this case a moving bar. For this purpose 24 healthy subjects were separated into two groups of the same size and an fMRI NF training with group-specific feedback types was carried out. The training process consisted of sessions on three following days with three runs each. One run comprised eight NF and nine baseline blocks (see Figure 5.1). During the NF blocks, subjects were supposed to regulate the dorsal anterior cingulate cortex (dACC), while in baseline blocks they were told to count backwards from 100. The subjects were instructed to recall positive emotional autobiographic memories or to imagine performing their hobby. Additionally, two transfer runs were carried out, one before and one after the NF runs, where the subjects were told to control the activity of the dACC, but no feedback was shown. This was done to test if the learned strategies could be transferred to a situation without feedback and to account for side effects due to the measurement situation.

The emotional feedback was given in form of a white male avatar, with either dark or fair hair. Depending on the degree of activation in the dACC, the smile of the avatar changed gradually. A change in the undesired direction led to a neutral looking face of the avatar, which was also shown during baseline periods.

The group, which received social feedback, reached a higher activation in the dACC during NF training than the group with the standard feedback. The results are depicted in Figure 5.2. In the transfer run, subjects of both groups were able to increase the activation of the sgACC significantly without feedback, but no significantly higher activation for social reward could be found after statistical error correction.

Although it is known that the ACC activation is raised by visual exposure to emotional pictures, such as smiling or sad faces [PWTL02], there is no mention of an additional activation due to feedback presentation. The authors do mention a possible impact on the signal due to the fact that the ACC is also part of the reward system and its activation could raise the measured activation signal of the ACC.

In the transfer runs, where no feedback was shown and a bias due to emotional triggers can be excluded, no significant improvement due to social feedback was found. This fact could be seen as indication, that at least some of the benefits originate in the additional stimulus due to feedback perception.

Another field of rising interest is the development of brain-computer interfaces (BCI) and in this context the interaction with so-called ‘virtual agents’, a virtual entity that interacts with the user via speech, facial expressions, or body movements. A study by

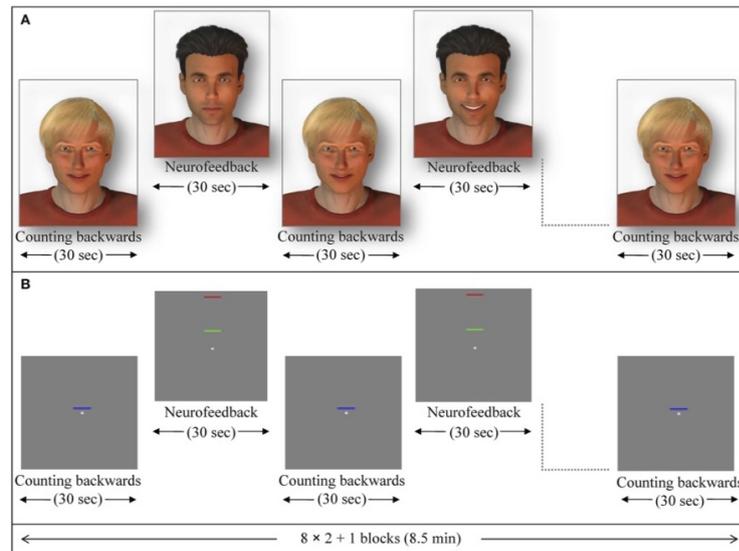


Figure 5.1: **Schematic structure of the study design in Mathiak et al.** NF blocks are alternated with baseline blocks. Two different feedback designs were compared: One group received social reward feedback in the form of a smiling avatar; the second group received standard feedback in the form of a moving bar [MAK<sup>+</sup>15].

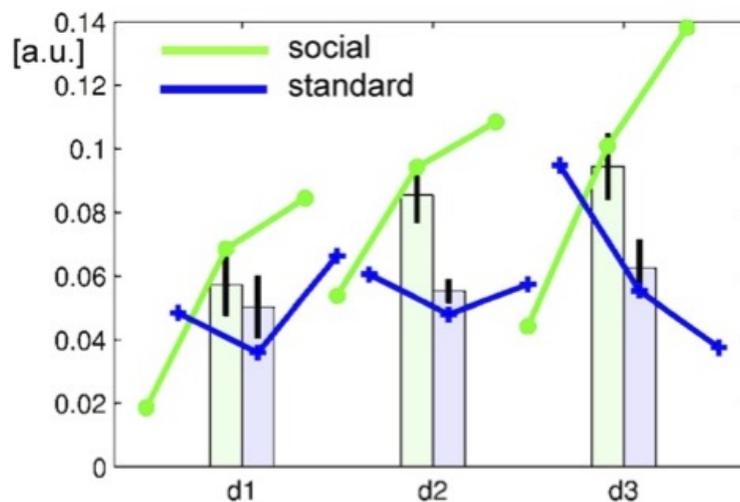


Figure 5.2: **Difference in regulation success between the two feedback designs.** For all three days regulation success was higher for the group that received social feedback (bars). The change in the course of the runs was also higher with social feedback (lines) [MAK<sup>+</sup>15].

Aranyi et al. [APC<sup>+</sup>16] aims at such an interaction with virtual agents and examines the question of whether it is possible that subjects can establish a mental connection with an agent only through their thoughts and mental strategies. More precisely, they wanted a positive approach between subject and agent to be achieved. The agent would react with a realistic posture and facial expressions to a measured sign of approach on the subject's part. This emotional feedback situation was part of a functional near-infrared spectroscopy (fNIRS) NF session. During the session the subjects were instructed to express positive feelings towards the agent in order to capture its interest, but to refrain from talking, frowning, or moving their limbs.

The subjects' attitude was captured by measuring the difference in activation in the left and right prefrontal cortex (PFC). The asymmetry between the two hemispheres is interpreted as the intensity of approach towards the agent. The agent, in the shape of a dark haired woman, was 'responding' to the subject with an expression matching the perceived interest in both valence and intensity: The asymmetry values were mapped linearly onto the agent's facial expression and the gaze was changed accordingly. This was integrated to account for the special role of the gaze in social engagement situations, in which gaze avoidance is one of the most prominent indicators for disengagement. Depending on the measured approach-intensity of the subject, four levels of engagement are defined on the agent's side:

- **Low engagement:** The agent gazes away, turns away from the subject and has a lip pout
- **Below average engagement:** Neutral facial expression with gaze and body oriented towards the subjects
- **Above average engagement:** Agent gazes directly at the subject and has a smile of mild intensity
- **High engagement:** The smile widens, crow's feet appear around the eyes, cheeks are raised and the head slightly tilts backwards

A depiction of the different levels can be seen in Figure 5.3.

A total number of 18 subjects were included in the study. The whole session consisted of eight blocks with three epochs: a resting epoch, a view epoch, and a NF epoch (Figure 5.4). In the view epoch the subjects were instructed to look at the agent and carry out a mental arithmetic task, such as counting backwards from 500. This epoch was taken as reference point for the NF evaluation.

About half of the NF blocks were successful, which means that the subjects were able to increase the average asymmetry in a statistically significant way during the NF epoch, compared to the view epoch of the same block. This was interpreted as the ability of subjects to engage with a virtual agent through mental strategies.

The asymmetry in PFC activity is known to be increased by the perception of emotional faces. Additionally, the left PFC is activated as a reaction to reward, whereas the right PFC as a reaction to punishment [AFAC08]. The facial expressions of the agent can thus

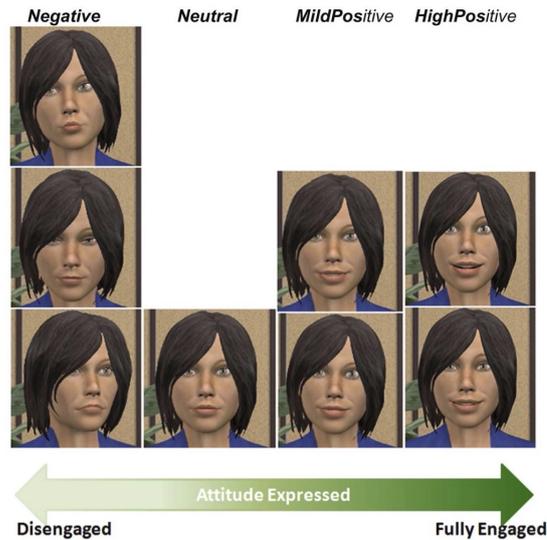


Figure 5.3: **Reaction of the virtual agent depending on the expressed engagement of the subject.** Different facial expression, gaze and head tilt are shown in four different levels of engagement [APC<sup>+</sup>16].

be assumed to stimulate the measured asymmetry, an effect, which is further reinforced by the fact, that the engaging behaviour of the agent should act as a reward. This possible side effect of the feedback perception is not mentioned or discussed by the authors.

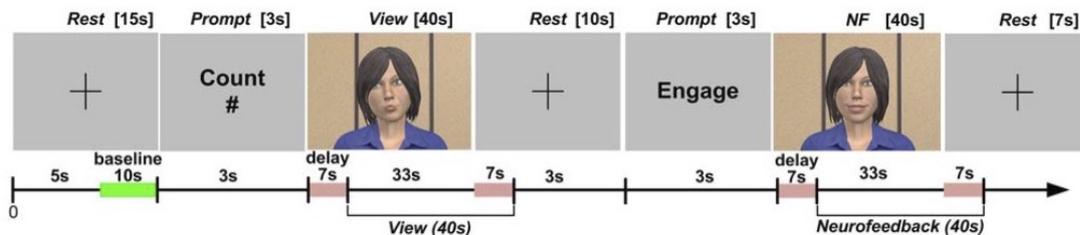


Figure 5.4: **Session design in Aranyi et al.** Three different epochs are carried out: Rest epoch, where the subjects look at a centred cross, View epoch, where the subjects look at the agent and count backwards, and an NF epoch, where the subjects try to engage with the virtual agent [APC<sup>+</sup>16].

In order to understand the mechanisms of mental disorders that are associated with the impairment or misperception of social interaction, such as autism spectrum disorder (ASD), the examination of affected brain regions with NF can provide crucial insights. A possible target for corresponding clinical studies could be the posterior superior temporal sulcus (pSTS), a core region in face perception and the imagery network.

To test this, Direito et al. [DLS<sup>+</sup>19] carried out a study that examined if the pSTS could be localized based on the presence of dynamic facial expressions (in contrast to static face signals or mere motion perception) and whether it can be regulated with the help of fMRI NF training.

During the functional localizer part the subjects were told to look at changing facial expressions of a white male avatar, while their brain activation was measured using fMRI. The presented images changed between a static image of a neutral face, a face morphing from neutral to happy, a face morphing from neutral to sad, a face with alternating expressions between sad and happy, and the presentation of moving dots. The neutral face was integrated to be able to subtract static aspects of face processing, the moving dots help subtract mere motion processing without emotional connotation. An exemplary stimulus sequence of a functional localizer run is shown in Figure 5.5.

This approach was interpreted as successful by Direito et al. since they showed that there exists a highly selective region in pSTS, encoding dynamic aspects of facial expressions. The area localized in this way was used as target area for the following NF training.

For this, they recruited 20 healthy subjects. A group of 10 subjects trained the localized area in pSTS, the other group trained another brain region non-related to the emotional task. Where one part of each group received auditory feedback in the form of a sound with varying frequency, the other part received visual emotional feedback in the form of changing facial expressions of the same avatar used in the functional localizer stage (see Figure 5.6). This feedback design was chosen due to the emerging trend of choosing rewards directly related to the goals of the intervention. The instruction for the group with visual feedback was to control the facial expression of the avatar, the group with auditory feedback was told to imagine corresponding facial expressions. The whole training session consisted of 25 blocks, where in 12 blocks the pSTS should be up-regulated (corresponding to thoughts or imagery of happy or sad faces) and in 13 blocks the target area should be down-regulated (corresponding to the imagery of neutral faces). After the NF training, a transfer run was carried out, in which no feedback was given, to test the learned ability of regulation without feedback.

For both feedback designs subjects were able to control the activation of the pSTS during the NF trainings. In the group with visual emotional feedback seven out of ten got significant results. In the group with auditory feedback this was only the case for four out of ten subjects. In the transfer run a total number of four participants succeeded to significantly increase the activation in pSTS without feedback.

Even if the same design is used for estimation of the target area and for the feedback calculation, no additional effect due to feedback perception is mentioned. The changing facial expressions given as feedback may have the same influence on the activation as the changing facial expressions supposed to activate the pSTS in the functional localizer run, so a further investigation of a feedback-induced increase in activation could be useful.

There are a few other studies in which social reward is used, for example Cordes et al. [CMD<sup>+</sup>15] and Shoji et al. [SPC17], but they use exactly the same feedback design as proposed in Mathiak et al. [MAK<sup>+</sup>15] and none of them mention a possible effect of

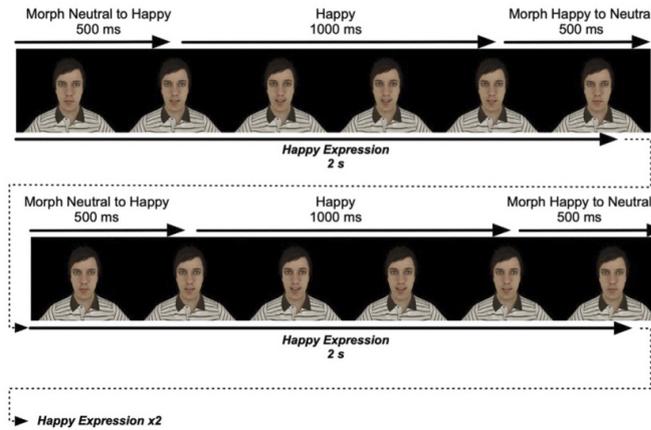


Figure 5.5: **Exemplary design of a functional localizer run in Direito et al.** A region encoding dynamic aspects of facial expressions is to be estimated, so the presented mood of the avatar is nearly continuously changing [DLS<sup>+</sup>19].

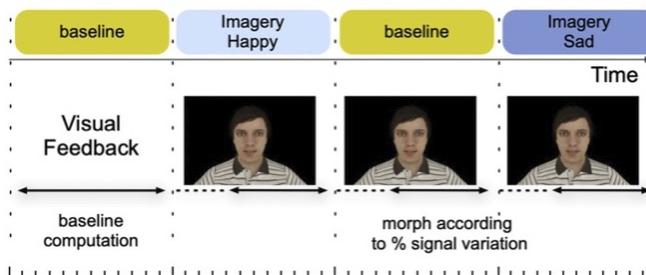


Figure 5.6: **Design of a NF run in Direito et al.** The facial expression of the avatar is gradually changing according to the measured activation in the pSTS. In baseline periods a neutral expression is presented [DLS<sup>+</sup>19].

emotional stimulation by the feedback, although areas involved in emotion processing are directly targeted.

One exception is the study by Klöbl et al. [KMG<sup>+</sup>20], where emotional feedback is used to train the sgACC in fMRI NF. Since the data of this study is the base of this work, it is separately presented in the next subsection.

In summary, it can be stated that emotional feedback is used for different approaches to the use of NF. Although in all the cases presented, brain areas are to be activated that are involved in emotion processing, the influence of feedback perception has not yet been investigated.

## 5.2 Previous research in this project

NF may offer therapy options for patients with manifestations of neurological or psychiatric diseases that have not yet been treatable with conventional approaches. Since it is a drug-free, non-invasive method, which addresses the self-control of subjects, NF could, e.g., provide a solution to the widespread problem of treatment-resistant depression. A project currently conducted at the Neuroimaging Labs of the General Hospital Vienna aims to develop such a therapeutic approach.

A first step towards this goal was to find the most effective feedback scheme, a problem investigated in an fMRI NF pilot study with 12 subjects [KMG<sup>+</sup>20]. Two feedback schemes were compared: One group containing of six healthy subjects received an empowering feedback, a picture of a smiling face, if they succeeded to regulate the target region in the desired direction and a neutral feedback if they failed. The second group (six healthy subjects) received a positive feedback for success as well, but a punishing feedback, in form of a sad face, in the undesired case. While most of the comparable studies are based on the first scheme, there are indications that positive punishment, in this case in the form of negative feedback, could also play a decisive and possibly even stronger role in the learning effect [FGA<sup>+</sup>08][SAL<sup>+</sup>11][Str14]. The aim of this pilot study was to determine whether these findings could also be transferred to fMRI NF settings.

The target area investigated in this study was the sgACC. Since the long-term goal is the development of a therapy for depression, a brain region was chosen that is known to be affected in mood disorders and therefore also an effective treatment target for deep brain stimulation [DST08]. This region of interest was localized by roughly adopting the paradigm used by Hamilton et al. [HGH<sup>+</sup>11] (compare Figure 5.7B): The voxel cluster that showed the most significant increase in activation, when the subject was shown strongly negatively affective images, was spatially averaged and used as basis for feedback calculation.

Referring to Mathiak et al. [MAK<sup>+</sup>15], implicit social reward was chosen as feedback signal, but instead of a white male avatar, an abstract face in the shape of a ‘smiley’ was used to avoid individual preferences and dislikes regarding gender, ethnicity, or age amongst other reasons.

Each training run consisted of eight regulation blocks (30 seconds), in which the subjects were supposed to decrease the activation of the sgACC, flanked by eight baseline or resting periods of the same duration. The subjects were informed about the involvement of the sgACC in emotion processing as a primary guidance, but no specific strategies were provided. During the regulation blocks continuous feedback was given; the degree of smiling indicated the percentual signal change of the activation. After each regulation block a resuming feedback was given depending on the regulation ability in the second half of the block (see Figure 5.8). Two training sessions were scheduled on two different days, which contained three training runs each. Additionally a transfer run was carried out before and after the training sessions. In these transfer runs no feedback was given, while the task to down-regulate the sgACC stayed the same, to evaluate if the strategies of the subjects could be transferred to non-feedback situations.

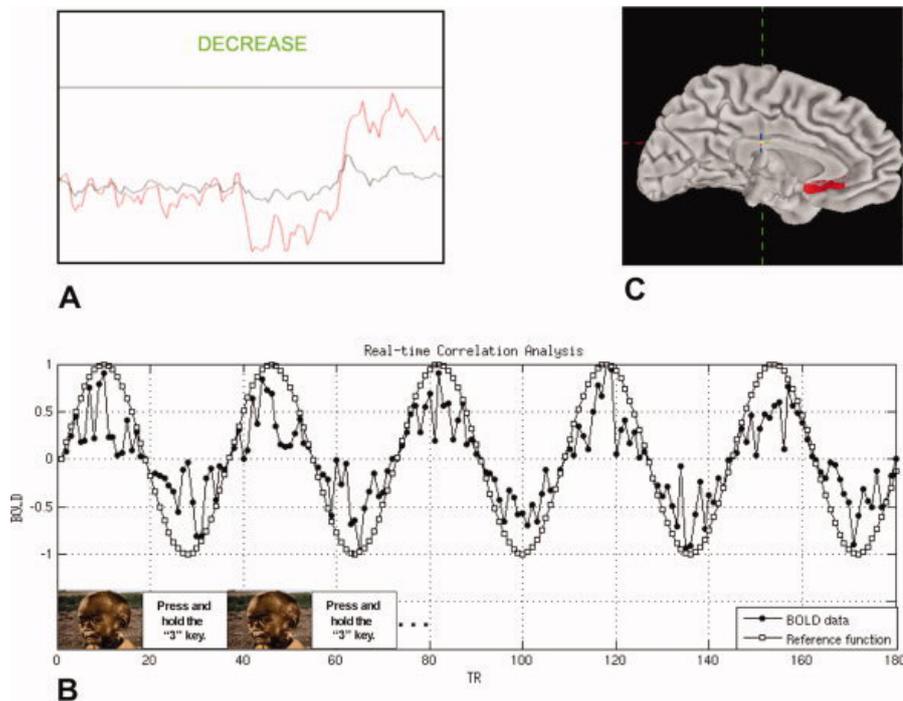


Figure 5.7: **Feedback scheme, Functional localizer and target area in Hamilton et al.** A: Feedback was shown in form of scrolling graphs. If the sgACC is successfully deactivated the red line (sgACC activation) falls under the black line (whole brain minus sgACC activation). B: The target area is localized by a contrast between the perception of a sad face and a baseline task. The voxel group with the highest activation was chosen as target area. C: Estimated target area as the part of the sgACC that is responsible for negative affection [HGH<sup>+</sup>11].

In the offline analysis each regulation block was separately modelled by a boxcar and a linear regressor, convolved with the HRF, as can be seen in Figure 5.8. This way it was possible to record changes between the single blocks as well as within a block, which would indicate learning effects.

It could be shown that the group, which perceived positive and negative feedback reached a higher regulation success in the first session and a higher ability to control the target area in the desired direction in the first runs of both sessions. These results indicate that a faster learning is possible in NF settings if the reward spectrum is expanded to positive punishment. In the transfer runs after the sessions, no significant effect could be found for any of the groups.

The possibility of an additional effect on emotion-related target areas due to emotionally connoted feedback is mentioned in the limitations of the study. The data of this study was provided as basis for the current work.

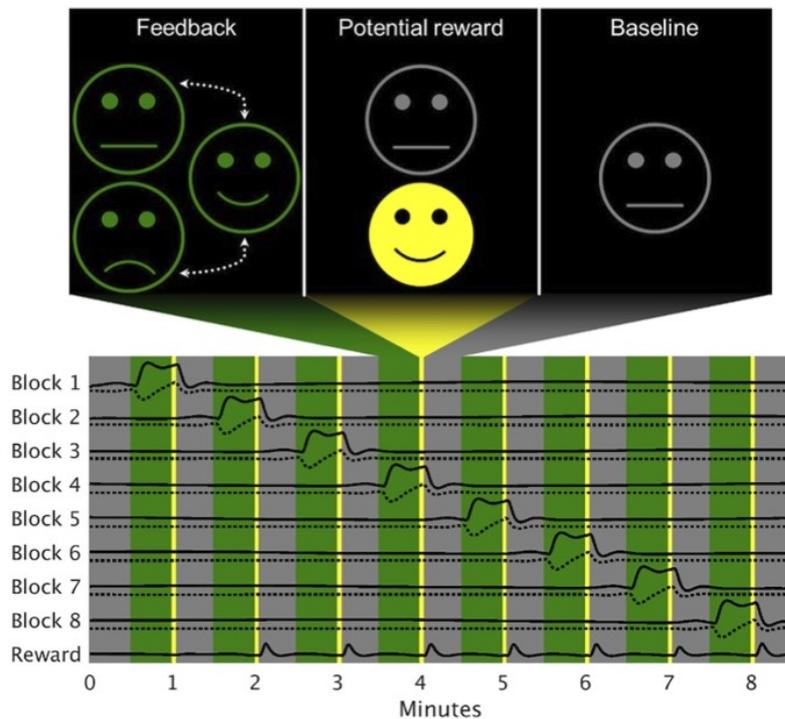


Figure 5.8: **Exemplary model design of a single neurofeedback run in Klöbl et al.** Grey areas represent baseline periods (a grey neutral looking face is shown), green areas represent the active regulation condition, where continuous feedback is given. At the time points marked with a yellow stripe, intermittent feedback was given depending on the performance of the second half of a block. Each regulation block is modelled by a boxcar and a linear regressor to record training effects between and within the blocks [KMG<sup>+</sup>20].

### 5.3 The sgACC in fMRI neurofeedback

Since the measurement of the sgACC activation as an exemplary brain region for emotion processing forms the data basis for this work, it is important to determine the role it has played in fMRI neurofeedback studies so far in order to be able to assess the relevance of its future investigation and possible integration into therapy approaches.

Although the anterior cingulate cortex (ACC) is the target of many NF studies [DA13], very few of them have used the subgenual part (sgACC). The probably first and only of these studies directly using sgACC activation was carried out by Hamilton et al. [HGH<sup>+</sup>11] and examines the general possibility of regulating the sgACC using fMRI neurofeedback.

The authors cite its involvement in the generation of affective states and in psychopathology (such as unipolar depressions, bipolar disorder, or obsessive-compulsive disorder) as the reason for considering the sgACC. The antidepressant effect of exogenous modulation

of sgACC activity with microelectrode stimulation could be shown in advance [MLV<sup>+</sup>05] and a similar effect following the NF training is suspected.

The sgACC is known to be spatially heterogeneous with respect to the type of affect it subserves. Since the authors are primarily interested in a possible intervention for major depression, only the part that processes negatively affected emotions is investigated. For this purpose a functional localizer is carried out before the NF training. The contrast between passive viewing of negatively connoted images, such as the picture of a crying child, and a baseline condition (pressing a button with a specific number) is analysed to localize the voxels with the highest activation that is correlated with the image perception (figure 5.7B).

To test the possibility to control this part of the sgACC (Figure 5.7C) fMRI NF training is carried out with 17 adult female subjects, separated into a NF group and a control group. The aim of the training was the successful down-regulation of the target region. The subjects were instructed to upregulate positive mood in any way, which was an instruction designed to be specific enough to make fast progress, but flexible enough to allow for personal strategies. The NF group received continuous feedback in the form of scrolling graphs. A graph with a red line indicated the sgACC activation signal, a black line the activation of the whole brain minus the sgACC (Figure 5.7A). A success was defined, if the red line falls below the black line, which means that the sgACC activation is lower than the averaged whole brain activation. The control group received sham feedback that was previously calculated by using the activation of a subject in the NF group.

The whole setting consisted of two NF training sessions, each with five downregulation blocks and five baseline blocks. Before and after the training a no-feedback or transfer run was carried out to test the transferability of the learned strategies and the regulation ability.

The activation of sgACC could be reduced during the NF training runs in the NF group, but not in the control group. During the transfer runs, neither group could achieve significant deactivation, which Hamilton et al. argue that the strategies learned cannot be verbalized enough to easily be called up and used during post-training scans.

In summary, they state that deactivation of the sgACC with the help of neurofeedback is in principle possible and should be investigated further in connection with a possible therapy for mental illnesses.

Another study with a therapeutic objective investigated the connectivity between the sgACC and the right anterior superior temporal cortex (ASTC) using fMRI NF [ZWB<sup>+</sup>19]. The rationale for this investigation are the findings of Green et al. [GRM<sup>+</sup>12], who stated that patients with MDD show a reduced connectivity between these two regions while feeling guilt or self-hate. This bias of over-generalizing self-blame is a clear clinical symptom in MDD, but the corresponding neural marker of reduced connectivity can even be seen in patients with remitted symptoms. Additionally, a particular vulnerability to MDD is found due to a bias to blame oneself for failure in an over-generalized way resulting in a decreased global self-esteem [AST78].

Therefore, there are both therapeutic and preventive purposes to test whether patients can use fMRI NF to gain control over this connectivity. Another aim of the study was to explore whether such training can lead to reduced self-hate and increased self-esteem. 28 subjects with remitted MDD, which in this case means that their last episode of depression was at least two months ago and they were currently not suffering from symptoms, were divided into two groups: an active intervention group and a control group. The members of both groups should evoke autobiographical memories that arouse in the first case guilt and in the second case indignation in them. These two emotions were contrasted with each other in the study as the first one is mostly self-centered while the other is more focused on the behaviour of others. During the training phase, self-chosen words were presented to the subjects to help them remember such a situation.

While experiencing these emotions, the subjects were told to simultaneously try to raise the level of a thermometer-like display. In the indignation phase, the level decreased when the connectivity changed in any direction; this procedure was intended to ensure a stable correlation between sgACC and ASTC as a comparison to the guilt phase. During the guilt phase, the two groups were given different tasks, since the study was double-blinded, neither the subjects nor the experimenter knew who belonged to which group. During the measurement in the active intervention group, the thermometer level could be raised by increased connectivity between sgACC and ASTC. In the control group, the connectivity should again remain as constant as possible in order to achieve a high thermometer level. The task of the control group was chosen in this way so that the same regions are addressed with the same emotions as in the active group, in order to prevent a mismatch between the training success and different tasks.

In total the training consisted of one session with four runs. In a first run, the 10 % most activated voxels in the target regions were estimated for guilt and indignation experience. These voxels were further taken to estimate the connectivity. The second and third runs were the actual training runs. Each run consisted of four guilt blocks and four indignation blocks, each block intermitted by a baseline block with a mental subtraction task to allow patients to distract themselves emotionally from the scenarios. The fourth run aimed at comparing the overall connectivity between sgACC and ASTC after the training to the one before the training.

The patients in the active intervention group were able to increase the level of connectivity in the guilt condition relative to the indignation condition. In the control run after the training, increased connectivity for the guilt versus indignation phase was found for the active group compared to the control group. An increased level of self-esteem could be found in the active group after the training, but no effect of reduced self-hate was achieved.

Although the sgACC has only been rarely studied in connection with fMRI NF, its suitability for this method has been clearly established. In combination with its strong involvement in mental disorders [DST08], this suitability makes it a useful area of investigation for the interplay of emotional or social feedback and emotion processing during NF application.

# Methods and data

## 6.1 Data origin

The data underlying this work was gained during the NF runs of Klöbl et al. [KMG<sup>+</sup>20]. The general study and feedback design are presented in Chapter 5.2.

### 6.1.1 Subject recruiting and technical details

Physically and mentally healthy volunteers from 18 to 35 years of age were recruited via postings on message boards and from a database of potential subjects. Further inclusion criteria consisted of right-handedness and signing of an informed consent form. Potential subjects were excluded due to MR incompatibilities, such as non-removable metal parts in the body, for example pacemakers, or known pregnancy. In addition, a later discovery of major internal, neurological or psychiatric illnesses or current substance abuse (including having smoked within two hours before the measurement) led to an exclusion from the study.

Of the 13 volunteers who were recruited for participation according to these criteria, one had to be subsequently excluded because, according to his own statements, the subject had tried to influence the activity in the target region by concentrating on his breathing. Since such a procedure can lead to a bias in the BOLD signal, the data of this subject could not be used for the analysis. Due to technical issues, single runs of two other subjects also could not be used.

In this work, a single run of another subject was also excluded; within this single run there was a sharp fluctuation in the baseline activation that a linear regression could only have been carried out with an intolerably high error variance. The groups for both feedback schemes were matched for sex as well as for mean and standard deviation of age.

The measurements were carried out using a Siemens Prisma 3T scanner (Siemens, Erlangen, Germany) with a 64-channel head coil. A multiband-accelerated echo-planar imaging (EPI) sequence with the following parameters was used: echo time (TE) of 30 ms, a repetition time (TR) of 483 ms and a multiband factor of 8. The field of view was set to 190 x 190 x 140 mm at 76 x 76 x 56 voxels, resulting in an isotropic resolution of 2.5 mm. The flip angle was 46° and the receiver bandwidth was 2630 Hz/Px. To avoid compressions of the frontal cortex, the phase-encoding direction was chosen as anterior-posterior.

### 6.1.2 Preprocessing

The data preprocessing during the measurement was carried out using the Turbo Brain-Voyager 4.0 beta software. Since there is only a very short period of time available in order to be able to present the feedback in real-time, this online preprocessing, in contrast to the offline analysis, only consists of few necessary steps.

An important step is the motion correction of the head during the measurement. All subsequent images are realigned to the first measured image via rigid-body transformations. For each spatial axis a translation and a rotation coefficient are determined and applied to the images. In short, this means that each newly measured volume is rotated and shifted until it corresponds as closely as possible to the first image (Figure 6.1).

In addition to the motion correction spatial smoothing is performed. The data is therefore convolved with a Gaussian kernel. To the grey value of each voxel a certain amount of the neighbouring voxels is added, which is determined depending on their distance to the central voxel using a three-dimensional Gaussian distribution, in this case it had a full width at half maximum (FWHM) of 5 mm. The central voxel consequently is aligned with the distribution mode, while voxels that are far away from it only make a very small contribution. The reason behind this process is the improvement of the spatial signal-to-noise ratio using a weighted mean. Additionally, outliers in single voxels and differences due to the spatial extent of the hemodynamic response are mitigated. The visual effect of this smoothing is shown in Figure 6.2.

After this preprocessing of the whole brain, the spatial mean over all voxels is formed, which were defined as target region (as described in Section 5.2.9). This spatially averaged time series is both the basis for the feedback estimation, which is described in the next section, as well as the basis for the offline analysis carried out in this thesis.

### 6.1.3 Feedback estimation

For the presentation of the feedback, the percentage change in the measured activation during the regulation blocks compared to the preceding baseline is calculated. This so-called 'percent signal change' (PSC) is calculated as follows:

$$PSC = \frac{\text{activation value} - \text{baseline value}}{\text{baseline value}} \cdot 100 \quad (6.1)$$

In this study, the *baseline value* is given by the median signal of the baseline period that precedes the respective regulation period, whereby the images of the first six seconds

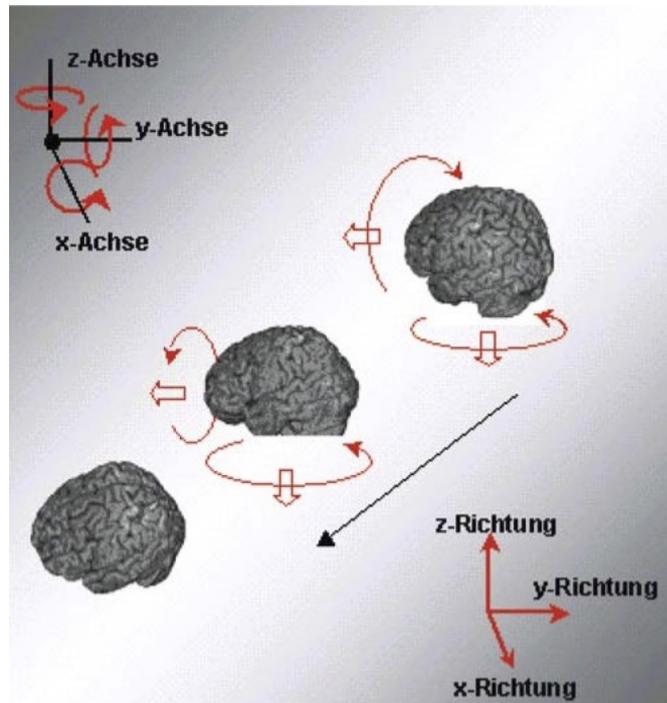


Figure 6.1: **Rigid-body transformation** Rotations and translations of the volume are possible for all spatial axes [WKH07].

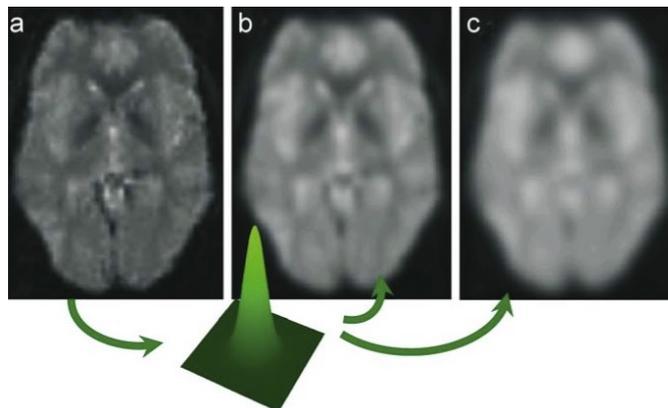


Figure 6.2: **Spatial smoothing.** One slice of a whole volume taken with the EPI method before (a) and after smoothing, where (b) shows a 6 mm FWHM kernel and (c) a 10 mm FWHM kernel [WKH07].

are excluded [KMG<sup>+</sup>20]. During this period there is a subsequent activation due to the hemodynamic delay of the previous regulation period, which is why this period cannot be rated as baseline.

During fMRI measurements of cognitive effects, usually a PSC of 1-3% is reached [JSJ<sup>+</sup>20], which is why a threshold value of 5% is set for the maximum activation or deactivation. This means that the subjects are shown a maximally smiling face if they reach a deactivation of -5% and, if negative feedback is shown, a maximally sad face if they reach an activation of +5%. The feedback was presented using the Psychtoolbox [Bra97] and MATLAB (The MathWorks, Natick, MA, United States).

## 6.2 Linear approaches of feedback influence

### 6.2.1 General approach

The presented feedback  $F_i$ , given at the time point  $i$ , is dependent on the measured activation  $A_{i-1}$  at the previous time point  $i - 1$ . For the determination of a possible stimulation through the feedback perception, an additive influence is assumed, i.e. a certain amount, which is a function  $f$  of the level of the feedback, is added to the ‘true’ activation  $A^*$  or, depending on the sign of the feedback, subtracted. Hence, the expected activation at time point  $i$  is given with:

$$A_i = A_i^* + f(F_{i-1}(A_{i-1})) \quad (6.2)$$

In this way a recursive sequence of activation measurement and feedback presentation is created, following the first activation measurement, during which no feedback can yet be presented:

$$\underbrace{A_1^*}_{A_1} \rightarrow F_1(A_1) \rightarrow \underbrace{A_2^* + f_2(F_1(A_1))}_{A_2} \rightarrow F_2(A_2, f_2(F_1)) \rightarrow \underbrace{A_3^* + f_3(F_2(A_2, f_2(F_1)))}_{A_3} \rightarrow \dots \quad (6.3)$$

Since a linear relationship between the neural response of a region and the measured BOLD signal is commonly assumed [PMN11], the additional influence is modelled as linearly dependent on the feedback level:

$$f_i(F_{i-1}) = r_i \cdot (F_{i-1} * HRF) + b_i \quad (6.4)$$

The feedback signal is convolved with an HRF to account for the delay of the BOLD signal after the perception of the stimulus. Under the assumption that  $r_i$  and  $b_i$  are constant for all  $i$ , which means that possible side effects on the emotion processing such as fatigue or habituation are neglected (as it is done for usual fMRI scan times),  $A_i$  reduces to

$$A_i = A_i^* + r \cdot (F_{i-1} * HRF) + b \quad (6.5)$$

This approach simplifies the integration into the analysis: The calculated feedback value  $F_{i-1}$  is returned by the NF software, can be convolved with the HRF and integrated

as an additional regressor into the main GLM, which already calculates  $A^*$  as a sum of different regressors that have an influence on the activation. The linear parameter  $r$  becomes an additional  $\beta$ -estimate of the GLM, the bias  $b$  is covered by the regression constant.

### 6.2.2 Choice of hemodynamic response function

The aim of fMRI measurements is to draw conclusions about the unmeasured neuronal activity from the BOLD level. To get information about magnitude, onset latency and duration of the evoked metabolic (re-)action in the brain, the amplitude, delay and duration of the used HRF can be evaluated. An accurate model of the evoked HRF thus plays an important role for the accuracy of further conclusions.

Ill-fitting statistical models of the HRF can lead to an increase of both false positive and false negative results. This has to be prevented by an accurate shape estimation of the HRF, since even minor amounts of mis-modelling can lead to a crucial loss in power and validity [LLW08].

There are a lot of models for the HRF with different possibilities to estimate amplitude, time-to-peak, and FWHM parameters [LLAW09]. In this work the most rigid model presented by Friston et al. [FFJ<sup>+</sup>98] is used, where the shape of the HRF is completely fixed and only the amplitude as the amplitude of the response can be varied, corresponding to the regression coefficient in the GLM analysis:

$$h(t) = A \cdot \left[ \frac{t^{\alpha_1-1} \beta_1^{\alpha_1} e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c \cdot \frac{t^{\alpha_2-1} \beta_2^{\alpha_2} e^{-\beta_2 t}}{\Gamma(\alpha_1)} \right] \quad (6.6)$$

with  $\alpha_1 = 6$ ,  $\alpha_2 = 16$ ,  $\beta_1 = \beta_2 = 1$ , and  $c = \frac{1}{6}$ .  $A$  denotes the variable amplitude and  $\Gamma$  the gamma function as a normalization factor. The shape of this HRF is depicted in Figure 6.3 for different amplitudes. The time-to-peak and FWHM parameters stay the same for all amplitude values.

This model is called canonical HRF and is used by default in Statistical Parametric Mapping (SPM) [Gro20], a common analysis software for fMRI measurements.

There are several reasons for choosing this model. The use of more parameters leads to an increased potential of mis-estimating them, whereas one parameter makes it easier and statistically more powerful to interpret resulting differences between measurement conditions [LLAW09]. Another advantage is the fact that the offline analysis of the used data is done with SPM, which means that the regulation regressors are also convolved with the same HRF model. Using this shape for the convolution with the feedback signal leads to an increased comparability with the other regressors.

According to Lindquist et al. [LLAW09] the canonical HRF provides adequate modelling if the duration of the stimulus does not exceed a period of approximately three seconds, which is the case in this setting, in which a new feedback stimulus is given with each new scan ( $TR \approx 0.5s$ ). For higher durations the amplitude of the stimulus is underestimated by this rigid canonical model, and a model with the possibility to vary also the time-to-peak and FWHM has to be chosen.

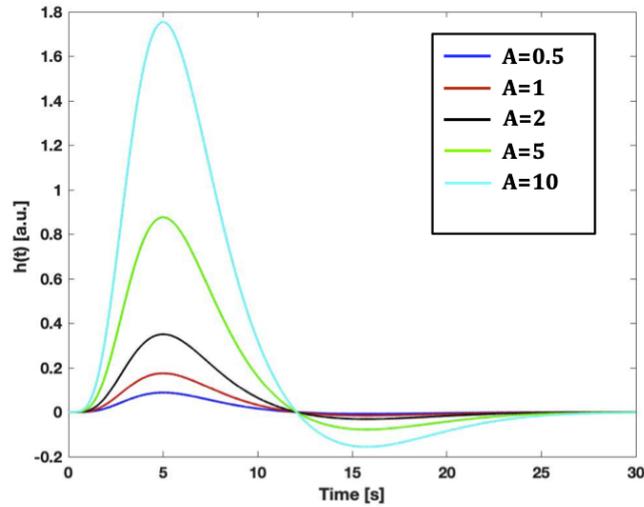


Figure 6.3: **Shape of the canonical HRF with different amplitudes.** Time-to-peak and FWHM stay the same for all amplitudes.

### 6.2.3 Valence-dependant models

The sgACC is involved in the processing of both positive and negative emotional stimuli with a spatial functional specialization separating the two valences [GSG<sup>+</sup>05]. With the help of the functional localizer, which is described in Section 5.2., the area of the sgACC that is responsible for the processing of negative emotions is determined. It could be possible that the perception of positive and negative feedback has a different influence on the target region due to this specialization. In particular, it can be assumed that the negative feedback has a higher influence.

To account for this possible asymmetry, positive and negative feedback values are treated as two separated regressors in an additional calculation. For this purpose, two feedback regressors were added in the GLM for a subject of the group that received both positive and negative feedback. To estimate the negative feedback regressor all negative feedback values were convolved with the HRF, all positive values were set to zero. The opposite was done for the positive feedback regressor.

The data of the two groups was analysed separately, to be able to compare the groups regarding differences in the feedback perception, depending on the range of feedback (only positive versus both positive and negative feedback).

### 6.2.4 Inclusions of the time derivatives of the HRF

As described in Section 6.2.2., the shape of the selected HRF is fixed concerning time-to-peak and FWHM parameters and only the amplitude of activation can be changed. This can lead to incorrect modelling if there are time delays in the physiological response of

the investigated region. Such time delays can arise, for example, from measuring an area in distinct neuronal layers or from spatially variable blood flow [CSPK04].

In this case the real BOLD response  $h^*(t)$ , as a function of the (post-stimulus) time  $t$ , is delayed by a small amount  $dt$  relative to the canonical response  $h(t)$ :

$$h^*(t) = \alpha \cdot h(t + dt) \quad (6.7)$$

where  $\alpha$  is a scaling factor. A first-order Taylor expansion results in:

$$h(t + dt) \sim h(t) + h'(t)dt \quad (6.8)$$

with  $h'(t)$  as the temporal derivative of the HRF. The real HRF  $h^*(t)$  can thus be expressed as:

$$h^*(t) = \beta_1 h(t) + \beta_2 h'(t) \quad (6.9)$$

where  $\beta_1 = \alpha$  and  $\beta_2 = \alpha \cdot dt$  [HPR<sup>+</sup>02]. To account for the temporal delay of the response it is useful to integrate both the canonical HRF and its derivative into the GLM as distinct regressors. The weights  $\beta_1$  and  $\beta_2$  are then the  $\beta$ -estimates for each regressor convolved with the HRF. The temporal delay  $dt$  can further be estimated as:

$$dt = \frac{\beta_2}{\alpha} = \frac{\beta_2}{\beta_1} \quad (6.10)$$

According to this relation, a positive  $\beta_2$  implies that the HRF peak occurs earlier than expected, whereas a negative  $\beta_2$  indicates a delayed peak [Gro20]. The time derivative of the HRF compared to the canonical HRF can be seen in Figure 6.4.

In an additional calculation, the convolution of the feedback level with the time derivative of the HRF is integrated into the GLM to test whether this leads to an improved estimation of the feedback influence. In a third approach the regressor that is convolved with the canonical HRF and the regressor that is convolved with the time derivative of the HRF are combined. It could be shown by Calhoun et al. [CSPK04] that statements about the significance of the non-derivated part of the model can be improved in their accuracy if a new amplitude is calculated by combining the canonical HRF and its time derivative. This new amplitude is determined as follows:

$$A = \text{sign}(\beta_1) \sqrt{\beta_1^2 + \beta_2^2} \quad (6.11)$$

In this work, this new regressor is estimated by combining the two regressors (feedback values convolved with the canonical HRF and its time derivative) in the above way. This new regressor is integrated into the GLM instead of the two others and a new  $\beta$ -value of this combination is estimated.

### 6.2.5 Overview of modelling approaches used in this work

In the table of Figure 6.5, all three modelling approaches that are utilized in this work are presented. To avoid misunderstandings about the implementation and results, the different approaches are numbered. This numbering will also be used in the following chapters to precisely identify the method described.

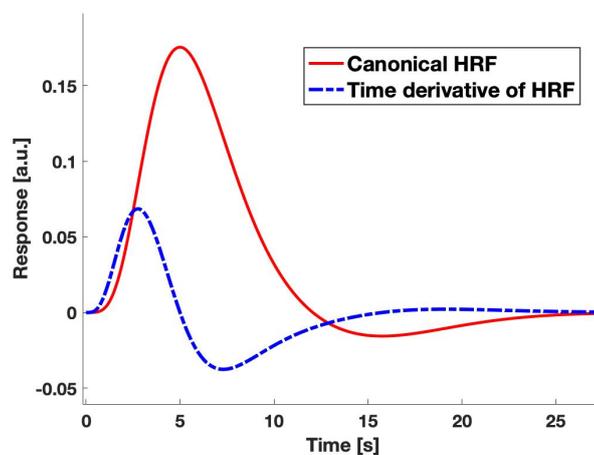


Figure 6.4: **Time course of canonical HRF and its time derivative.** Both are convolved with the feedback values and integrated into the GLM

	Group 1 - Only positive feedback	Group 2 - Positive and negative feedback	
		Pos.+neg. feedback combined	Pos. + neg. feedback separated
Only canonical HRF	1 a)	2 a)	3 a)
Canonical HRF + time derivative	1 b)	2 b)	3 b)
Canonical HRF + time derivative comined as one regressor	1 c)	2 c)	3 c)

Figure 6.5: **Numbering of the different approaches investigated in this work.**

## 6.3 Design of the GLM

### 6.3.1 Design matrix

One run consisted of a sequence of eight regulation blocks, in which the task was to reduce the activation of the sgACC. The subjects received both continuous feedback about their success as well as an overall feedback in explicit ‘reward blocks’ at the end of each regulation block. Each of the latter was flanked by a baseline block in which no feedback was given. The whole training took place on a total of two days with sessions of three runs each.

In order to achieve the highest possible comparability, the design matrix for modelling the regulation of the target area is taken from Klöbl et al. [KMG<sup>+</sup>20]. The reward blocks are not modelled as they are separated from the next regulation by a baseline block and, hence, not expected to have an influence on it.

Each of the eight regulation blocks per run is individually modelled with a boxcar and a sawtooth function resulting in 16 different regressors plus one constant term modelling the baseline.

All regulation regressors are convolved with the canonical HRF implemented in SPM. For all approaches except for 1a), 2a) and 3a) the regressors are also convolved with the time derivative of the HRF.

All regressors of a single regulation block are orthogonalized to one another in order to avoid collinearity. The used regulation regressors are exemplary shown for one block in Figure 6.6.

While all regulation regressors, which are implemented according to Klöbl et al. [KMG<sup>+</sup>20], are integrated into the design matrix by using SPM’s batch module ‘fMRI model specification’, the feedback regressors are prepared separately and added manually to the design matrix. The single steps of the preparation are presented in the following sections.

### 6.3.2 Convolution with the HRF

The time series of the presented feedback values can be seen as a function  $F(n)$  or  $F(t)$  of the frame number  $n$  or time  $t$ . The convolution for continuous functions is defined as follows:

$$(h * F)(t) = \int h(\tau)F(t - \tau)d\tau \quad (6.12)$$

Since  $F(n)$  is a discrete time series, the convolution results in:

$$(h * F)(n = i) = \sum_{m=1}^M h(m)F(i - m) \quad (6.13)$$

where  $M$  is the number of frames. In Figure 6.7, a raw feedback time series is presented on the left side, on the right side the resulting regressors can be seen, i.e. the feedback time series convolved with the canonical HRF and its time derivative. The convolution

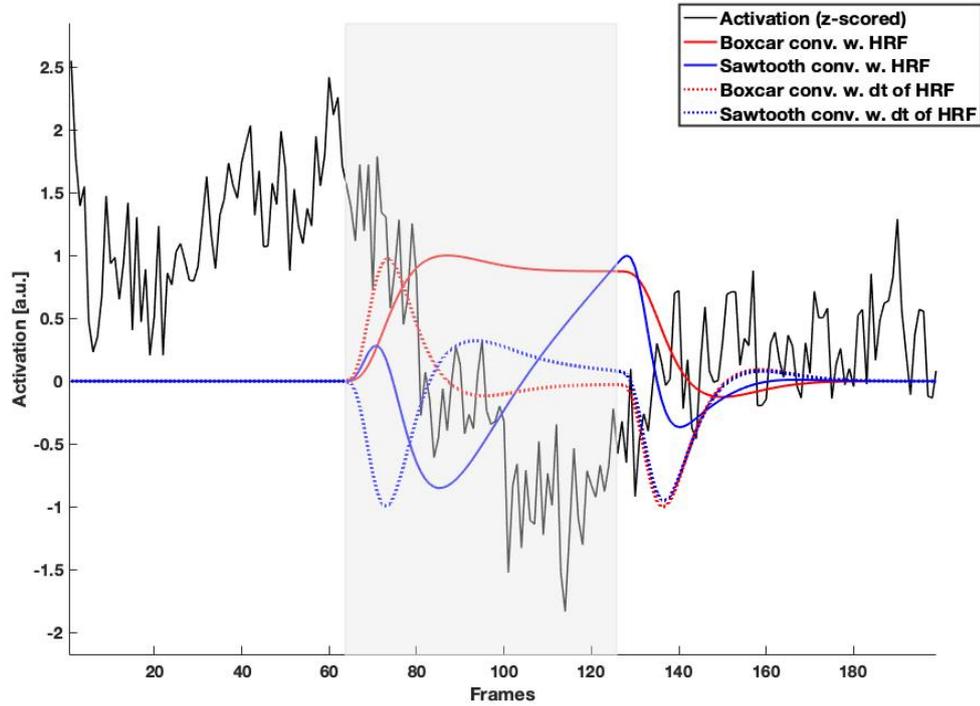


Figure 6.6: **Four different regulation regressors are used to model the activation timeline (black)** The presented regressors are the boxcar condition convolved with the HRF (solid red) and the time derivative of the HRF (dotted red), as well as the sawtooth condition convolved with the HRF (solid blue) and the time derivative of the HRF (dotted blue). The grey area denotes the onset of the regulation condition.

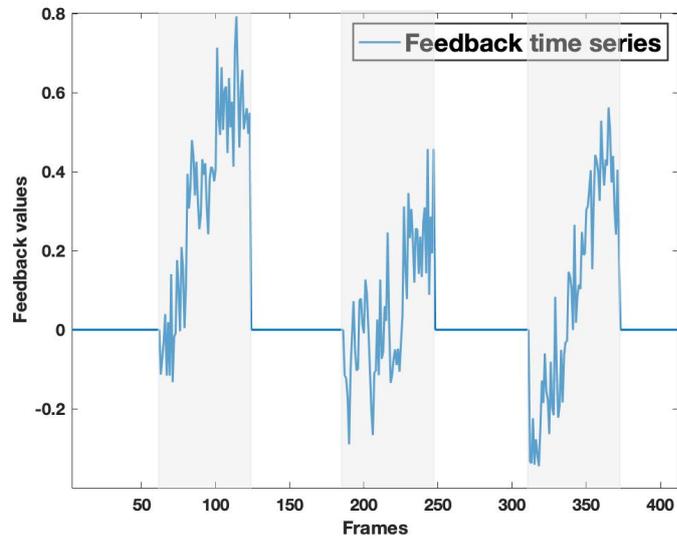
leads to a smoothing of the time course and a non-zero influence of the feedback that goes beyond the regulatory periods marked as grey areas.

### 6.3.3 Orthogonalization

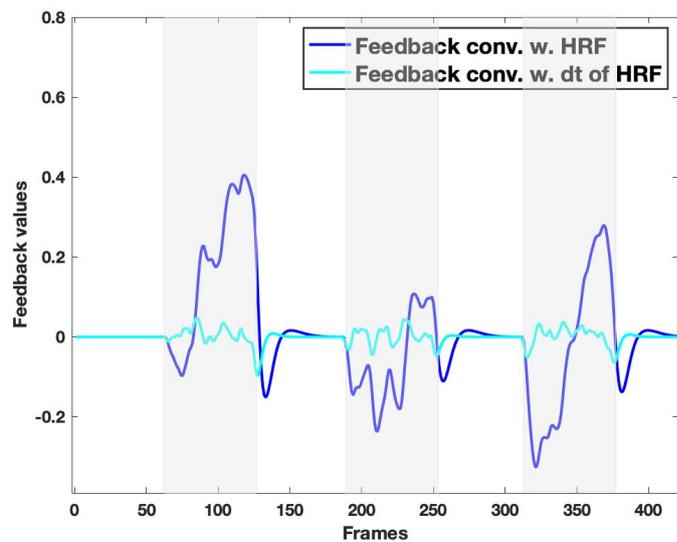
The feedback signal is necessarily correlated with the regulation regressor due to the fact that the feedback is calculated from the activation time series, which in turn is modelled by the regulation regressor. Therefore it must be assured that both regressors are orthogonalized to avoid an underestimation of the regulation coefficient due to shared variances. On Figure 6.8 (a), the correlation between the regressors, denoted by Pearson's Correlation Coefficient

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6.14)$$

is shown before their orthogonalization. As first regressor the regulation regressor is used, as second one the feedback regressor convolved with the canonical HRF, and as third



(a)



(b)

Figure 6.7: **Raw feedback time series and convolution with HRF.** (a): The raw time series for three regulation blocks are shown. (b): The same time series is convolved with the HRF and its time derivative, resulting in a smoother time course and a delayed feedback influence. The grey areas represent regulation blocks.

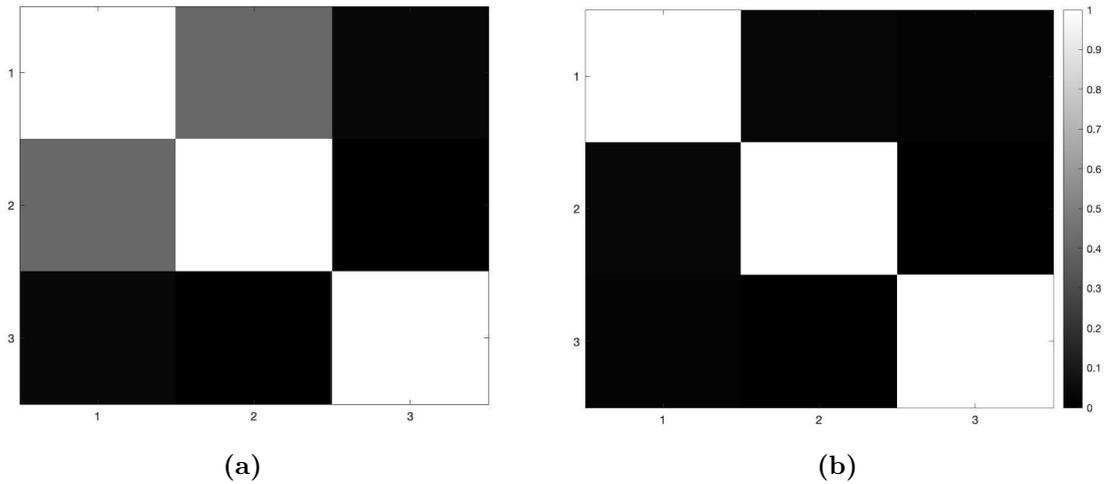


Figure 6.8: **Correlation between the three regressors before (a) and after (b) orthogonalization.** After application of Gram-Schmidt orthogonalization, the correlation between feedback and regulation regressor vanishes. **1:** Regulation regressor, **2:** Feedback regressor convolved with HRF, **3:** Feedback regressor convolved with time derivative of HRF

one the feedback regressor convolved with the time derivative of the HRF. As can be seen, the correlation between the first feedback regressor and the regulation regressor has a value of around 0.4, whereas the time derivative regressor is uncorrelated to both of the other regressors.

To resolve this problem, the feedback regressors are orthogonalized to the regulation regressor by a Gram-Schmidt procedure. This algorithm transforms the (non-orthogonal, but linearly independent) vectors  $w_1, w_2$  and  $w_3$  into the new vectors  $v_1, v_2$  and  $v_3$ , which are pairwise orthogonal:

$$\begin{aligned}
 v_1 &= w_1 \\
 v_2 &= w_2 - \frac{\langle v_1, w_2 \rangle}{\langle v_1, v_1 \rangle} v_1 \\
 v_3 &= w_3 - \frac{\langle v_1, w_3 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle v_2, w_3 \rangle}{\langle v_2, v_2 \rangle} v_2
 \end{aligned} \tag{6.15}$$

where  $\langle \cdot, \cdot \rangle$  denotes the scalar product of two vectors. It can be seen, that the first vector stays the same, and all other vectors are orthogonalized in a serial manner. Since the regulation regressor should not be influenced by the additional existence of the feedback regressors, it is important, that this one is the first regressor in the orthogonalization order.

In Figure 6.8 (b), the correlation between the regressors after the orthogonalization is presented. The pairwise correlation between all three regressors is around zero, so the shared variance will be minimized.

### 6.3.4 Amplitude normalization

In order to ensure an evenly distributed influence of the regressors and to prevent the values of the  $\beta$ -estimates from reflecting the (arbitrary) magnitude of the amplitude of the regressors, normalization is necessary.

Therefore all values of a regressor  $\vec{r}$  are divided by its maximum deflection:

$$\vec{r}_{norm} = \frac{\vec{r}}{\max(\text{abs}(\vec{r}))} \quad (6.16)$$

where  $\text{abs}(\vec{r})$  are the absolute values of the regressor elements. After this transformation, all regressors have an amplitude in the interval  $[-1,1]$ .

### 6.3.5 Residual correction

Considering the residuals of the model, linear or even quadratic trends can be seen in some runs (see Figure 6.9 (a)). These trends can be explained by measurement artefacts such as magnetic drift or physiological changes. In order to model these artefacts, a linear and a quadratic term are added to the design matrix as regressors of no interest. As can be seen in Figure 6.9 (b), these additional regressors result in an almost trendless distribution of the residuals around zero.

Apart from this step, the residuals are not further tested for normality. Common tests for normal distribution, such as the Kolmogorov-Smirnov test, are too sensitive for a sample of this size. Instead, the law of large numbers applies and a normal distribution can be assumed. Hypothesis tests, as used in this work, also show a high degree of robustness against the violation of the normal distribution criterion of the residuals [GPS72][LKK96], which is why incorrect results cannot be assumed due to the existing distribution. Auto-correlation of the residuals, which is usually an issue in subject-level fMRI models, is not further taken into account as it is of no particular importance for group-level analyses.

## 6.4 Linear mixed-effects model design

### 6.4.1 Estimation of hierarchies in the data

The  $\beta$ -estimates are calculated for single runs. For each of the 12 subjects there are therefore six individual estimates, which originate from two sessions with three runs each. It can be assumed that the reaction to the presented feedback differs from person to person, i.e. that there is variance between individual subjects. It is also conceivable that there is a grouping with regard to the sessions and the runs. For example, it would be possible that the response to the feedback is stronger during the first run than during the following runs, as the subjects are more attentive. On the other hand, it would also be possible that the influence of the feedback increases with increasing run number due to, for instance, frustration effects. The same applies to the sessions. In the first session, in which the measurement procedure is still new and unfamiliar, the feedback on the

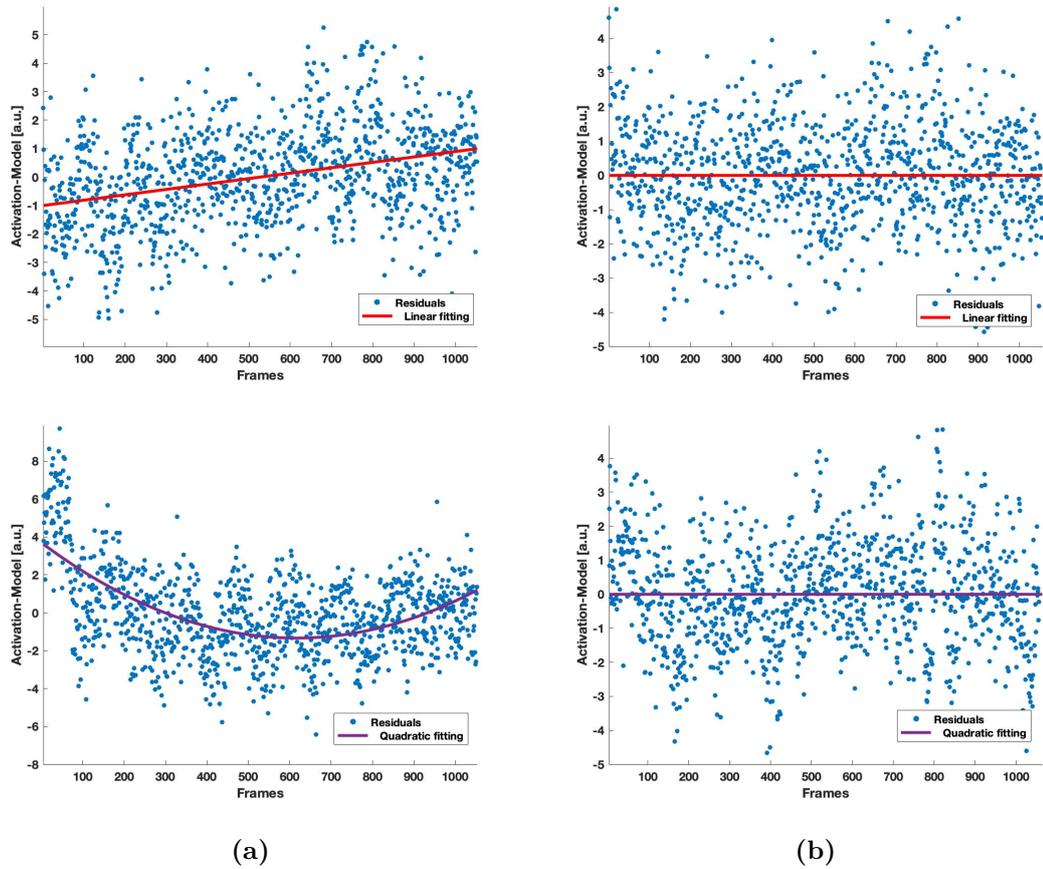


Figure 6.9: **Distribution of the residuals before (a) and after (b) the integration of additional regressors.** While on the left side, a linear or even quadratic trend can be found, the residuals are approximately normally distributed around zero after the integration of a linear and a quadratic regressor into the design matrix.

regulation success could have a different effect than in the second session, where the scenario is already known.

In order to test whether a hierarchical treatment of the data in the context of a mixed-effects model is useful, the ICC is calculated as described in Section 4.2.1. for the grouping into subjects, runs, and sessions. The ICC values of this grouping are shown in Figure 6.10 for the different model approaches. Regarding these results it can be concluded that the variance between the individual subjects explains a significant part of the total variance. 'Subject' should therefore be included in the mixed-effects model as a random effect. In contrast, the variances of the single runs and sessions explain only a negligible part of the total variance. These groupings are therefore not taken into account in the mixed-effects model. The differences in the ICC can also be seen in Figure 6.11. The  $\beta$ -estimates are clustered according to subject (exemplary for approach 1a), run (approach 1b), and session number (approach 3a) and the respective mean value is

Approach Nr.	Subjects	Runs	Sessions
1a)	0.1695	0.0176	0.0049
1b)	0.179/0.1253*	0.0278/0.016*	5.1643e-05/2.106e-04*
1c)	0.2008	0.0256	7.42E-05
2a)	0.1556	1.99E-04	0.1193
2b)	0.1602/0.4169*	0.0057/0.0078*	0.0876/0.0089*
2c)	0.2417	0.0033	0.0363
3a)	<b>0.0625</b> / <u>0.1497</u>	<b>0.0276</b> / <u>0.0728</u>	<b>0.1037</b> / <u>1.6521e-04</u>
3b)	<b>0.1445</b> / <u>0.1235*</u> / <u>0.0273</u> / <u>0.1635*</u>	<b>0.0134</b> / <u>0.0269*</u> / <u>0.0284</u> / <u>0.0086*</u>	<b>0.016</b> / <u>0.0021*</u> / <u>0.1089</u> / <u>0.0019*</u>
3c)	<b>0.1601</b> / <u>0.071</u>	<b>0.0109</b> / <u>0.0268</u>	<b>0.0069</b> / <u>0.0654</u>

Figure 6.10: ICC values of the groupings into six different subjects, three different runs and two different sessions for each model approach. The values marked with a \* denote the  $\beta$ -estimates for the time derivative of the HRF, bold font denotes the estimates of the positive feedback, underlined numbers denote the ones of the negative feedback, for the approaches where they are treated separately. As can be seen, the values in the 'Subjects' column justify the inclusion of this grouping into the LME, whereas the values of the grouping of 'Runs' and 'Sessions' are negligible.

displayed. While the mean values for the individual subjects differ obviously from each other, the mean values for the runs and sessions are approximately the same. In addition, both the variance and the interval, in which the estimates are located, within the run and session groups are approximately the same, so that a separate treatment of the data would not lead to a different result than complete pooling.

#### 6.4.2 Fixed effects and random effects

The  $\beta$ -estimates, which are determined with the help of the GLMs for the individual runs, should be combined into a common coefficient, so that statements about the entire group can be made and comparisons drawn. This common coefficient is a fixed effect, which is necessary for generalizing the results of the investigation. Since it is assumed to be constant over time and should not in any way depend on parameters of interest, this is a fixed intercept. As such it describes the generally valid influence of the feedback on the activation of the target area in this group. By testing the significance of such a fixed intercept, it is investigated whether such a constant influence is generally present.

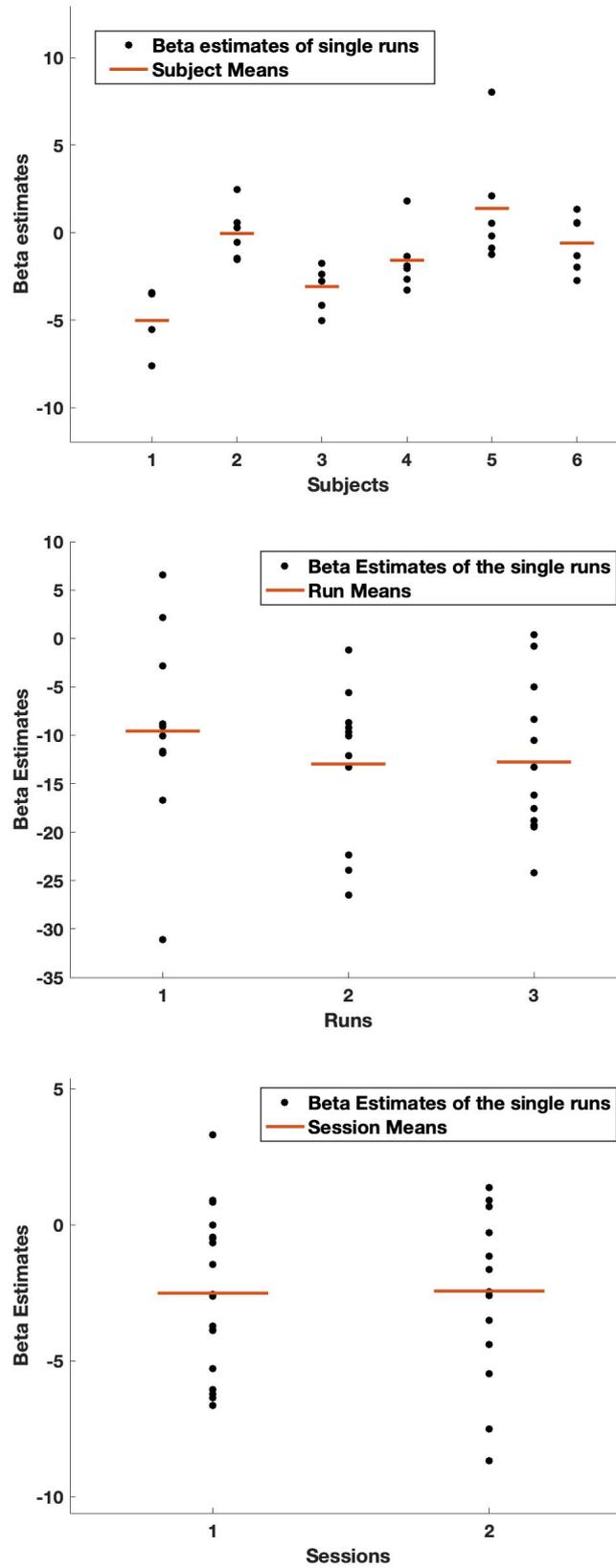


Figure 6.11: Clustering of the  $\beta$ -estimates according to the subject, run and session they are estimated from. The red lines denote the respective means. The means for the subjects vary strongly, whereas the distributions over the single runs and sessions are similar to each other.

The group's fixed intercept is biased by the influence of the random intercepts, which are caused by the variation of the  $\beta$ -estimates for the individual subjects.

The modelling of the  $\beta$ -estimate  $\beta_{ij}$  of the  $i$ -th subject in group  $j$  is done by summing up the fixed intercept  $\bar{\beta}_j$  of the  $j$ -th group and a subject-dependent random intercept  $u_i$ :

$$\begin{aligned}\beta_{ij} &= \bar{\beta}_j + u_i \\ i &\in [1, \text{number of subjects in group } j]\end{aligned}\tag{6.17}$$

The  $\beta$ -values are not dependent on the time or any other predictor, so there is neither a fixed nor a random slope.

To determine the significance of the feedback influence, also comparing the different modelling approaches, the fixed intercept  $\bar{\beta}_j$  determined in this way is used, i.e. the value adjusted for influences and fluctuations between the single subjects.

## 6.5 Cross-validation for single subjects

In order to test whether the influence of the emotional feedback on the sgACC activation is generalizable, a cross-validation is carried out for each model. For each subject, the  $\beta$ -value that describes the influence of the feedback is replaced by the one that was calculated as the fixed intercept  $\bar{\beta}_j$  of the other subjects in the group using the LME model. For the first subject, the  $\beta$ -estimates of subject two to six are combined and the fixed intercept  $\bar{\beta}_j^{-S_1}$  is estimated ( $-S_1$  in this case denotes that  $\bar{\beta}_j$  is calculated without the first subject). This value is then used as the new  $\beta$ -value, which describes the feedback influence of the first subject, and new model values  $\vec{y}_{\bar{\beta}_j^{-S_1}}$  are calculated. In order to determine the quality of this new model containing a generalized  $\beta$ -estimate, the concordance correlation coefficient  $\rho_c(\vec{y}, \vec{y}_{\bar{\beta}_j^{-S_1}})$  between the measured data  $\vec{y}$  and the new model data  $\vec{y}_{\bar{\beta}_j^{-S_1}}$  is calculated. A possible improvement or deterioration of this new model compared to the originally determined model  $\vec{y}$  of the subject is estimated by comparing the two correlation coefficients  $\rho_c(\vec{y}, \vec{y})$  and  $\rho_c(\vec{y}, \vec{y}_{\bar{\beta}_j^{-S_1}})$ . In addition, the correlation  $\rho_c(\vec{y}, \vec{y}_{\beta_{FB=0}})$  between measurement data and model data is determined for the case that the feedback is assumed not to have any influence at all, i.e. the  $\beta$ -value in the weight vector is set to 0.

This procedure is repeated for all subjects of a single approach and the mean value of the correlation coefficients is formed in order to be able to compare the approaches with one another. To test if the correlation between different approaches differ significantly, the p-values of the differences are estimated according to Eq. 4.17 with

$$\mathbf{L} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

## 6.6 Implementation in Matlab and SPM

All necessary steps were implemented with the help of Matlab, a software for solving mathematical problems and for graphically displaying the results. The software focuses on the calculation of numerical problems using matrix operations. It is also possible to read in data and manipulate them using a number of Matlab's own commands. These properties make Matlab a widely used tool in teaching and research in natural sciences. The software SPM is a toolbox for Matlab, which is specially designed to analyse brain-imaging data and to create spatially extended statistical processes for carrying out hypothesis tests on this data. With this tool, which is widely used in functional imaging, the complete statistical evaluation of the fMRI data can be carried out, but in this work it is only used to create the design matrix, as a number of manual manipulations are necessary and the statistical evaluation is limited to the significance of the feedback regressor.

The implementation of the steps described so far is presented in more detail below in order to offer the interested reader the opportunity to review and reproduce all calculations.

### 6.6.1 Reading of the data

The activation and feedback time series are read from the measurement log file. This file contains the following information for each of the 1056 frames:

- Activation value of the target region (preprocessed in the way that is described in Section 6.1.2.). This time series will represent the dependent variable  $\vec{y}$ .
- Current condition, that is, the information whether the regulation is carried out in the current frame or not. Depending on the condition, this variable has the value 1 or 0 and is used as boxcar regressor.
- The feedback value that is currently shown to the subject and is calculated from the previous activation and baseline

The three time series obtained in this way, arising from 1056 single fMRI frames, are stored in vectors for further use. Figure 6.12 shows the Matlab graphical user interface (GUI) when reading and storing the data.

### 6.6.2 Creation of the design matrix

In a next step, the design matrix  $\mathbf{X}$  has to be created. In order to avoid manually creating a matrix with 1056 measured values and thus 1056 rows, SPM is used. Here it is possible to create a design matrix with any number of regressors by entering various parameters, for example the number of measured frames and the time interval between them.

#### 6.6.2.1 Addition of regulation regressors and baseline

The regulation regressors are added to the design matrix using the SPM function 'fMRI model specification (design only)'. The SPM graphical editor is not used for this step,

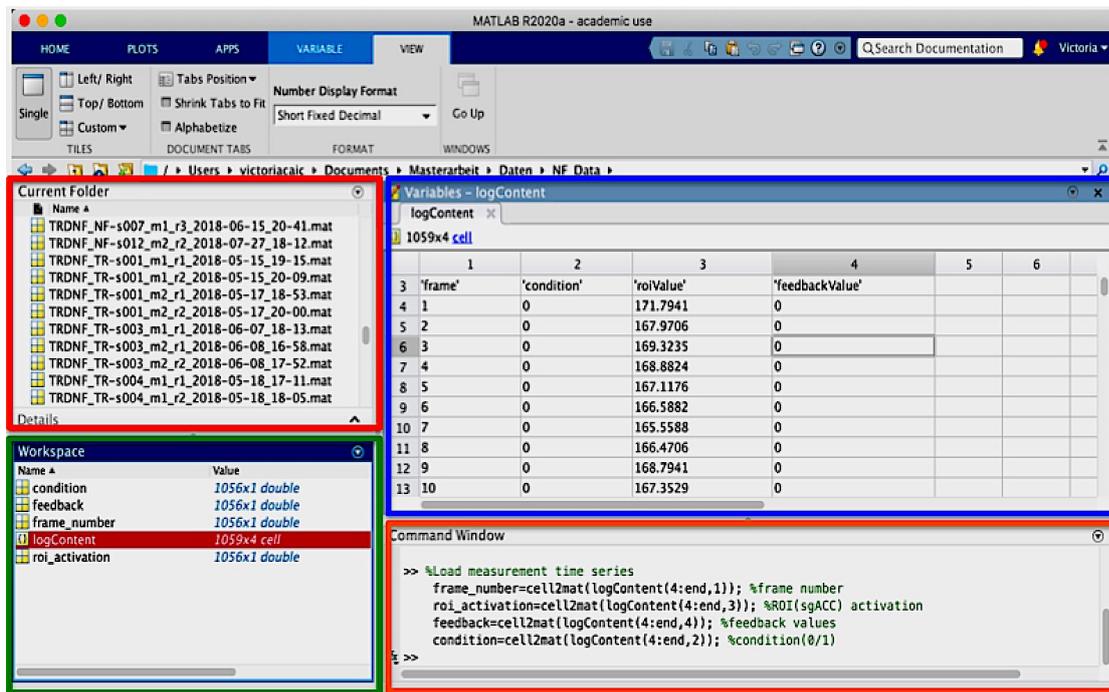


Figure 6.12: Graphical user interface of Matlab during loading and storing of the measurement data. For each subject, a logfile of the measurement exists (red), which contains the frame numbers, the current condition, the measured activation of the target region and the feedback value (blue). These time series are read from the logfile and stored into vectors (orange). An overview of all existing vectors and their sizes are displayed in the window marked in green.

but the parameters necessary for the function to run are specified using a Matlab script, as can be seen in Figure 6.13. This is done using the expression

$$\text{matlabbatch}\{1\}.\text{spm}.\text{stats}.\text{fmri\_design}$$

and the specification of the corresponding parameters as subnodes. In this way, all necessary parameters are stored in a variable called ‘matlabbatch’. The SPM function is then executed by the command

$$\text{spm\_jobman}('run', \text{matlabbatch})$$

All standard parameters for creating a design matrix are retained. The following parameters are set according to the measurement:

- Units: ‘Scans’ instead of ‘seconds’
- Repetition time: 0.483s

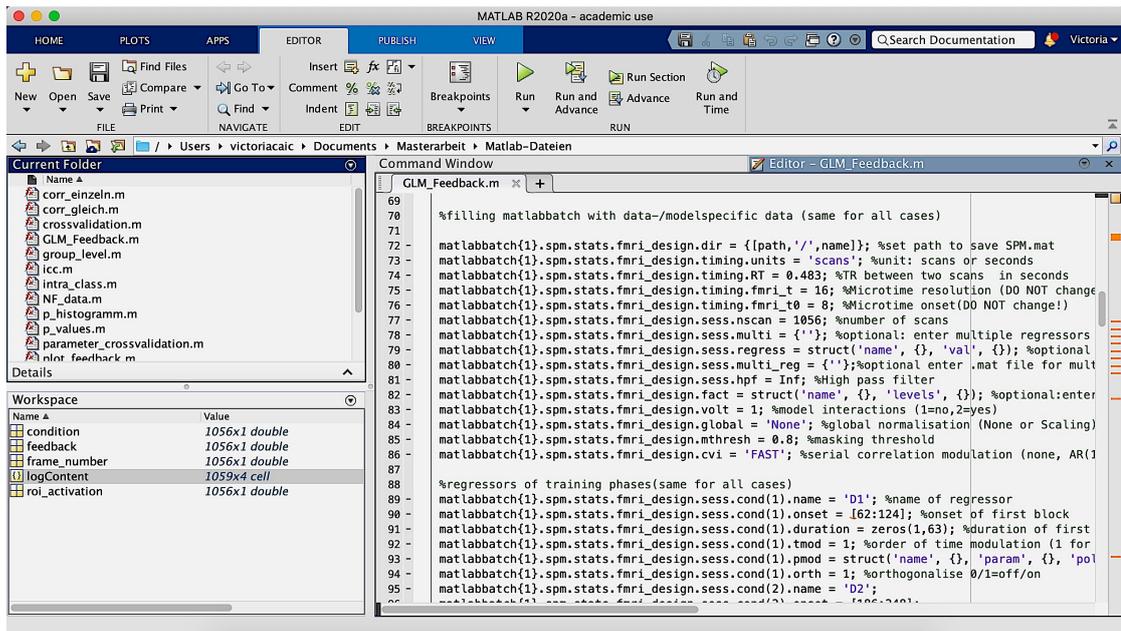


Figure 6.13: **Matlab Editor with the script defining the parameters for SPM.** In Matlab's editor window, a script is prepared, that specifies all necessary parameters for SPM to create the design matrix by storing the parameters in a variable called '*matlabbatch*'. The script is then executed by the command '*spm\_jobman*'.

- Number of scans: 1056
- High pass filter: Inf

The high pass filter is set to an infinitely high value, which means that frequency components should not be filtered out. This setting is necessary in order to maintain trends in learning behaviour.

In the next step, the run-specific regulation regressors are defined. According to the original analysis in Klöbl et al. [KMG<sup>+</sup>20] eight different regressors are used, one for each regulation block. To create a regressor in SPM it is necessary to define the onset and the duration of the 'on' condition. All other frames will be automatically set by SPM to be in the 'off' condition. For all frames belonging to the 'on' condition, SPM enters a '1' in the design matrix, for each frame belonging to the 'off' condition a '0' is entered. The onset of each block can be identified from the 'condition' time series. Each time the condition value changes from '0' to '1', a new regulation block starts. The actual frame numbers are manually identified in the 'condition' vector and entered as parameters into the SPM function.

A single regulation block consists of 63 frames, so the duration parameter could be specified in SPM as '63' for each regulation block. Instead, the duration is determined by a vector, which consists of 63 zeros. A duration of length '0' is the coding for a stimulus or 'on' condition that has the length of only one frame, i.e. corresponds to a Dirac peak.

If 63 stimuli with the length of one frame are combined as in this data set [KMG<sup>+</sup>20], this equals to a block of length 63. This step is necessary since the time modulation, which is carried out in the next step for the creation of the sawtooth regressor, is only performed correctly by SPM in this way.

In the SPM function it is possible to insert an additional regressor that modulates the boxcar regressor with a specified power of the time. This power is set to '1' in our case in order to obtain a sawtooth function that increases linearly during the regulation period (i.e., corresponding to a linear trend). The orthogonalization parameter is set to '1', which is the parameter value that ensures that the modulated regressor is orthogonalized to the base one.

By default, all regressors of the design matrix are convolved with the HRF described in Section 6.2.2. by SPM. In addition, it is possible to add the convolution both with the time derivative and dispersion derivative of the HRF. This option is not selected for the model approaches 1a), 2a) and 3a). For all other approaches, the convolution with the time derivative is activated by changing the corresponding parameter value to '1'.

Since a baseline value must be determined for each regression, regardless of the specific model, this baseline regressor is automatically added to the design matrix by SPM. The baseline is a constant value, so a constant function in the form of 1056 ones is used for this purpose, which is then adjusted to the specific baseline value by the regression.

### 6.6.2.2 Addition of feedback regressors

Since positive and negative feedback are considered separately in the approaches 3a)–c), two different time series are created from the individual feedback, each containing only one valence. In this way, the negative-feedback time-series contains only negative feedback values, the positive time series only positive values, at all other times the feedback has the value zero (Figure 6.14). Depending on the modelling approach, the next step is to convolve the combined or separated feedback time series with the HRF. For this, a canonical HRF is created and convolved with the feedback using the Matlab function *conv()*:

$$feedback\_hrf = conv(feedback, hrf)$$

If included, the convolution with the time derivative of the HRF is conducted the same way. The command window for this step is shown in Figure 6.15.

For the orthogonalization of the feedback with the eight regulation regressors, the latter are first combined to form a regressor that contains all points in time at which regulation took place. This regulation regressor, the feedback (convolved with the HRF), and possibly the feedback convolved with the time derivative of the HRF are then orthogonalized using the SPM function

$$spm\_orth(regulation, feedback\_hrf, feedback\_hrf\_dt)$$

As can be seen from the description of the Gram-Schmidt method in Section 6.3.3., the order of the regressors is important here. Since the first vector remains unchanged by this

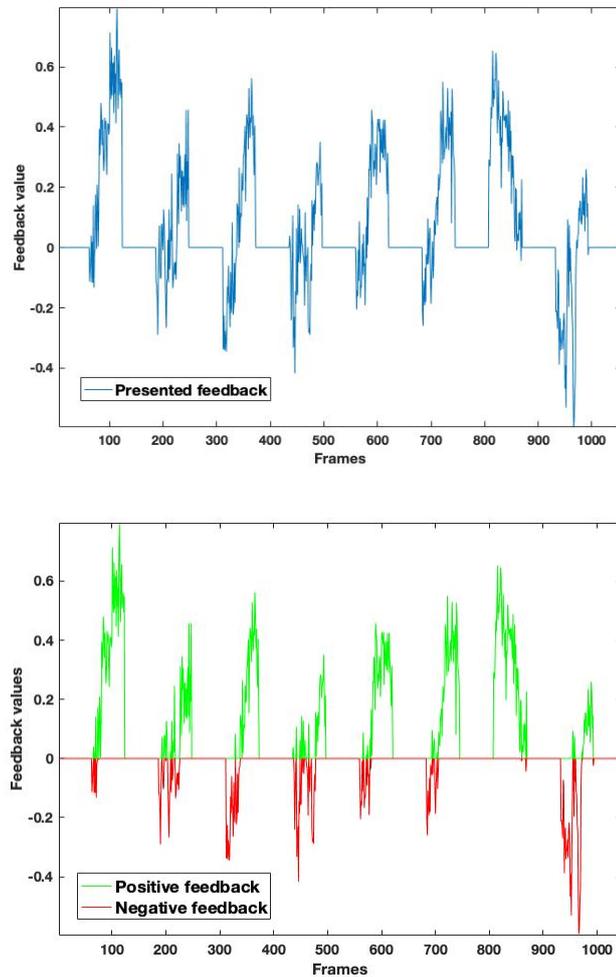


Figure 6.14: **Separation of feedback time series.** All positive feedback values belong to the positive feedback time series (green) while all negative feedback values belong to the negative feedback time series (red)

Name	Value
condition	1056x1 double
feedback	1056x1 double
feedback_hrf	2111x1 double
frame_number	1056x1 double
hrf	1056x1 double
hrf_dt	1056x1 double
logContent	1059x4 cell
roi_activation	1056x1 double

```

fx >> %Convolving the feedback with the hemodynamic response
hrf=(frame_number.^5).*exp(-frame_number)/gamma(6)-1/6*(frame_number.^5).* ...
exp(-frame_number)/gamma(16);
hrf_dt=1/gamma(6)*(5*(frame_number.^4).*exp(-frame_number)+(frame_number.^5).*...
exp(-frame_number)*(-1))-1/(6*gamma(16))*(15*(frame_number.^14).*exp(-frame_number)...
+(frame_number.^15).*exp(-frame_number)*(-1));

feedback_hrf=conv(feedback,hrf);
if dt~=0 %time derivative
    feedback_hrf_dt=conv(feedback,hrf_dt);
end

```

Figure 6.15: Matlab's command window during convolution of the feedback regressors with the HRF and its time derivative.

method, it is necessary to use the combined regulation regressor as the first argument, since this one should not be influenced by the insertion of an additional feedback regressor. The feedback regressor is set as the second argument and for all approaches except 1a), 2a), and 3a), the time derivative regressor is orthogonalized to both previous regressors.

For the model approaches 1c), 2c), and 3c), the feedback regressors and the time derivative regressors are combined into one regressor in the next step, according to the Formula 6.11 In the last step, all regressors created in this way are amplitude-normalized and appended to the previous design matrix as additional columns.

### 6.6.2.3 Trend regressors

The last additional regressors are the two regressors to model the linear and quadratic trends in the signal. The frame number is used as the linear regressor and the square of the frame number is used as quadratic regressor. To prevent these regressors from becoming disproportionately large, and thus leading to numeric inaccuracies (e.g. the corresponding  $\beta$ -weights becoming so small, that they are treated as zero by Matlab) they are also amplitude-normalized before being inserted.

### 6.6.3 Estimation of $\beta$ -values

Since the design matrix is no longer singular after orthogonalization, the weight vector  $\vec{\hat{\beta}}$  can be determined analytically according to Equation 4.7. Matlab offers the function *fitlm*( $\mathbf{X}, y$ ) for this purpose, where  $y$  in this case is given by the activation time series and  $\mathbf{X}$  by the previously created design matrix. The function does not only return the estimates for the  $\beta$ -weights, but also information about

- the estimate of the standard error  $\hat{\sigma}_{\beta_j}$  of the  $\beta$ -weights, which can be used to determine the confidence intervals of the estimate
- the value  $t_j$  (compare Equation 4.18 )
- the p-value of the respective weight ( a  $\beta$ -weight with a value under 0.05 is called 'significant') and
- the degrees of freedom of the error  $df_e$ .

The output of *fitlm* is shown by way of example for a subject of the model approach 1a) in Figure 6.16.

### 6.6.4 Correction of activation time series

After determining the weight of the feedback regressors, the entire influence of the feedback on the activation of the target area can be determined and the 'true' activation  $A_i^*$  at time point  $i$  can be calculated for the single runs:

$$A_i^* = A_i - (\vec{F}_{i-1} * HRF) \cdot \vec{r} \quad (6.18)$$

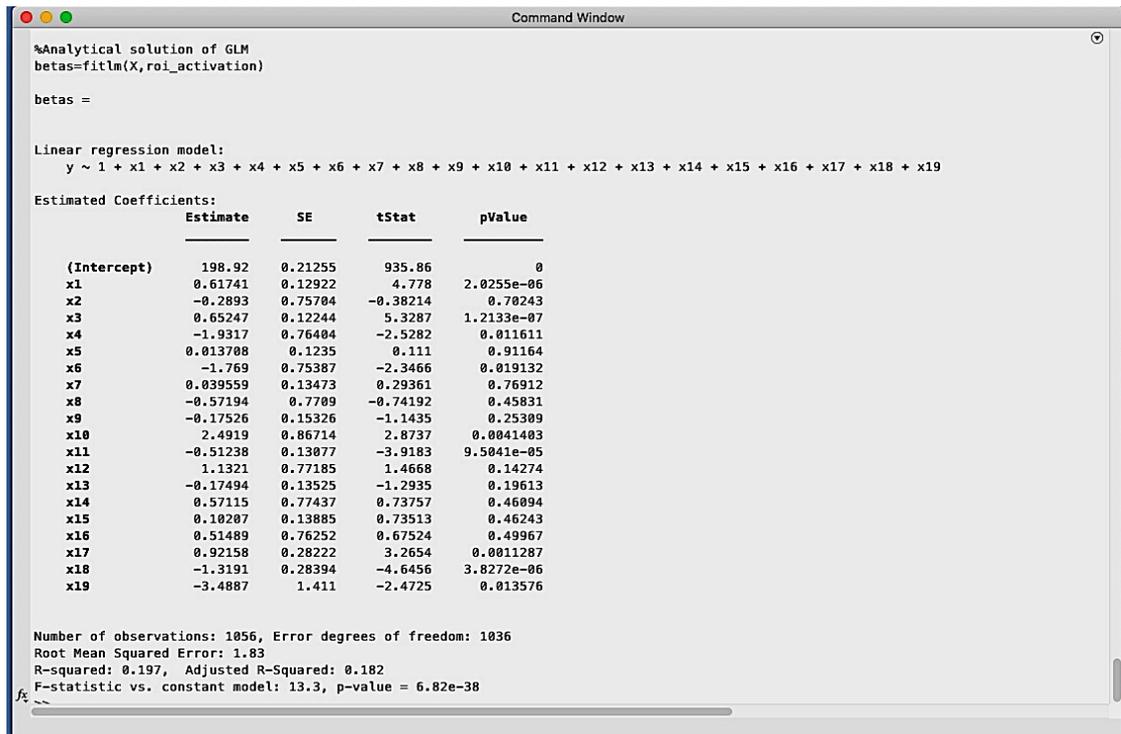


Figure 6.16: Output of Matlab’s linear fitting function *fitlm* for one subject of model approach 1a). The function returns the estimates of the single  $\beta$ -weights and a wide range of statistical information such as the standard error (SE) or the t-Value (tStat) about these estimates and the linear fit.

The term  $\vec{F}_{i-1} * HRF$  is a  $(1 \times n_F)$  vector, with  $n_F$  as number of feedback regressors, which can be read from the design matrix. The  $(n_F \times 1)$  vector  $\vec{r}$  consists of the  $\beta$ -weights that describe the influence of the respective regressors.

### 6.6.5 Group level estimates

The fixed intercept  $\bar{\beta}$  of the LME model, which was set up in Section 6.4.2., is calculated using the Matlab function for creating an LME model, *fitlme*. To do this, a table must first be created that contains both the single run  $\beta$ -weights of the individual subjects under the name ‘Parameter’, as well as the respective subject number to which the value belongs (see Figure 6.17):

```
tbl = table(single_run_betas, subject_number, 'VariableNames', {'Parameter', 'Subject'})
```

In addition, the model equation must be defined, which in this case results in

$$'Parameter \sim 1 + (1|Subject)', 'FitMethod', 'ReML')$$

The Matlab function returns a wide range of information, as can also be seen in Figure 6.18:

```

>> group_level('/Users/victoriacaic/Documents/Masterarbeit/Daten/Ergebnisse/Ergebnisse_ohne_dt_Gruppe_1.xlsx','Parameter')
%Linear Mixed Effect Model:
%Create a table
tbl=table(T(:,parameter).Variables,T.Subject,VariableNames,{Parameter,Subject})

tbl =
    34x2 table
      Parameter    Subject
      _____    _____
      -3.4887         1
      -5.5291         1
      -3.44           1
      -7.6219         1
      -1.4617         3
      0.28217         3
      -1.5196         3
      -0.56568         3
      0.55554         3
      2.4509          3
      -2.3565         8
      -4.1706         8
      -1.7604         8
      -2.7826         8
      -2.3885         8
      -5.0233         8
      -2.6568         10
      -1.3667         10
      -2.046          10
      -3.2763         10
      1.7988          10
      -1.9031         10
      8.0367          13
      0.53918         13
      -1.2556         13
      2.1103          13

```

Figure 6.17: Creation of a tyble in Matlab with all necessary parameters for the LME design.

- **Model fit statistics:** Among other statistics, the function returns the AIC, the Bayesian Information Criterion (BIC) and the (minimized) *logLikelihood*, i.e. the logarithm of the likelihood, for the chosen model.
- **Fixed effects coefficients:** For the selected fixed effect, in this case only the intercept  $\bar{\beta}$ , the best estimate, the standard error, the value  $t_j$ , the degrees of freedom (DF), the p-Value and the 95 % confidence interval is returned.
- **Random effects covariance parameter:** For all random effects, in this case the random intercept  $u_i$  per subject  $i$ , the estimates of the standard deviation  $\sqrt{\sigma_u^2}$  and the 95% confidence intervals of  $\sqrt{\sigma_u^2}$  are given. In the event that there are several random effects, the covariance and the confidence interval of the covariance between the individual effects are also returned.
- **Error distribution:** The standard deviation  $\sqrt{\sigma_\varepsilon^2}$  and the 95% confidence interval of  $\sqrt{\sigma_\varepsilon^2}$  are returned for the resulting error term of the model.

This way it can be determined whether the feedback coefficient, which has been cleared of the subject-specific fluctuations, has a significant influence on the activation of the target area at group level.

```

0.5435    14
-1.3115    14
 1.3435    14

%LME Design:
fitlme(tbl,Parameter~1+(1|Subject),FitMethod,REML)

lme =

Linear mixed-effects model fit by REML

Model information:
Number of observations      34
Fixed effects coefficients    1
Random effects coefficients   6
Covariance parameters       2

Formula:
Parameter ~ 1 + (1 | Subject)

Model fit statistics:
AIC      BIC      LogLikelihood  Deviance
160.78   165.27   -77.389       154.78

Fixed effects coefficients (95% CIs):
Name      Estimate  SE      tStat    DF      pValue   Lower    Upper
{'(Intercept)'} -1.4476  0.91432 -1.5833  33     0.1229  -3.3078  0.41258

Random effects covariance parameters (95% CIs):
Group: Subject (6 Levels)
Name1      Name2      Type      Estimate  Lower    Upper
{'(Intercept)'} {'(Intercept)'} {'std'}    2.0593    0.9719  4.3631

Group: Error
Name      Estimate  Lower    Upper
{'Res Std'} 2.0768   1.5967  2.7013

```

Figure 6.18: **Design of the LME in Matlab.** A wide range of statistical information as well as an estimate for the fixed effects is returned.

### 6.6.6 Cross-validation

The cross-validation is carried out in several steps. In the first step, the  $\beta$ -estimates and the measured activations are read in for a single run of a subject. The original model value  $\vec{y}$  is calculated according to Equation 4.2. The model value  $\vec{y}_{\beta_{FB=0}}$  is determined by setting all  $\beta$ s that belong to a feedback regressor to zero and by calculating new model values using this modified weight vector and the design matrix. For the model value  $\vec{y}_{\beta_j-s_i}$  with generalized coefficients, an LME model is created with *fitlme*, but only the values of all other subjects are used. This group value is then used for the corresponding  $\beta$ -weights of the excluded subject. All weights that do not describe the feedback remain unchanged. The model value  $\vec{y}_{\beta_j-s_i}$  is again calculated by multiplying the new weight vector with the design matrix. With the help of these model values, the concordance correlation coefficients  $\rho_c(\vec{y}, \vec{y})$ ,  $\rho_c(\vec{y}, \vec{y}_{\beta_j-s_1})$ , and  $\rho_c(\vec{y}, \vec{y}_{\beta_{FB=0}})$  can be determined. This procedure is repeated for all runs of the subject.

In the second step, subject-specific correlation coefficients can then be calculated by forming the mean value of the correlation coefficients of the individual runs.

In the last step, the correlation coefficients for the entire group, which means the entire model approach, are determined by averaging the coefficients of the individual subjects and the p-Values of the differences are estimated.

# Results

## 7.1 Single run analysis

In a first step, the  $\beta$ -weights of the single runs are calculated and tested for significance. For the model approaches 3a)-c) the significance of the different valences (positive and negative feedback) are compared. This way, estimations can be made as to whether and, if so, which feedback has a stable influence on the activation of the sgACC. This could be useful information for future experiment designs.

Subsection 7.1.3. shows the time series of the ‘true’ activation, i.e. the measured values that have been corrected for the influence of the feedback. Here too, a comparison is made between the individual model approaches.

### 7.1.1 Significance of $\beta$ -estimates

For visualization, the p-values of the  $\beta$ -weights of the single runs are binned with a span of 0.05 (as shown in Figure 7.1). It is noticeable that most values fall between 0 and 0.05, and thus can be described as significant. This is particularly distinctive in the model approaches of modelling group 2, where up to 43% of the values, which describe the influence of the feedback, are below 0.05. In the first modelling group, this is around 24% of the values, similarly for all model approaches. In addition, other bins also contain up to 15% of the p-values (for example the group of p-values between 0.4 and 0.45 in Model 1a)). Regarding the percentage of significant  $\beta$ -weights for the influence of the feedback, no outstanding difference can be seen between the models for the first modelling group. In the second modelling group, the results for model approach 2a) and 2b) are similar, while in 2c) slightly more significant values were determined.

The parameter describing the feedback convolved with the temporal derivation of the HRF achieved a high proportion of significant weights. Both in Model 1b) and 2b), i.e., in both approaches in which the temporal derivation was considered separately, a clearly higher percentage of values is significant than in the convolution with the canonical HRF.

In Model 2b) this is even well over 50%, that is, there are more significant single run  $\beta$ -weights than non-significant ones.

In the Models 1c) and 2c) the convolution with both the canonical HRF and its time derivative have been combined into one parameter. This single parameter does, however, not represent the significance of the temporal derivative shown in Models 1b), 2b) and 3b).

### 7.1.2 Valence dependance of $\beta$ -estimates

For the comparison between positive and negative feedback in the Modells 3a)-c) histograms of the  $\beta$ -weights' p-Values are presented in Figure 7.2. As with the non-valence-separated models, the group of significant values with a p-value below 0.05 is the largest. The observations regarding the significance of the time derivative are only partially repeated in model 3b). While the time derivative parameter for the negative feedback values is with about 58% almost three times as high as the canonical HRF, the feedback parameter for positive feedback is unambiguously more often significant than its temporal derivation.

In all three model approaches, the p-values for the positive feedback remain much more often below 0.05 than those for the negative feedback. While this effect is not as strong in Model 3a), the percentage for positive feedback in Model 3b) is twice as high as for negative feedback, and in Model 3c) it is almost three time as high.

Model 3a) has the highest percentage of significant  $\beta$ -values, as 50% of the positive and 30% of the negative feedback weights have a p-value below 0.05. In the other models it lies about 40% and below 20% respectively.

### 7.1.3 Activation time series corrected for feedback influence

After the weights of the feedback influence have been estimated, the time series of the activation for the single runs can be corrected according to Equation 6.18. The influence of the feedback is subtracted from the measured activation. The original and the corrected time series are presented in the following four figures.

In Figure 7.3, the first 250 frames of an exemplary run from Group 1 are shown in black, as well as the corrected activation in red. A difference is clearly visible at the times when feedback was given (areas highlighted in grey).

On this large scale, there is no discernible difference between the models in Group 1, so in Figure 7.4, all models are shown together in one diagram, this time only for a portion of 35 frames. There is virtually no difference between Model 1a) and 1c) even in the enlarged view. However, a clear difference can be seen between these two models and Model 1b). It partly leads to a larger variation in the correction, which is usually expressed in a greater deviation from the original measurement time series, but in some cases also in a closer approximation than the other two.

Figure 7.5 shows an exemplary time series of Group 2. Compared to Group 1, a distinctly higher difference to the original time series can be seen in this zoom. In addition, there

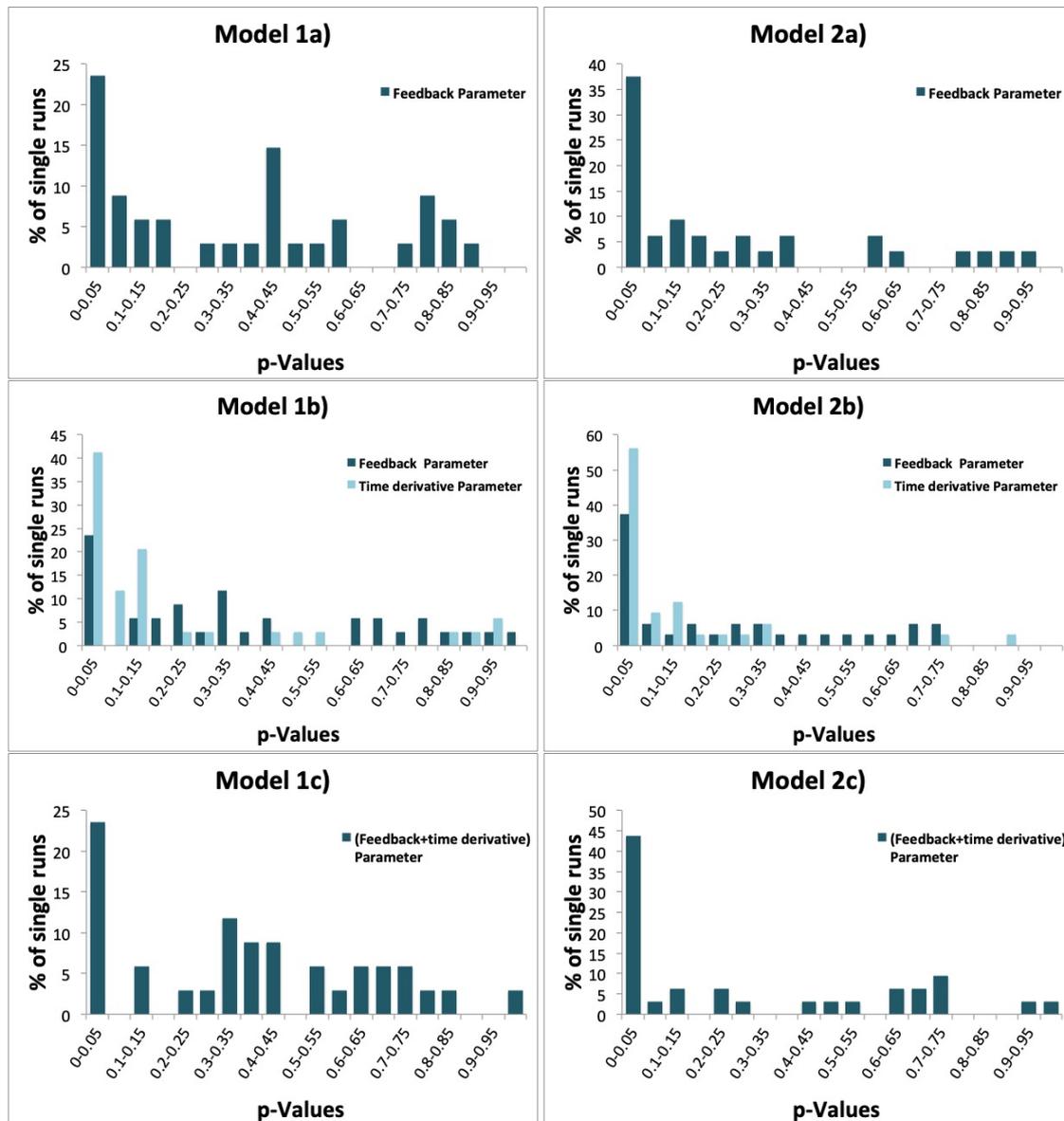


Figure 7.1: p-Values of the  $\beta$ -weights describing the influence of the feedback on the subgenual anterior cingulate cortex activation for the single runs. The values are divided into groups of span 0.05. In the Models 1b) and 2b) the darker colours denote the convolution of the feedback with the canonical HRF, the lighter colour refers to the convolution with the time derivative of the HRF. In the Models 1c) and 2c) these parameters are combined in the way described in Section 6.2.4.

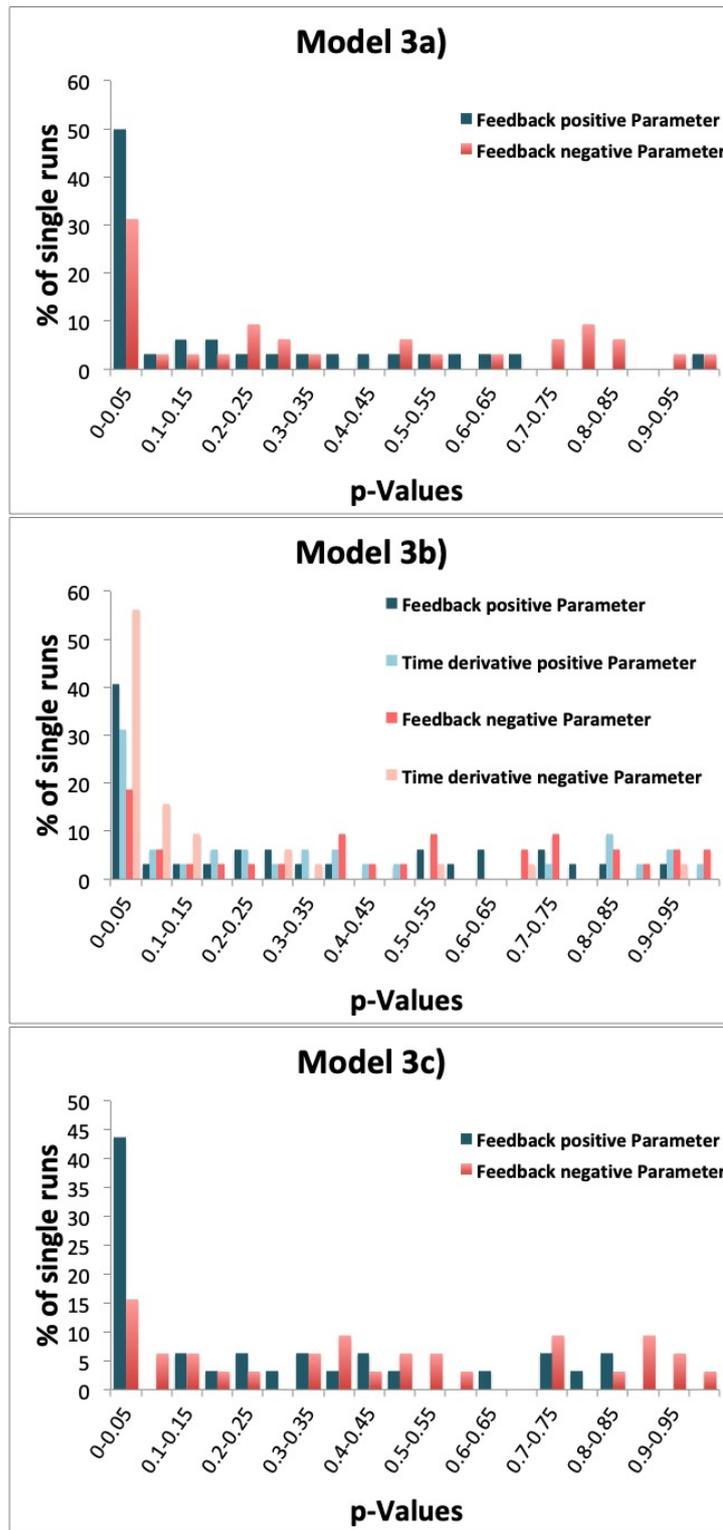


Figure 7.2: p-Values of the models that separate between positive (green) and negative (red) feedback. The values are again divided into groups of span 0.05. In Model 3b) the convolution of the feedback with the time derivative of the HRF is separately considered, its p-Values are shown in lighter colours than these of the convolution of the feedback with the canonical HRF. In Model 3c) they are again combined in the same way as in 1a) and 2a).

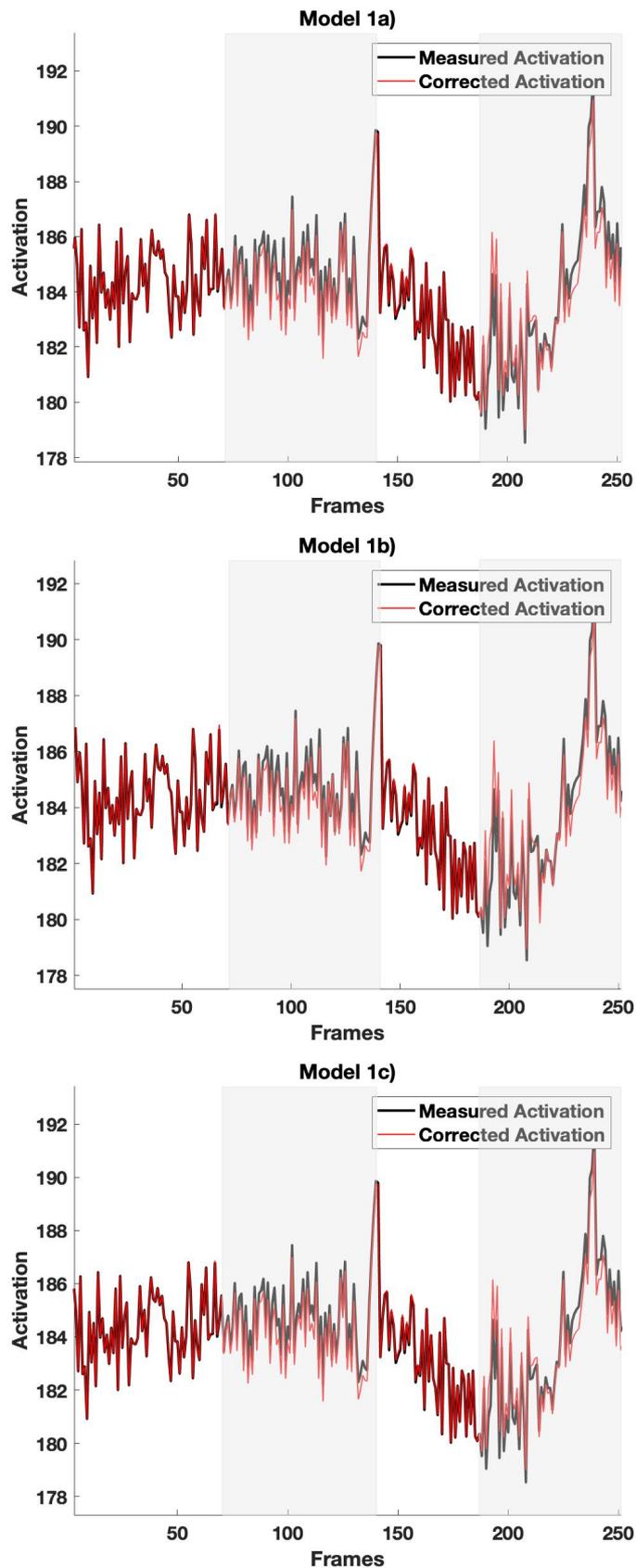


Figure 7.3: Original (black) and corrected (red) time series for the first subject of **Group 1**. On this scale, it looks as if the whole time series is shifted down and a difference between the model approaches cannot be observed.

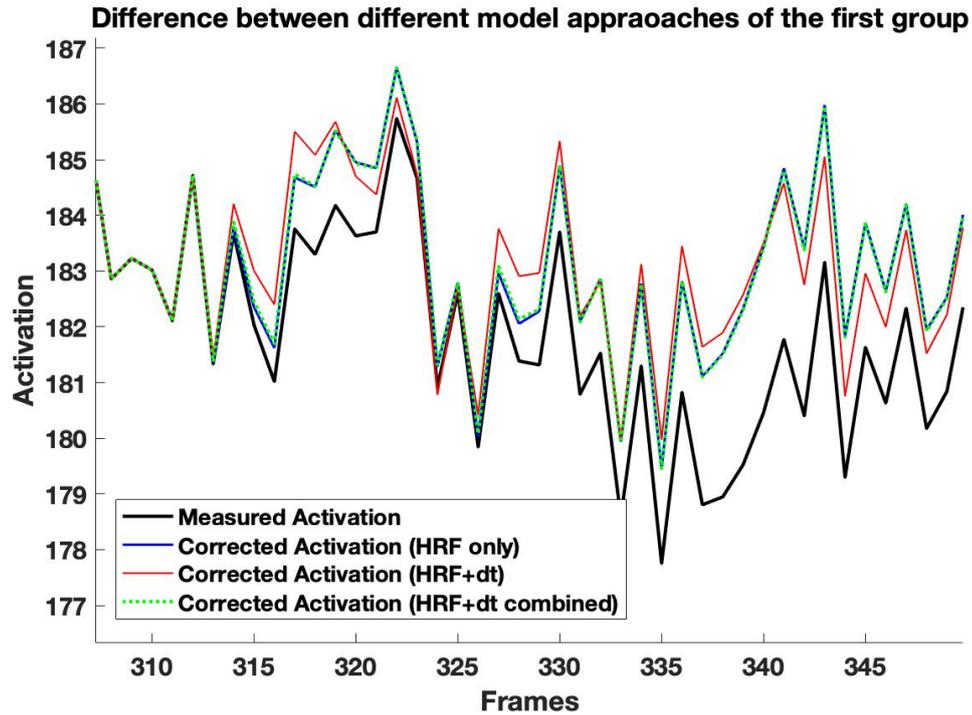


Figure 7.4: **Original (black) and all of the corrected time series in a larger view.** The difference between Model 1a) (blue) and 1c) (green) is again not visible. Model 1b) (red) leads to a different correction than the other two.

is no such clear spike amplification. Instead, entire portions have been shifted, as can be seen between frames 80 and 130, for example.

To compare the corrections of the single models, all of them are again combined in one diagram (Figure 7.6). As for the models in Group 1, approaches a) and c) cannot be clearly distinguished. This applies to both the combined- and the separate-valence models. Model b) also shows a deviation from the other two models here. A difference between Model 2 and 3 can be observed. The model with the separate valences approaches, Models a) and c), are more closely than in the non-separated case.

It is noticeable that the corrected time series are closer to each other than to the original measurement series. The differences between certain ways of correction are therefore not as big as the difference whether there is correction at all.

## 7.2 Group level analysis

In order to test the generalizability of the findings for the single runs, group results were calculated with all subjects using an LME model. The results of this analysis are listed in Figure 7.7. This table contains the fixed effects  $\bar{\beta}_j$  values, i.e. the cross-subject feedback

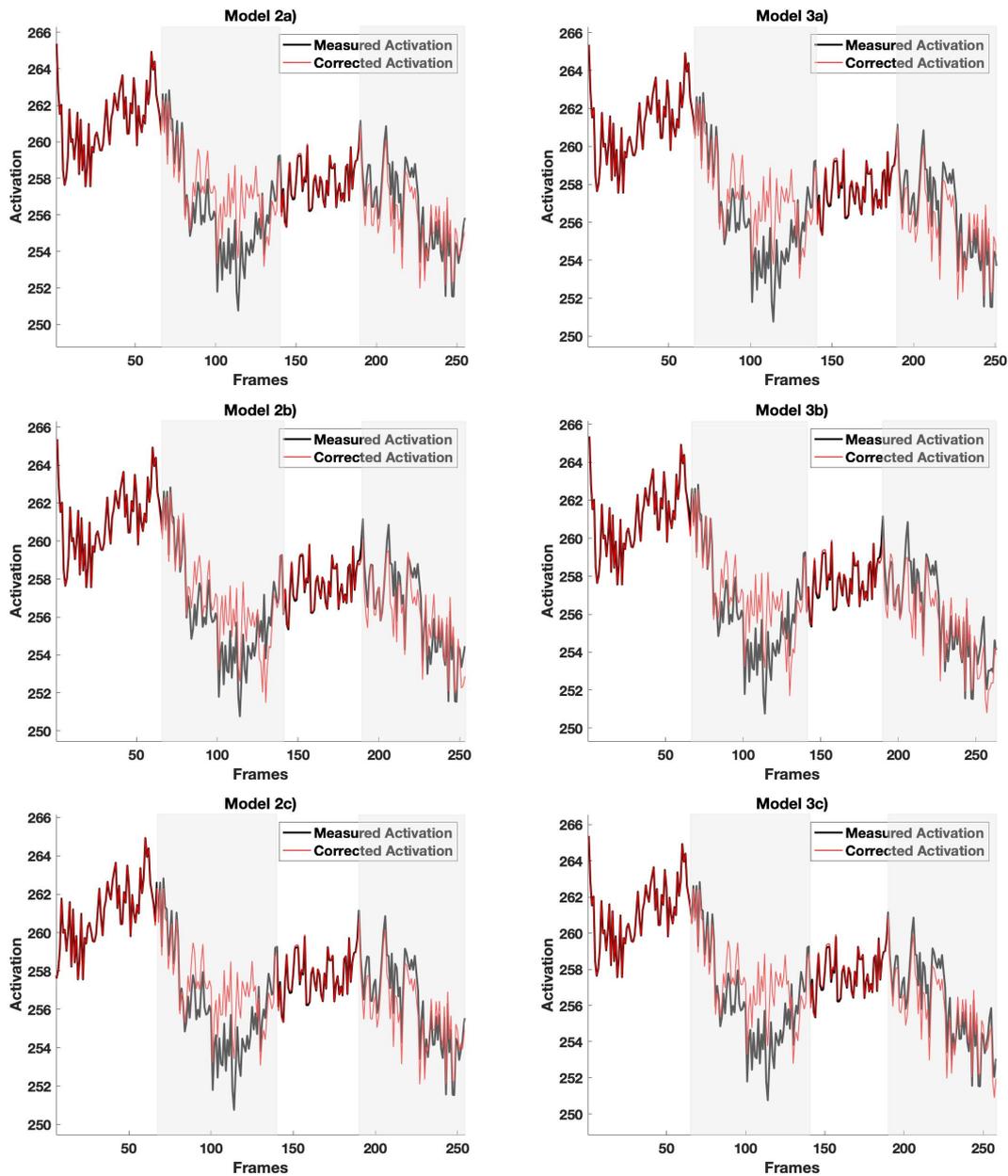


Figure 7.5: **Original (black) and corrected (red) time series for the fourth subject of Group 2.** Again whole areas are shifted by the correction, but compared to the models of the first group, the difference between correction and original data is clearly visible on a large scale. The grey areas denote all frames in which feedback was given. To compare the different model approaches, a zoomed in view is necessary.

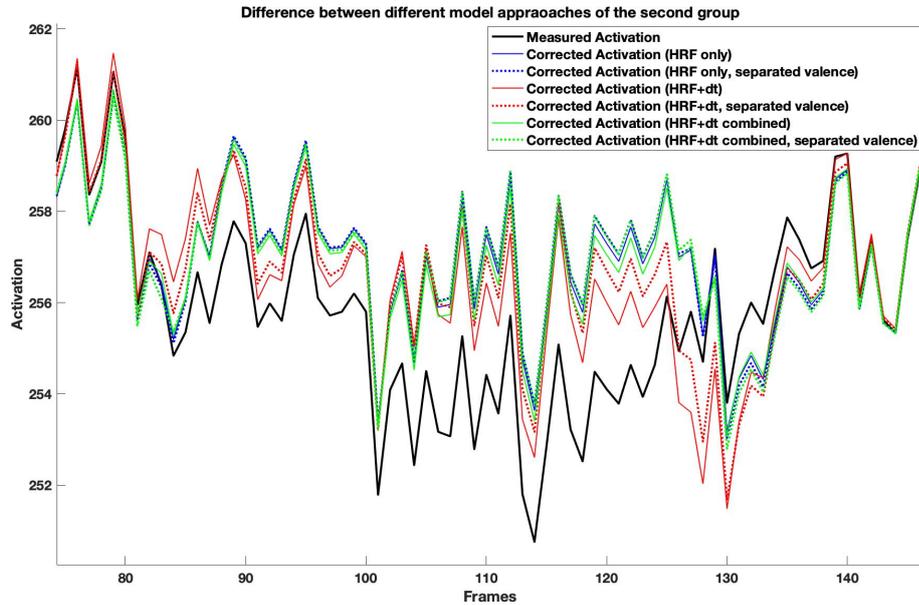


Figure 7.6: **Original (black) and corrected time series for all models of the second group.** Again, there is no decisive contrast between the Models a) and c) (blue and green). Model 2b) and 3b) shows differences to the other models as well as between models each other, i.e. whether positive and negative feedback was considered separately (dotted line). Variant 3b) approaches the other variants more closely than 2b).

weight of the  $j$ -th group after the subject-specific fluctuations were considered separately in the random effects. In addition, the respective standard error and the p-Values of the  $\bar{\beta}_j$  estimates are listed, which allow for significance statements.

Finally, the standard deviations of the random effects  $\sigma_u$ , i.e. the individual intercepts of the single subjects, as well as the AIC of the respective LME are given.

Since the aim of this thesis is not to find the ideal model for describing the feedback influence, but primarily to determine whether there are any differences between the models, a correction for multiple testing was neglected.

All calculated  $\bar{\beta}_j$ -values are negative because the task during the measurement was to reduce the activation of the sgACC. The better this succeeded, the more positive feedback was given. This results in an anti-parallel relation, the higher the activation, the lower the feedback, which makes negative weights necessary.

Six of the 16 estimated weights are significant, including two that describe the convolution of the feedback with the canonical HRF. All parameters describing a convolution of the feedback with the time derivative of the HRF are significant (Models 1b), 2b) and 3b) (for both valences)). With the combination of the two parameters in the models 1c), 2c), and 3c) on the other hand, no significance can be determined.

Model	Parameter	Estimate of fixed effect $\beta$	Standard error of $\beta$	p-Value of $\beta$	Standard deviation of random effects	AIC
1a)	HRF	-0.40323	0.27132	0.14672	0.62247	74.118
1b)	HRF	-0.06837	0.26909	0.80100	0.62229	70.153
	Time derivative of HRF	-0.66437	0.27074	0.00300	0.48036	52.183
1c)	HRF + time derivative combined	-0.22715	0.28314	0.42814	0.65819	70.987
2a)	HRF	-1.09730	0.48445	0.03065	1.12300	99.588
2b)	HRF	-0.72322	0.38424	0.06922	0.86089	96.996
	Time derivative of HRF	-1.08030	0.44039	0.02000	1.06310	60.052
2c)	HRF + time derivative combined	-1.00820	0.53861	0.07068	1.24880	106.090
3a)	HRF positive feedback	-0.74222	0.39590	0.07028	0.84906	108.290
	HRF negative feedback	-0.23068	0.09289	0.01862	0.17524	33.188
3b)	HRF positive feedback	-0.83712	0.41545	0.05265	0.91780	105.350
	Time derivative of positive feedback	-0.61246	0.23848	0.01530	0.49966	80.561
	HRF negative feedback	-0.01840	0.05705	0.74921	0.00046	27.322
	Time derivative of negative feedback	-0.88611	0.38632	0.02875	0.89120	87.653
3c)	HRF + time derivative combined positive	-0.94881	0.47806	0.05609	1.08890	105.590
	HRF + time derivative combined negative	-0.09223	0.08551	0.28908	0.13156	38.761

Figure 7.7: **Results of the LME for each model approach.** Significant fixed effects results are marked in red.

### 7.3 Cross-validation

In order to test the model approaches for their generalizability on the one hand, and, on the other hand, to make a comparison of the models in this regard, a cross-validation is carried out.

The concordance correlations  $\rho_c(\vec{y}, \hat{y})$ ,  $\rho_c(\vec{y}, \vec{y}_{\beta_j^{-s_i}})$ ,  $\rho_c(\vec{y}, \vec{y}_{\beta_{FB=0}})$  are determined for each model. The first value describes the correlation between the measured values and the previously estimated model values. The second value indicates the correlation between measured values and the generalized feedback influence, that is, weighted with the  $\beta$ -values of the other subjects, as described in Section 6.5. The last value determines the correlation between the measured values and a model in which all feedback influences are set to zero, i.e. in which it is assumed that the feedback has no influence at all. These results are shown in Figure 7.8.

Model	$\rho_c(\vec{y}, \vec{\hat{y}})$	$\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i})$	$\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_{FB=0}})$	p-Value of $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i}) - \rho_c(\vec{y}, \vec{\hat{y}})$	p-Value of $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i}) - \rho_c(\vec{y}, \vec{\hat{y}}_{\beta_{FB=0}})$
1a)	0.42 ( $\pm 0.13$ )	0.41 ( $\pm 0.13$ )	0.40 ( $\pm 0.12$ )	0.9829	0.9834
1b)	0.46 ( $\pm 0.14$ )	0.45 ( $\pm 0.13$ )	0.44 ( $\pm 0.13$ )	0.9655	0.9807
1c)	0.45 ( $\pm 0.13$ )	0.44 ( $\pm 0.13$ )	0.44 ( $\pm 0.13$ )	0.9823	0.9937
2a)	0.50 ( $\pm 0.17$ )	0.49 ( $\pm 0.15$ )	0.47 ( $\pm 0.15$ )	0.9562	0.9244
2b)	0.55 ( $\pm 0.15$ )	0.54 ( $\pm 0.13$ )	0.52 ( $\pm 0.13$ )	0.9602	0.9246
2c)	0.54 ( $\pm 0.15$ )	0.53 ( $\pm 0.12$ )	0.52 ( $\pm 0.12$ )	0.9527	0.9386
3a)	0.50 ( $\pm 0.17$ )	0.49 ( $\pm 0.15$ )	0.47 ( $\pm 0.15$ )	0.9441	0.9346
3b)	0.55 ( $\pm 0.15$ )	0.54 ( $\pm 0.13$ )	0.52 ( $\pm 0.13$ )	0.9446	0.9339
3c)	0.54 ( $\pm 0.15$ )	0.53 ( $\pm 0.12$ )	0.52 ( $\pm 0.12$ )	0.9383	0.9528

Figure 7.8: Correlation between the measured data, the run-specific model, the generalized model, and a model in which the feedback influence is set to zero. The values in brackets denote the standard deviation of the correlation coefficients. On the right side the p-Values of the contrasts between the correlations are presented.

To test whether there are significant differences between the correlations, the p-values of the contrast are calculated. This gives information about whether the difference is significantly different from zero. If the p-value is higher than 0.05 the measurement distributions are not considered to be different. The contrast between  $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i})$  and  $\rho_c(\vec{y}, \vec{\hat{y}})$ , as well as between  $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i})$  and  $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_{FB=0}})$  are examined. This is to test whether the generalized model is significantly different from the model with the run-specific parameters and whether the generalized model is significantly different from a model in which the feedback does not play a role. These results are also shown in Figure 7.8.

At first glance, the values for all three coefficients are very similar. This impression is confirmed by the p-Values. This means that for the modelling quality it makes no difference whether run-specific parameters are used for the feedback, generalized ones, or none at all.

To find out whether one of the models can generalize better than the others, p-Values of the contrasts of the  $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i})$  coefficients are calculated. These p-Values are listed in Figure 7.9. The rows indicate the model of the first  $\rho_c(\vec{y}, \vec{\hat{y}}_{\beta_j - s_i})$ , the columns the model of the second. It can easily be seen that all of the p-Values are well above 0.05, so no model is able to generalize significantly better than the others. Since none of the values was uncorrected significant, a correction for multiple testing has been omitted.

	1a)	1b)	1c)	2a)	2b)	2c)	3a)	3b)	3c)
1a)	1	0.8547	0.8726	0.7152	0.5213	0.5619	0.7188	0.5259	0.5702
1b)	0.8547	1	0.9818	0.8534	0.6445	0.6892	0.8571	0.6497	0.6982
1c)	0.8726	0.9818	1	0.8358	0.6286	0.6728	0.8396	0.6336	0.6817
2a)	0.7152	0.8534	0.8358	1	0.7856	0.8324	0.9962	0.7910	0.8418
2b)	0.5213	0.6445	0.6286	0.7856	1	0.9518	0.7820	0.9944	0.9423
2c)	0.5619	0.6892	0.6728	0.8324	0.9518	1	0.8287	0.9574	0.9904
3a)	0.7188	0.8571	0.8396	0.9962	0.7820	0.8287	1	0.7873	0.8381
3b)	0.5259	0.6497	0.6336	0.7910	0.9944	0.9574	0.7873	1	0.9478
3c)	0.5702	0.6982	0.6817	0.8418	0.9423	0.9904	0.8381	0.9478	1

Figure 7.9: **P-values of the contrasts between the  $\rho_c(\vec{y}, \vec{y}_{\beta_j - s_i})$ -coefficients of the different model approaches.** The rows denote the model of the first coefficient, the columns the model of the second one. Since all the p-values are well above 0.05 none of the models is able to generalize significantly different than the others.

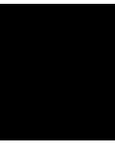
## 7.4 Test-retest reliability

In order to assess whether the calculated influence of the feedback is reliable, or whether the subject-specific fluctuation of this influence can also be expected in future studies, the subjects' intra-class correlation coefficient (ICC) is used. The ICCs of the single model approaches are calculated using the empirical between-subject variance  $s_b^2$  and the empirical total variance  $s_{tot}^2$  of the sample:

$$ICC = \frac{s_b^2}{s_{tot}^2} \quad (7.1)$$

and are presented in the first column of Figure 6.10. With one exception, the coefficients are all in the same range of magnitude between 0.03 (Model 3b) and 0.24 (Model 2c) and therefore have 'poor reliability' ( $ICC < 0.4$ ) according to the categorization by Cichetti and Sparrow [CS81]. The one salient value is obtained in Model 2b) for the time derivative parameter, and with a level of 0.42 it just falls into the 'fair reliability' category.





# Discussion

The aim of this work was to investigate whether the presentation of emotional feedback in the form of a schematic face with changing facial expressions can lead to additional activation or deactivation of the sgACC when used as target region in fMRI NF. It was also of interest whether a positive feedback (laughing face) has a different influence than a negative feedback (sad face).

Overall, the results show an inhomogeneous picture. Looking at the single runs, the feedback shows a significant influence in a decisive fraction of cases. The time series corrected for the feedback influence, especially in the subject group that received positive and negative feedback, show clear deviations from the original data. However, these findings could not be generalized.

Only the change in feedback over time still shows a significant group effect, likely because it also has the clearest influence on the individual level.

## 8.1 Single run analysis

It is noticeable that the bin of p-values between 0 and 0.05 (what would be considered significant for a single run) contains the highest percentage of  $\beta$ -weights in all models, while in all other intervals there are distinctly fewer single-run results. For the data underlying the current analyses, a significant influence of the feedback could be determined in a percentage between 15% and 57% of the runs, depending on the model approach. Thus, the feedback indeed seems to have a certain influence on the activation of the sgACC during a single run.

As can be seen from the calculation of the ICC (Figure 6.10), the presence and magnitude of this influence does not show any grouping with regard to sessions or runs. This speaks against any systematic symptoms such as fatigue, habituation, or other time-dependant reactions to the feedback. However, the weights for the feedback clearly vary between the

individual subjects. There are also strong fluctuations and outliers in the magnitude of the  $\beta$ -weights within the subjects, so there seem to be single runs, in which the feedback has a particularly strong or weak influence.

The individual reactions to the feedback might depend on various factors, for example, how sensitive the subjects react to feedback, how empathically they react to facial expressions, or how much they get involved in the task. This dependence on manifestations of personality traits [KS19] and also neuroanatomical constitutions, such as the relation of volumes between different brain areas [WEE20], can generally be observed in the success of neurofeedback training and also seems to play a role in this case.

For the within-subject fluctuations, the current condition as well as the chosen strategy and the satisfaction or annoyance about the feedback received on this strategy, could play a role besides general noise. The sgACC is also responsible for a number of other related tasks [GDK<sup>+</sup>15] [GKP<sup>+</sup>95] [MLB<sup>+</sup>99] [RD08], such as emotion reappraisal, decision-making, and processing of rewards. Since the subjects were required not to think of anything in particular during the baseline blocks these additional tasks are also conducted while receiving feedback. Thus, they could cause activation simultaneously with the perception of the emotional stimuli, which is also picked up by the feedback regressor and provides additional variability.

In the modelling approaches of Group 2, which received positive and negative feedback, a larger percentage of the results are significant than in those of Group 1. Since in Klöbl et al. [KMG<sup>+</sup>20] it has been concluded that this scheme also leads to a higher success in neurofeedback training and thus will be used in future studies, it is particularly important to take potential feedback influences on neural activation into account. However, between the models a), b) and c), i.e. in the different ways in which the time derivative is taken into account, there does not seem to be any particular qualitative difference in the frequency of significant results in any group.

Comparing the influence of the positive and negative feedback on the activation of the sgACC, it is primarily noticeable that the positive feedback shows a significant influence more often than the negative feedback. However, there are more significant  $\beta$ -weights in the time derivative of the negative feedback than in the positive case. This points towards the assumption that the positive feedback and the temporal change of the negative feedback could have a decisive influence on the second feedback loop. The amplitude of the negative feedback and the change in positive feedback, in contrary, have a lower influence. This conclusion supports the result that the feedback in the experimental scheme with positive and negative feedback has a greater influence than the presentation of only positive feedback [KMG<sup>+</sup>20]. Since the change in feedback generally shows a stronger influence on the activation than the level of feedback and this change in negative feedback is more often significant than in positive feedback, the additional presentation of negative feedback leads to a stronger activation.

The comparably high percentage of significant results for the weights that describe the convolution of the feedback with the time derivative of the HRF is particularly noticeable. It can be concluded from this observation that the sgACC is primarily activated by a change in the presented facial expressions and to a lesser extent by the absolute level of the feedback.

In contrast to other regions, such as the amygdala, the inferior frontal gyrus or the fusiform gyrus, the activation of the sgACC by dynamic facial expressions has not yet been confirmed in the literature [TFM09],[YS06], but it has been shown that a dynamic presentation generally improves the recognition of the emotional content [Fri53], [HHS99]. This preference can easily be explained by the fact that static emotional images perform poorly in imitating real situations from everyday life [Fri53]. Such increased attention to changes in the facial expressions of the smileys is therefore also conceivable for the sgACC. An activating effect of the feedback could presumably be prevented by a static presentation, for example, by giving intermittent feedback, i.e. a single feedback in the form of a static facial expression that reflects the performance of the entire block. However, following the stimulating influence of the feedback, the positive effects of social feedback might be comparably weak. Since the static facial expression of a schematic face would hardly be perceived as a real emotion, the naturalness and thus intuitiveness of the feedback would be lost. It therefore seems to be more useful to take the activation of the target region due to a changing feedback into account in the analysis by adding an additional regressor.

Another possible explanation why a significant activation of the sgACC was found related to the temporal change in feedback, can be derived from the works of Barrett and Simmons [BS15]. They suggest that the sgACC makes predictions about required physiological changes based on perceived changes in the environment, which then result in a top-down influence on the hypothalamus and the periaqueductal grey and trigger changes in the anticipation of physiological responses. This suggestion is consistent with earlier findings that the sgACC plays a central role in the modulation of physiological arousal (e.g. changes in heart rate or blood pressure) [BKFP00], [Sap02], [TC02]. Since the BOLD signal does not directly represent neural activity, but rather the hemodynamic response to it, it seems possible that the sgACC shows increased activation due to initiating a change in the response as a reaction to changed feedback. This assumption would also explain why a change in the negative feedback has a greater impact than a change in the positive feedback. The need to change something in the hemodynamic response is greater if the feedback is negative than if it is positive. Such an explanation of the observed reaction of the sgACC would underline Weiskopf's concerns that it is not certain whether neurofeedback actually modulates neuronal activity or only the blood flow through the corresponding target area [Wei12].

On the other hand, it could also be possible that the different reactions to positive and negative feedback are caused by the functional localizer. The sgACC is sensitive to both valences, but spatially separated [GSG<sup>+</sup>05] and the localizer used in the underlying study only included negative stimuli [HGH<sup>+</sup>11]. To gain clarity about this it would be useful

either to investigate the complete sgACC or only the part that is responsible for positive emotion processing.

The difference between the original and the corrected time series is distinctly more visible in Group 2, i.e. in the group that includes positive and negative feedback, and confirms that the feedback influence is larger when negative feedback is also presented. The first model group mainly shows a shift in the curve, i.e. all values are less negative after the measured data are corrected. This effect can be explained by the fact that the examined part of the sgACC is activated by negative emotions and deactivated by positive emotions, so the positive feedback leads to an additional deactivation. Since the aim of the training was also to deactivate the sgACC, the perception of the positive feedback leads to an over-estimation of the training successes. A simultaneous correction of the feedback influence would therefore lead to a lower feedback, since the sgACC would be deactivated less strongly.

In the second and third modelling groups, the correction shows an increase in the amplitude: At periods that generally show a lower activation (e.g. in Figure 7.6, frames 100-120), positive feedback is given, which results in a corrected time series shifted upwards, as described in Model 1 described. At periods that show a higher activation (e.g. Figure 7.6, frames 125-140), negative feedback is given. Since the perception of the sad smiley activates the sgACC and thereby shifts the curve upwards, the corrected curve is below the measurement values here. In summary, it can be stated that if the performance was good, the feedback was displayed too high, and if the performance was poor, the feedback was too negative.

Similar to the distribution of the p-Values of the weights, there is virtually no visible difference between the models a) and c). Apparently, it does not seem to make a difference in any model for the specific influence of the feedback whether the time derivative is not taken into account, as in approach a), or whether it is combined with the HRF convolution to form a single regressor. The findings of Calhoun et al. [CSPK04] that a model approach, in which the convolution with a canonical HRF and its time derivative are combined to a single regressor, improves the fit to the data, could not be confirmed in this work. However, the results achieved with modelling groups b) significantly deviate from the others. In the b) models, the influence of the feedback is also different, depending on whether the valences are examined separately or not. This difference may be based on the possibility, which has already become apparent, that the influence of the positive and negative feedback concerning the time derivative is different. In the event of a decision to implement one of the models, this only has to be made between models a) and b), since the c) models are no different from the a) ones.

## 8.2 Group level analysis and cross-validation

At the group level, the majority of model approaches have no significant  $\beta$ -weights, although the percentage of significant single run weights is quite high for all of them. A

possible explanation for this phenomenon can again be found in the strong inter-individual variability.

In each of the b) models, the time derivative regressor also has a significant influence on the activation of the target area at group level. This finding confirms the assumption that the change in the feedback over time really has a noticeable influence on the activation. This influence is strong enough to be maintained at group level despite subject-specific fluctuations.

In order to test the models for their generalizability outside of the examined data set, a cross-validation was carried out. The correlations between the measurement data and the individual model, the generalized model and a model in which the feedback does not play a role are virtually equal. Also in no case a significant difference between these three correlation coefficients can be found. Additionally, there seems to exist no difference between the single model approaches in terms of their correlation between the measurement data and the generalized model, so none of the models is able to generalize better than the others.

Since it does not seem to make any difference for the correlation with the measurement data whether the feedback is taken into account as a regressor or not, these findings suggest that the influence of the feedback on the activation of the sgACC might be negligible in magnitude. However, since the results of this work so far indicated a certain influence of the feedback and, above all, its change over time, other causes for the results of the cross-validation can be assumed. On the one hand, there is the possibility that, due to the strong individuality of both the measurement data and the found feedback influence, generalization is not possible and each case must be considered separately. In this case, a generalizing effect could not be confirmed by a cross-validation, either. Another explanation could be that there is a power problem, i.e. that the sample size with the six subjects per model approach is too small for a significant effect to be found in the correlation.

### 8.3 Reliability of the results

None of the modelling approaches shows a good test-retest reliability of the feedback influence. It therefore seems reasonable to conclude that there is a high degree of variability within the subjects, which makes up a decisive part of the total variance.

These results support the assumption that a strong within-subject variability leads to the fact that no significant results on the group level could be found. This poor generalization ability of the group parameters was confirmed by the cross-validation.

Again, the  $\beta$ -weight with the highest reliability is the one that describes the influence of the change in feedback over time. The previous finding that this time derivative has an explicit influence on the activation of the sgACC is again confirmed.

In comparison with the test-retest reliability of other fMRI measurement series, the values achieved in this case are not noticeably lower. In a meta study by Elliott et al. [EKI<sup>+</sup>20], 56% of all examined studies showed poor reliability. If only the studies are considered that investigated various aspects of emotion processing, and thus are comparable to the task in the study examined in this work, even 85% have an ICC below 0.4. This could be due to the fact that the ICC can be undermined by habituation effects and rapid fluctuations, especially in transient psychological states such as momentary emotions or thoughts [HMK17]. In addition, the small number of subjects and measurements per subject can cause that single outliers in the feedback influence have a particularly high influence. Thus both the between-subject and the within-subject variability are not representative of the phenomenon. This problem Elliott et al. [EKI<sup>+</sup>20] also named as one of the most important reasons for the low reliability in fMRI investigations.

## 8.4 Limitations

This work is subject to some limitations. The first possible source for an imprecise determination of the feedback influence is the modelling of the same as a linear function of the neural response. The fact that this linearity is not always clearly given [PMN11] and that the neural response itself does not always follow linearly on a stimulus, has also been established for other cases of sensory perception of a (more or less constant) stimulus over a longer period of time. For example, if a longer series of sounds were exposed to a subject, instead of sustained neural activity in the auditory system, phasic outbursts at the beginning and end of the series of measurements were found [HM02]. Neural adaptation to a recurring stimulus also seems to occur, especially if they are presented in intervals of less than two seconds [DB97]. Since such a rapid presentation of the feedback also occurs with neurofeedback if continuous feedback is given (in the study underlying this work the repetition time was also well under a second), a non-linear modelling of the feedback influence could lead to more precise results.

In the same context, there is the simplified assumption in the modelling that the influence of the feedback is constant from the start, and the proportionality factor  $r$  in Formula 6.18 is independent of time. In addition to short-term, neural adaptations during a single run, it is possible that longer-term, time-dependant factors such as fatigue, loss of concentration, or frustration influence the perception of the feedback. In this data set, the ICCs concerning *runs* and *sessions* did not indicate a trend regarding level and distribution of the  $\beta$ -coefficients, which describe the influence of the feedback, over the course of the two training sessions. So the assumption of the time-independent influence does not seem to lead to a decisive bias.

The error variance of the models could have been reduced by additionally orthogonalizing the feedback regressor with the sawtooth regressor of the single runs. Although this would not have had any general influence on the regression coefficients, the p-Values of the models could have changed. This orthogonalization has not been carried out in this work, since such a sawtooth regressor is typically not used and was needed only for the original investigations in Klöbl et al. [KMG<sup>+</sup>20]. However, it is possible that due to this

inaccuracy, any significance in the  $\beta$ -weights of the feedback was not taken into account or was incorrectly determined.

When examining the hierarchies in the data in the course of creating the linear mixed effects model, it should be noted that the number of different class elements was far too low to be able to make statistically meaningful statements. With a run number of three and a session number of only two, it is not possible to form an intra-class variance that enables statements to be made as to whether a run-specific or session-specific random effect exists. A larger number of measurements would be necessary to investigate this phenomenon further. However, random effects were primarily used as a correction for dependencies within the data.

In the models 1b), 2b), and 3b) the amount of the presented feedback was also convolved with the time derivative of the HRF and treated as an explicit regressor. The sign of the corresponding coefficient provides additional information about the time-to-peak of the underlying hemodynamic response. A positive coefficient indicates an earlier peak than expected, while a negative  $\beta$ -weight indicates a later peak than expected [Gro20]. Since all  $\beta$ s describing the time derivative are significantly negative, this leads to the assumption that the feedback may generally cause a delayed hemodynamic response compared to the canonical HRF used by default in SPM. In this case, it could be useful for further investigations to opt for another HRF model that offers the possibility of also modelling the time-to-peak [LWK<sup>+</sup>12].

## 8.5 Outlook

Since the feedback apparently has a significant influence on the activation of the sgACC in single runs, but this influence fluctuates very individually between subjects and measurements, it seems reasonable to include the influence of the feedback as an additional regressor during online analysis. This would have the advantage that the feedback in the next presentation step could be shown corrected by the previous feedback influence. The effect found in this thesis that good performance might be rewarded and bad performance punished too much, could be avoided in this way. Too strong punishment leads to frustration, but a too high reward can lead to the performance potential not being fully exploited. Overall, such corrected feedback should lead to improved learning outcomes. An additional implementation of regressors is possible in many software for neurofeedback, e.g. OpenNFT [KAP<sup>+</sup>17]. The next step here would be to test the influence on the processing time in the online analysis. In addition, future studies must show whether such a correction actually leads to improved learning outcomes.

A further move towards a personalized approach is particularly important, should a neurofeedback design similar to the one presented in this work be used as a therapeutic approach for affective disorders. Depressed patients show increased activity in the sgACC when processing emotional information [SSD11] and generally have a biased emotion processing compared to healthy subjects [BDP10]. On the one hand, the feedback might be adapted to the changed perception of facial expressions [EZDA11], as depressed patients often perceive friendly faces as neutral and neutral faces as sad. On the other hand, when

regressing out the influence via the second feedback loop, the used HRF must be varied in such a way that the longer processing time of emotional content in depressed people [GKYJ04][LRSK07][SDLM<sup>+</sup>04] is taken into account. However, additional optimizations are still necessary. Particularly in the case of mental disorders, individual differences are pronounced as a result of various factors, such as medication, the disease severity, etc. and require special attention in order to enable optimal therapeutic success.

An influence of the emotional feedback not only could exist in the sgACC, but also in other regions that are involved in emotion processing. This should be investigated in future studies as it might bias offline analyses. Whether intermittent feedback schemes would also benefit from such a correction, remains an open question. However, carry-over effects to the regulation periods appear unlikely if the separating baselines are chosen long enough.

The results of this work can contribute to the fact that more attention should generally be paid to an appropriate selection of feedback. In addition to a mathematical integration into the analysis software, there are other ways of avoiding a possible bias due to the given feedback. The respective properties of the target region should be taken into account. In the case of the sgACC, for example, using auditory feedback would be a possibility, since this region shows no activation in response to audibly received emotional stimuli [PWTL02]. Even if the advantage of visual versus auditory feedback [HNP<sup>+</sup>04] is not to be ignored, there is the possibility to rely on non-emotional stimuli. Many regions that are involved in emotion processing, for example the ACC or various areas of the PFC, are not activated by non-emotional images [THC<sup>+</sup>99]. The adaptation of the feedback to the respective target area is also advantageous for non-emotional areas of the brain, e.g. auditory or tactile feedback could be used for regions that react to dynamic images such as the inferior occipital gyrus [SKY<sup>+</sup>04]. However, in this case the benefits of using emotional feedback would also disappear, so careful consideration of the goal and framework is necessary.

## Conclusion

Following previous successful attempts to promote learning in NF by utilizing social reward, a recent study employed a smiley face as feedback mechanism for the sgACC – a brain region important in MDD. As the sgACC reacts to emotional stimuli, a second feedback loop, besides that of volitional control, might have been present. The emotional feedback that was presented during the neurofeedback training seems to have had a significant influence on the activation of the sgACC. However, this influence can primarily be observed in the analysis of single runs; the values vary highly between individual subjects. Thus, almost no significant effect could be found at a group level. The influence was mainly evident for the change of feedback over time, which was discernible at group level. Positive feedback seems to trigger additional activation, whereas negative feedback especially exerted influence on the sgACC through its temporal variation. After correcting the measurement series about the influence of the feedback, there was the tendency visible that good performance was rewarded too strongly, while poor performance was punished more than would have been appropriate.

Since significant influences can be found, but vary highly between individual subjects, it seems reasonable to correct for them during the NF training, which could lead to improved learning outcomes. Such an individual approach is particularly necessary when dealing with patients suffering from mental disorders, as individuality is reinforced by various factors, such as the severity of the illness or medication. Future studies must show whether such an integration of the feedback influence actually leads to improved learning and whether the findings of this work also translate to other brain regions that are involved in emotion processing.

# List of Figures

2.1	Entries available on GoogleScholar up to the respective year for the keyword "Neurofeedback". . . . .	7
2.2	Brain areas activated by different tasks concerning language [PMN11]. . .	9
2.3	Classic simple feedback schemes. . . . .	12
2.4	Feedback design that has a content-related connection to the trained brain region. . . . .	13
2.5	Social feedback using changing facial expressions of a human avatar [MAK <sup>+</sup> 15].	14
2.6	Regulation of a game character as feedback. . . . .	14
2.7	Protons act like little magnets because of their spin property [Goe07]. . .	16
2.8	An electromagnetic pulse with the according resonance frequency leads to an excitement of the proton spins [Goe07]. . . . .	17
2.9	The signal amplitude of the measured MRI signal decays exponentially with the time constant $T_2^*$ [Goe07]. . . . .	18
2.10	The process of the dephasing can be reversed [Goe07]. . . . .	18
2.11	The principle of frequency encoding using glasses of water with different heights as example [Goe07]. . . . .	19
2.12	From neuronal activity to BOLD signal [Goe07]. . . . .	21
2.13	The hemodynamic response to neuronal activation [PMN11]. . . . .	22
3.1	Mood-inducing pictures with different facial expressions. . . . .	28
3.2	Regions playing a role in emotions, in lateral, anterior and posterior view.	30
3.3	Neuronal correlates of the single emotions <i>happiness</i> , <i>sadness</i> , <i>disgust</i> , <i>fear</i> , and <i>anger</i> in a meta-analysis of Phan et al. [PWTL02]. . . . .	31
3.4	Results of the meta-analysis of Phan et al. [PWTL02] concerning cognitive participation in emotion processing. . . . .	35
3.5	The subgenual part of the ACC. . . . .	37
4.1	Correlation between two-dimensional value pairs. [Moo11] . . . . .	40
4.2	Example of a multidimensional regression. . . . .	41
4.3	Model errors called 'residuals' for a linear regression in two dimensions. .	42
4.4	Error distribution of the sample. [Moo11] . . . . .	44
4.5	T-distributions for different degrees of freedom. . . . .	44
4.6	T-distribution with critical t-value. . . . .	45
4.7	Exemplary time series of the activation of a voxel measured with fMRI and the condition regressor . . . . .	47
4.8	Complete pooling versus no pooling. . . . .	49
5.1	Schematic structure of the study design in Mathiak et al.[MAK <sup>+</sup> 15] . . .	59
5.2	Difference in regulation success between the two feedback designs in Mathiak et al. [MAK <sup>+</sup> 15] . . . . .	59

5.3	Reaction of the virtual agent depending on the expressed engagement of the subject. [APC <sup>+</sup> 16]	61
5.4	Session design in Aranyi et al.[APC <sup>+</sup> 16]	61
5.5	Exemplary design of a functional localizer run in Direito et al. [DLS <sup>+</sup> 19]	63
5.6	Design of a NF run in Direito et al. [DLS <sup>+</sup> 19]	63
5.7	Feedback scheme, Functional localizer and target area in Hamilton et al. [HGH <sup>+</sup> 11]	65
5.8	Exemplary model design of a single neurofeedback run in Klöbl et al. [KMG <sup>+</sup> 20]	66
6.1	Rigid-body transformation [WKH07]	71
6.2	Spatial smoothing. [WKH07]	71
6.3	Shape of the canonical HRF with different amplitudes	74
6.4	Time course of canonical HRF and its time derivative.	76
6.5	Numbering of the different approaches investigated in this work.	76
6.6	Four different regulation regressors are used to model the activation timeline	78
6.7	Raw feedback time series and convolution with HRF.	79
6.8	Correlation between the three regressors before and after orthogonalization.	80
6.9	Distribution of the residuals before and after the integration of additional regressors.	82
6.10	ICC values of the groupings into six different subjects, three different runs and two different sessions for each model approach	83
6.11	Clustering of the $\beta$ -estimates according to the subject, run, and session they are estimated from.	84
6.12	Graphical user interface of Matlab during loading and storing of the measurement data.	87
6.13	Matlab Editor with the script defining the parameters for SPM.	88
6.14	Separation of feedback time series.	90
6.15	Matlab's command window during convolution of the feedback regressors with the HRF and its time derivative.	90
6.16	Output of Matlab's linear fitting function <i>fitlm</i> for one subject of model approach 1a).	92
6.17	Creation of a tyble in Matlab with all necessary parameters for the LME design.	93
6.18	Design of the LME in Matlab.	94
7.1	p-Values of the $\beta$ -weights describing the influence of the feedback on the sgACC activation for the single runs.	97
7.2	p-Values of the models that separate between positive (green) and negative (red) feedback.	98
7.3	Original (black) and corrected (red) time series for the first subject of Group 1.	99
7.4	Original (black) and all of the corrected time series in a larger view.	100
7.5	Original (black) and corrected (red) time series for the fourth subject of Group 2.	101

7.6	Original (black) and corrected time series for all models of the second group.	102
7.7	Results of the LME for each model approach. . . . .	103
7.8	Correlation between the measured data, the run-specific model, the generalized model, and a model in which the feedback influence is set to zero. . . . .	104
7.9	p-Values of the contrasts between the $\rho_c(\vec{y}, \vec{y}_{\beta_j}^{s_i})$ -coefficients of the different model approaches. . . . .	105

# Bibliography

- [ADFS16] T Auer, WI Dewiputri, J Frahn, and R Schweizer. Higher-order brain areas associated with real-time functional MRI neurofeedback training of the somato-motor cortex. *Neuroscience*, 378:22–33, May 2016.
- [AdRS<sup>+</sup>09] M Arns, S de Ridder, U Strehl, M Breteler, and A Coenen. Efficacy of neurofeedback treatment in ADHD: the effects of inattention, impulsivity and hyperactivity: a meta-analysis. *Clinical EEG and Neuroscience*, 40(3):180–89, July 2009.
- [AFAC08] NT Alves, SS Fukusima, and JA Aznar-Casanova. Models of brain asymmetry in emotional processing. *Psychology and Neuroscience*, 1(1):63–66, May 2008.
- [Agg92] JP Aggleton, editor. *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction*, pages 355–377. Wiley-Liss, New York, 1992.
- [APC<sup>+</sup>16] G Aranyi, F Pecune, F Charles, C Pelachaud, and M Cavazza. Affective interaction with a virtual character through an fNIRS Brain-Computer Interface. *Frontiers in Computational Neuroscience*, 10:70, July 2016.
- [ASK<sup>+</sup>16] K Amano, K Shibata, M Kawato, Y Sasaki, and T Watanabe. Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. *Current biology*, 26(14):1861–66, July 2016.
- [AST78] LY Abramson, ME Seligman, and JD Teasdale. Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology*, 87(1):49–74, February 1978.
- [ATD03] R Adolphs, D Tranel, and AR Damasio. Dissociable neural systems for recognizing emotions. *Brain and Cognition*, 52(1):61–9, June 2003.
- [Ban97] A Bandura. *Self Efficacy: The Exercise of Control*. Worth Publishers, New York, 1997.

- [BC07] N Birbaumer and LG Cohen. Brain-computer interfaces: Communication and restoration of movement in paralysis. *Journal of Physiology*, 579(3):621–35, March 2007.
- [BDDA94] A Bechara, AR Damasio, H Damasio, and SW Anderson. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1-3):7–15, April-June 1994.
- [BDP10] C Bourke, K Douglas, and R Porter. Processing of facial emotion expression in major depression: a review. *Australian and New Zealand Journal of Psychiatry*, 44(8):681–96, August 2010.
- [BFF09] AR Brunoni, R Fraguas, and F Fregni. Pharmacological and combined interventions for the acute depressive episode: Focus on efficacy and tolerability. *Therapeutics and clinical risk management*, 5:897–910, November 2009.
- [Bir06] N Birbaumer. Breaking the silence: Brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43:517–532, November 2006.
- [BKFP00] R Bandler, KA Keay, N Floyd, and J Price. Central circuits mediating patterned autonomic activity during active vs. passive emotional coping. *Brain Research Bulletin*, 53(1):95–104, September 2000.
- [BMCN08] AD Boes, LM McCormick, WH Coryell, and P Nopoulos. Rostral anterior cingulate cortex volume correlates with depressed mood in normal healthy children. *Biological Psychiatry*, 63(4):391–97, February 2008.
- [BMF<sup>+</sup>99] RJ Blair, JS Morris, CD Frith, DI Perrett, and RJ Dolan. Dissociable neural responses to facial expressions of sadness and anger. *Brain*, 122:883–893, 1999.
- [BMSD04] B Brierley, N Medford, P Shaw, and AS David. Emotional memory and perception in temporal lobectomy patients with amygdala damage. *Journal of Neurology, Neurosurgery and Psychiatry*, 75(4):593–9, April 2004.
- [Bow81] GH Bower. Mood and memory. *American Psychologist*, 36(2):129–148, 1981.
- [Bra97] DH Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4):433–6, 1997.
- [BRD<sup>+</sup>02] KN Botteron, ME Raichle, WC Drevets, AC Heath, and RD Todd. Volumetric reduction in left subgenual prefrontal cortex in early onset depression. *Biological Psychiatry*, 51(4):342–44, February 2002.

- 
- [BRSE87] AT Beck, AJ Rush, BF Shaw, and G Emery. *Cognitive Therapy of Depression*. Guilford Press, New York, 1987.
- [BS15] LF Barrett and WK Simmons. Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7):419–429, July 2015.
- [BSO07] S Bray, S Shimojo, and JP O’Doherty. Direct instrumental conditioning of neural activity using functional magnetic resonance imaging-derived reward feedback. *Journal of Neuroscience*, 27(28):7498–507, July 2007.
- [Bud09] T Budzynski. *Introduction to quantitative EEG and neurofeedback: Advanced theory and applications*. Elsevier: Academic Press, Amsterdam, 2nd edition, 2009.
- [Cal03] AJ Calder. Disgust discussed. *Annals of Neurology*, 53(4):427–8, April 2003.
- [CDP<sup>+</sup>00] H Critchley, E Daly, M Phillips, M Brammer, E Bullmore, S Williams, T Van Amelsvoort, D Robertson, A David, and D Murphy. Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. *Human Brain Mapping*, 9(2):93–105, February 2000.
- [CFWA<sup>+</sup>01] B Crespo-Facorro, AK Wiser, NC Andreasen, DS O’Leary, GL Watkins, LL Boles Ponto, and RD Hichwa. Neural basis of novel and well-learned recognition memory in schizophrenia: a positron emission tomography study. *Human Brain Mapping*, 12(4):219–31, April 2001.
- [CGT<sup>+</sup>10] TF Collura, J Guan, J Tarrant, J Bailey, and F Starr. EEG biofeedback case studies using live z-score training and a normative database. *Journal of Neurotherapy*, 14(1):22–46, February 2010.
- [CHZ<sup>+</sup>09] A Caceres, DL Hall, FO Zelaya, SCR Williams, and MA Mehta. Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage*, 45:758–768, April 2009.
- [CJH95] RW Cox, A Jesmanowicz, and JS Hyde. Real-time functional magnetic resonance imaging. *Magnetic Resonance in Medicine*, 33:230–236, February 1995.
- [CMD<sup>+</sup>15] JS Cordes, KA Mathiak, M Dyck, EM Alawi, TJ Gaber, FD Zepf, M Klasen, M Zvyagintsev, MC Gur, and K Mathiak. Cognitive and neural strategies during control of the anterior cingulate cortex by fMRI neurofeedback in patients with schizophrenia. *Frontiers in Behavioral Neuroscience*, 9:169, June 2015.

- [CS81] DV Cicchetti and SA Sparrow. Developing criteria for establishing inter-rater reliability for specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2):127–37, September 1981.
- [CSPK04] VD Calhoun, MC Stevens, GD Pearlson, and KA Kiehl. fMRI analysis with the general linear model: Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *NeuroImage*, 22:252–257, May 2004.
- [CYR+96] AJ Calder, AW Young, D Rowland, DT Perrett, JR Hodges, and NL Etcoff. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, 13(5):699–745, 1996.
- [CZD+01] T Canli, Z Zhao, JE Desmond, E Kang, J Gross, and JD Gabrieli. An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behavioral Neuroscience*, 115(1):33–42, February 2001.
- [DA13] WI Dewiputri and T Auer. Functional magnetic resonance imaging (fMRI) neurofeedback: Implementations and application. *Malaysian Journal of Medical Science*, 20(5):5–15, October 2013.
- [DB97] AM Dale and RL Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5(5):329–40, 1997.
- [DDOB12] AM Dias, AM Van Deusen, E Oda, and MR Bonfirm. Clinical efficacy of a new automated hemoencephalographic neurofeedback protocol. *Spanish Journal of Psychology*, 15(3):930–41, November 2012.
- [deC07] RC deCharms. Reading and controlling human brain activation using real-time functional magnetic resonance imaging. *Trends in cognitive science*, 11(11):473–81, November 2007.
- [DEYH05] HA Demaree, DE Everhart, EA Youngstrom, and DW Harrison. Brain lateralization of emotional processing: Historical roots and a future incorporating 'dominance'. *Behavioral and Cognitive Neuroscience Review*, 4(1):3–20, March 2005.
- [DGB+00] AR Damasio, TJ Grabowski, A Bechara, H Damasio, LLB Ponto, J Parvizi, and RD Hichwa. Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature neuroscience*, 3:1049–1056, October 2000.
- [DH95] RJ Davidson and K Hughdahl. *Brain Asymmetry*, chapter 'Cerebral asymmetry, emotion and affective style', pages 361–387. The MIT Press, Cambridge, 1995.

- 
- [DLS<sup>+</sup>19] B Direito, J Lima, M Simoes, A Sayal, T Sousa, M Lührs, C Ferreira, and M Castelo-Branco. Targeting dynamic facial processing mechanisms in superior temporal sulcus using a novel fMRI neurofeedback target. *Neuroscience*, 406:97–108, May 2019.
- [dMG<sup>+</sup>05] RC deCharms, F Maeda, GH Glover, D Ludlow, JM Pauly, D Soneji, JDE Gabrieli, and SC Mackey. Control over brain activation and pain learned by using real-time functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51):18626–31, December 2005.
- [DMRB13] G Dragoi, PR Melinte, L Radu, and O Buda. The paradigm of cingulate cortex globalization. psychopathologic and forensic implications. *Romanian Journal of Legal Medicine*, 21:55–66, 2013.
- [Dol02] RJ Dolan. Emotion, cognition, and behavior. *Science*, 298(5596):1191–94, November 2002.
- [DP05] WC Drevets and JL Price. *Biology of Depression: From Novel Insights to Therapeutic Strategies*, chapter 'Neuroimaging and Neuropathological Studies of Mood Disorders', pages 427 – 465. John Wiley and Sons, Ltd, New York, January 2005.
- [DPJ<sup>+</sup>97] WC Drevets, JL Price, JR Simpson Jr., RD Todd, T Reich, M Vannier, and ME Raichle. Subgenual prefrontal cortex abnormalities in mood disorders. *Nature*, 386(6627):824–7, April 1997.
- [DPNP02] RJ Davidson, D Pizzagalli, JB Nitschke, and K Putnam. Depression: perspectives from affective neuroscience. *Annual Review of Psychology*, 53:545–74, 2002.
- [DRB<sup>+</sup>07] WC Drevets, N Ryan, W Bogers, B Birmaher, D Axelson, and R Dahl. Subgenual prefrontal cortex volume decreased in healthy humans at high family risk for mood disorders. Abstract presented at Annual Meeting of the Society for Neuroscience, 2007.
- [DST08] WC Drevets, J Savitz, and M Trimble. The subgenual anterior cingulate cortex in mood disorders. *CNS Spectrums*, 13(8):663–81, August 2008.
- [DTTC17] ML Dixon, R Thiruchselvam, R Todd, and K Christoff. Emotion and the prefrontal cortex: an integrative review. *Psychological Bulletin*, 143(10):1033–1081, October 2017.
- [EAM11] C Escalano, M Aguilar, and J Minguez. EEG-based upper alpha neurofeedback training improves working memory performance. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2327–2330, 2011.

- [EKG<sup>+</sup>03] S Erk, M Kiefer, J Grothe, AP Wunderlich, M Spitzer, and H Walter. Emotional context modulates subsequent memory effect. *Neuroimage*, 18(2):439–47, February 2003.
- [EKI<sup>+</sup>20] ML Elliott, AR Knodt, D Ireland, ML Morris, R Poulton, S Ramrakha, ML Sison, TE Moffitt, A Caspi, and AR Hariri. What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, 31(7):792–806, July 2020.
- [EKS<sup>+</sup>16] K Emmert, R Kopel, J Sulzer, AB Brühl, BD Berman, DEJ Linden, SG Horowitz, M Breimhorst, A Caria, S Frank, S Johnston, Z Long, C Paret, F Robineau, R Veit, A Bartsch, CF Beckmann, D Van de Ville, and S Haller. Meta-analysis of real-time fMRI neurofeedback studies using individual participant data: How is brain regulation mediated. *Neuroimage*, 124:806–812, January 2016.
- [ENGGC<sup>+</sup>14] C Escalano, M Navarro-Gil, J Garcia-Campayo, M Congedo, D De Ridder, and J Minguez. A controlled study on the cognitive effect of alpha neurofeedback training in patients with major depressive disorder. *Frontiers in Behavioral Neuroscience*, 8:296, September 2014.
- [ET11] WA Edmonds and G Tenenbaum. *Case studies in applied psychophysiology: Neurofeedback and biofeedback trainings for advances in human performance*. John Wiley and Sons, New York, 2011.
- [EZDA11] R Elliott, R Zahn, JFW Deakin, and IM Anderson. Affective cognition and its disruption in mood disorders. *Neuropsychopharmacology*, 36(1):153–82, January 2011.
- [FADT11] JS Feinstein, R Adolphs, A Damasio, and D Tranel. The human amygdala and the induction and experience of fear. *Current Biology*, 21(1):34–8, January 2011.
- [FFJ<sup>+</sup>98] KJ Friston, P Fletcher, O Josephs, A Holmes, MD Rugg, and R Turner. Event-related fMRI: Characterizing differential responses. *NeuroImage*, 7(1):30–40, January 1998.
- [FGA<sup>+</sup>08] T Fernandez, F Garcia, RAP Alcala, E Santiago, and H Belmont. Positive vs. negative reinforcement in neurofeedback applied to learning disabled children. *Clinical Neurophysiology*, 119(9), September 2008.
- [Fis12] RA Fisher. An absolute criterion for fitting frequency curves. *Messenger of Math*, 41:155–160, 1912.
- [FLP<sup>+</sup>12] S Frank, S Lee, H Preissl, B Schultes, N Birbaumer, and R Veit. The obese brain athlete: Self-regulation of the anterior insula in adiposity. *PLoS One*, 7(8):e42570, August 2012.

- [Fri53] NH Frijda. The understanding of facial expression of emotion. *Acta Psychologica*, 9:294–362, 1953.
- [FWSN14] EVC Friedrich, G Wood, R Scherer, and C Neuper. Mind over brain, brain over mind: Cognitive causes and consequences of controlling brain activity. *Frontiers in Human Neuroscience*, 8:348, May 2014.
- [GB09] M Gendron and LF Barrett. Reconstructing the past: a century of ideas about emotion in psychology. *Emotion Review*, 1(4):316–339, October 2009.
- [GDK<sup>+</sup>15] M Groene, M Dyck, Y Koush, S Bergert, KA Mathiak, EM Alawi, M Elliott, and K Mathiak. Upregulation of the rostral anterior cingulate cortex can alter the perception of emotions: fMRI-based neurofeedback at 3 and 7 T. *Brain Topographie*, 28(2):197–207, March 2015.
- [GFR<sup>+</sup>16] MI Gerin, H Fichtenholz, A Roy, CJ Walsh, JH Krystal, S Southwick, and M Hampson. Real-time fMRI neurofeedback with war veterans with chronic ptsd: A feasibility study. *Frontiers in Psychiatry*, 7:111, June 2016.
- [GH07] A Gelman and J Hill. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, New York, 2007.
- [GKP<sup>+</sup>95] MS George, TA Ketter, PI Parekh, HB Horwitz, P Herscovitch, and RM Post. Brain activity during transient sadness and happiness in healthy women. *American Journal of Psychiatry*, 152(3):341–351, March 1995.
- [GKP<sup>+</sup>97] MS George, TA Ketter, PI Parekh, N Rosinsky, HA Ring, PJ Pazzaglia, LB Marangell, AM Callahan, and RM Post. Blunted left cingulate activation in mood disorder subjects during a response interference task (the stroop). *Journal of Neuropsychiatry and Clinical Neuroscience*, 9(1):55–63, Winter 1997.
- [GKYJ04] IH Gotlib, E Krasnoperova, D Neubauer Yue, and J Joormann. Attentional biases for negative interpersonal stimuli in clinical depression. *Journal of abnormal psychology*, 113(1):121–35, February 2004.
- [Goe07] R Goebel. *Clinical Functional MRI: Presurgical functional neuroimaging*, chapter 'Localization of Brain Activity using Functional Magnetic Resonance Imaging', pages 11–21. Springer Verlag, Heidelberg, 2007.
- [GPS72] GV Glass, PD Peckham, and JR Sanders. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42(3):237–288, 1972.

- [GRM<sup>+</sup>12] S Green, MA Lambon Ralph, J Moll, JF Deakin, and R Zahn. Guilt-selective functional disconnection of anterior temporal and subgenual cortices in major depressive disorder. *JAMA Psychiatry*, 69(10):1014–21, October 2012.
- [Gro20] The FIL Methods Group. *SPM12 Manual*. Functional Imaging Laboratory, Wellcome Centre for Human Neuroimaging, London, 2020.
- [GSG<sup>+</sup>05] IH Gotlib, H Sivers, JHE Gabrieli, S Whitfield-Gabrieli, P Goldin, KL Minor, and T Canli. Subgenual anterior cingulate activation to valenced emotional stimuli in major depression. *Neuroreport*, 16(16):1731–4, November 2005.
- [GSK04] R Goebel, B Sorger, and J Kaiser. BOLD brain pong: Self-regulation of local brain activity during synchronously scanned, interacting subjects. Presented at: 34th Annual Meeting of the Society for Neuroscience, 2004.
- [GST<sup>+</sup>02] RC Gur, L Schroeder, T Turner, C McGrath, RM Chan, BI Turetsky, D Alsop, J Maldjian, and RE Gur. Brain activation during facial emotion processing. *NeuroImage*, 16(3.1):651–62, July 2002.
- [GTPS<sup>+</sup>01] ML Gorno-Tempini, S Pradelli, M Serafini, G Pagnoni, P Baraldi, C Porro, R Nicoletti, C Umita, and P Nichelli. Explicit and incidental facial expression processing: an fMRI study. *NeuroImage*, 14(2):465–73, August 2001.
- [HGB<sup>+</sup>16] JP Hamilton, GH Glover, E Bagarinao, C Chang, S Mackey, MD Sacchet, and IH Gotlib. Effects of salience-network-node neurofeedback training on affective biases in major depressive disorder. *Psychiatry research. Neuroimaging*, 249:91–96, March 2016.
- [HGH<sup>+</sup>11] JP Hamilton, GH Glover, JJ Hsu, RF Johnson, and IH Gotlib. Modulation of subgenual anterior cingulate cortex activity with real-time neurofeedback. *Human Brain Mapping*, 32(1):22–31, January 2011.
- [HHS99] NK Harwood, LJ Hall, and AJ Shinkfield. Recognition of facial emotional expressions from moving and static display by individuals with mental retardation. *American Journal of Mental Retardation*, 104(3):270–278, May 1999.
- [HKK<sup>+</sup>05] U Habel, M Klein, T Kellermann, NJ Shah, and F Schneider. Same or different? Neural correlates of happy and sad mood in healthy males. *Neuroimage*, 26(1):206–214, May 2005.
- [HM02] MP Harms and JR Melcher. Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology*, 88(3):1433–1450, September 2002.

- [HMK17] G Hajcak, A Meyer, and R Kotov. Psychometrics and the neuroscience of individual differences: Internal consistency limits between-subjects effects. *Journal of Abnormal Psychology*, 126(6):823–834, August 2017.
- [HNP<sup>+</sup>04] T Hinterberger, N Neumann, M Pham, A Köbler, A Grether, N Hofmayer, B Wilhelm, H Flor, and N Birbaumer. A multimodal brain-based feedback and communication system. *Experimental brain research*, 154(4):521–526, February 2004.
- [HPR<sup>+</sup>02] RNA Henson, CJ Price, MD Rugg, R Turner, and KJ Friston. Detecting latency differences in event-related BOLD responses: Application to words versus nonwords and initial versus repeated face presentations. *NeuroImage*, 15(1):83–97, January 2002.
- [HS07] U Habel and F Schneider. *Funktionelle MRT in Psychiatrie und Neurologie*, chapter 'Emotionen', pages 362–379. Springer Medizin Verlag, Heidelberg, 2007.
- [HSS<sup>+</sup>99] Y Hirayasu, ME Shenton, DF Salisbury, JS Kwon, CG Wible, IA Fisher, D Yurgelun-Todd, C Zarate, R Kikinis, FA Jolesz, and RW McCarley. Subgenual cingulate cortex volume in first episode psychosis. *American Journal of Psychiatry*, 156(7):1091–93, July 1999.
- [HTF17] T Hastie, R Tibshirani, and J Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Verlag, Cham, 2nd edition, 2017.
- [HWD<sup>+</sup>07] U Habel, C Windischberger, B Derntl, S Robinson, I Kryspin-Exner, RC Gur, and E Moser. Amygdala activation and facial expressions: Explicit emotion discrimination versus implicit emotion processing. *Neuropsychologia*, 45(10):2369–77, June 2007.
- [JHL<sup>+</sup>12] KA Johnson, K Hartwell, T LeMatty, J Borckardt, PS Morgan, K Govindarajan, K Brady, and MS George. Intermittent "real-time" fMRI feedback is superior to continuous presentation for a motor imagery task: a pilot study. *Journal of Neuroimaging*, 22(1):58–66, January 2012.
- [JSJ<sup>+</sup>20] XZ Jia, JW Sun, GJ Ji, W Liao, YT Liu, J Wang, Z Wang, H Thang, DQ Liu, and YF Zang. Percent amplitude of fluctuation: A simple measure for resting-state fMRI signal at single voxel level. *PLoS One*, 15(1), January 2020.
- [KAP<sup>+</sup>17] Y Koush, J Ashburner, E Prilepin, R Sladky, P Zeidman, S Bibikov, F Scharnowski, A Nikonorov, and D Van de Ville. OpenNFT: An open-source Python/Matlab framework for real-time fMRI neurofeedback training based on activity, connectivity and multivariate pattern analysis. *Neuroimage*, 156:489–503, August 2017.

- [KBL<sup>+</sup>17] F Krause, C Benjamins, M Lührs, J Eck, Q Noirhomme, M Rosenke, S Brunheim, B Sorger, and R Goebel. Real-time fMRI-based self-regulation of brain activation across different visual feedback presentations. *Brain-Computer Interfaces*, 4(1-2):87–101, April 2017.
- [KMG<sup>+</sup>20] M Klöbl, P Michenthaler, GM Godbersen, S Robinson, A Hahn, and R Lanzenberger. Reinforcement and punishment shape the learning dynamics in fMRI neurofeedback. *Frontiers in human neuroscience*, 24:304, July 2020.
- [KS19] KC Kadosh and G Staunton. A systematic review of the psychological factors that influence neurofeedback learning outcomes. *Neuroimage*, 185:545–555, January 2019.
- [KVE<sup>+</sup>15] L Konicar, R Veit, H Eisenbarth, B Barth, P Tonin, U Strehl, and N Birbaumer. Brain self-regulation in criminal psychopaths. *Scientific Reports*, 5:9426, March 2015.
- [KWA<sup>+</sup>01] ML Kessler-West, AH Andersen, CD Smith, MJ Avison, CE Davis, RJ Kryscio, and LX Blonder. Neural substrates of facial emotion processing using fMRI. *Cognitive Brain Research*, 11(2):213–26, April 2001.
- [KWN<sup>+</sup>13] SE Kober, M Witte, M Ninaus, C Neuper, and G Wood. Learning to modulate one’s own brain activity: the effect of spontaneous mental strategies. *Frontiers in Human Neuroscience*, 7:695, October 2013.
- [LaC11] SM LaConte. Decoding fMRI brain states in real-time. *Neuroimage*, 56(2):440–54, May 2011.
- [LEJ<sup>+</sup>03] J Levesque, F Eugene, Y Joannette, V Paquette, B Mensour, G Beaudoin, JM Leroux, P Bourgouin, and M Beauregard. Neural circuitry underlying voluntary suppression of sadness. *Biological Psychiatry*, 53(6):502–10, March 2003.
- [LHJ<sup>+</sup>12] DEJ Linden, I Habel, SJ Johnston, S Linden, R Tatineni, L Subramanian, B Sorger, D Healy, and R Goebel. Real-time self-regulation of emotion networks in patients with depression. *PLoS One*, 7(6):e38115, June 2012.
- [Lin89] LI Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–68, March 1989.
- [LKK96] LM Lix, JC Keselman, and JJ Keselman. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66(4):579–619, 1996.

- [LLAW09] MA Lindquist, JM Loh, LY Atlas, and TD Wager. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage*, 45:S187–S198, March 2009.
- [LLC<sup>+</sup>12] KM Lee, J Lee, CY Chung, S Ahn, KH Sung, TW Kim, HJ Lee, and MS Park. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopedic research. *Clinics in orthopedic surgery*, 4(2):149–55, June 2012.
- [LLW08] JM Loh, MA Lindquist, and TD Wager. Residual analysis for detecting mis-modeling in fMRI. *Statistica Sinica*, 18(4):1421–1448, October 2008.
- [Log08] NK Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–78, June 2008.
- [LRB<sup>+</sup>97] RD Lane, EM Reiman, MM Bradley, PJ Lang, GL Ahern, RJ Davidson, and GE Schwartz. Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia*, 35(11):1437–44, November 1997.
- [LRSK07] L Leyman, R De Raedt, R Schacht, and EHW Koster. Attentional biases for angry faces in unipolar depression. *Psychological medicine*, 37(3):393–402, March 2007.
- [LS76] JF Lubar and MN Shouse. EEG and behavioral changes in a hyperkinetic child concurrent with training of the sensorimotor rhythm (SMR): a preliminary report. *Biofeedback and Self Regulation*, 1(3):293–306, September 1976.
- [LWK<sup>+</sup>12] KA Lindquist, TD Wager, H Kober, E Bliss-Moreau, and L Feldmann Barrett. The brain basis of emotion: A meta-analytic review. *Behavioral and brain sciences*, 12:121–202, 2012.
- [LWY<sup>+</sup>03] K Lange, LM Williams, AW Young, ET Bullmore, MJ Brammer, SCR Williams, JA Gray, and ML Phillips. Task instructions modulate neural responses to fearful facial expressions. *Biological Psychiatry*, 53:226–232, February 2003.
- [MAK<sup>+</sup>15] KA Mathiak, EM Alawi, Y Koush, M Dyck, JS Cordes, TJ Gaber, FD Zepf, N Palomero-Gallagher, P Sarkheil, S Berger, M Zvyagintsev, and K Mathiak. Social reward improves the voluntary control over localized brain activity in fMRI-based neurofeedback training. *Frontiers in behavioral neuroscience*, 9:136, June 2015.
- [May97] HS Mayberg. Limbic-cortical dysregulation: a proposed model of depression. *Journal of Neuropsychiatry and Clinical Neuroscience*, 9:471–481, Summer 1997.

- [MBT<sup>+</sup>00] HS Mayberg, SK Brannan, JL Tekell, JA Silva, RK Mahurin, S McGinnis, and PA Jerabek. Regional metabolic effects of fluoxetine in major depression: Serial changes and relationship to clinical response. *Biological Psychiatry*, 48(8):830–43, October 2000.
- [MBZ<sup>+</sup>15] M Misaki, J Barzigar, V Zotev, R Phillips, S Cheng, and J Bodurka. Real-time fMRI processing with physiological noise correction-comparison with off-line analysis. *Journal of Neuroscience Methods*, 256:117–21, December 2015.
- [MG15] J Mishra and A Gazzaley. Closed-loop cognition: the next frontier arrives. *Trends in Cognitive Science*, 19(5):242–3, May 2015.
- [MHN05] IJ Mitchell, H Heims, EA Neville, and H Rickards. Huntington’s disease patients show impaired perception of disgust in the gustatory and olfactory modalities. *Journal of Neuropsychiatry and Clinical Neuroscience*, 17(1):119–21, 2005.
- [MJM<sup>+</sup>16] M Marxen, MJ Jacob, DK Müller, S Posse, E Achley, L Hellrung, P Riedel, S Bender, R Epple, and MN Smolka. Amygdala regulation following fMRI-neurofeedback without instructed strategies. *Frontiers in human neuroscience*, 10:183, April 2016.
- [MLB<sup>+</sup>99] HS Mayberg, M Liotti, SK Brannan, S McGinnis, RK Mahurin, PA Jerabek, JA Silva, JL Tekell, CC Martin, JL Lancaster, and PT Fox. Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *American Journal of Psychiatry*, 156(5):675–682, May 1999.
- [MLV<sup>+</sup>05] HS Mayberg, AM Lozano, V Voon, HE McNeely, D Seminowicz, C Hamani, JM Schwalb, and SH Kennedy. Deep brain stimulation for treatment-resistant depression. *Neuron*, 45(5):651–60, March 2005.
- [MMM16] H Marzbani, HR Marateb, and M Mansourian. Neurofeedback: A comprehensive review on system design, methodology and clinical applications. *Basic and Clinical Neuroscience*, 7(2):143–158, April 2016.
- [Moo11] H Moosbrugger. *Lineare Modelle. Regressions- und Varianzanalysen*. Verlag Hans Huber, Bern, 4th edition, 2011.
- [MSPJG08] S Markovska-Simonska, N Pop-Jordanova, and D Georgiev. Simultaneous EEG and EMG biofeedback for peak performance in musicians. *Prilozi*, 29(1):239–52, July 2008.
- [MSS13] B Mirkovic, M Stevanovic, and A Savic. EEG controlled Ni lego robot: Feasibility study of sensorimotor alpha rhythm neurofeedback in children. *Biomedizinische Technik*, 58(Suppl 1), August 2013.

- [MYKI15] F Megumi, A Yamashita, M Kawato, and H Imamizu. Functional MRI neurofeedback training on connectivity between two regions induces long-lasting changes in intrinsic functional network. *Frontiers in Human Neuroscience*, 9:160, March 2015.
- [OKK<sup>+</sup>00] EA Osuch, TA Ketter, TA Kimbrell, MS George, BE Benson, MW Willis, P Herscovitch, and RM Post. Regional cerebral metabolism associated with anxiety symptoms in affective disorder patients. *Biological Psychiatry*, 48(10):1020–1023, November 2000.
- [OSB14] KN Ochsner, JA Silvers, and JT Buhle. Functional imaging studies of emotion regulation: A synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, 1251:E1–E24, August 2014.
- [PBH<sup>+</sup>98] ML Phillips, ET Bullmore, R Howard, PW Woodruff, IC Wriath, SC Williams, A Simmons, C Andrew, M Brammer, and AS David. Investigation of facial recognition memory and happy and sad facial expression perception: an fMRI study. *Psychiatry Research*, 83(3):127–38, September 1998.
- [PBK<sup>+</sup>02] S Pilling, P Bebbington, E Kuipers, P Garety, J Geddes, G Orbach, and C Morgan. Psychological treatments in schizophrenia: I. meta-analysis of family intervention and cognitive behaviour therapy. *Psychological medicine*, 32(5):763–82, July 2002.
- [PDRL03] ML Phillips, WC Drevets, SL Rauch, and R Lane. Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biological Psychiatry*, 54(5):515–528, September 2003.
- [PLD08] ML Phillips, CD Ladouceur, and WC Drevets. A neural model of voluntary and automatic emotion regulation: Implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Molecular Psychiatry*, 13(9):833–57, September 2008.
- [PMN11] RA Poldrack, JA Mumford, and TE Nichols. *Handbook of functional MRI Data Analysis*. Cambridge University Press, New York, 2011.
- [PWG<sup>+</sup>99] S Posse, S Wiese, D Gembris, K Mathiak, C Kessler, ML Grosse-Ruyken, B Elghahwagi, T Richards, SR Dager, and VG Kiselev. Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic resonance in medicine*, 42(1):87–97, July 1999.
- [PWTL02] KL Phan, T Wagner, SF Taylor, and I Liberzon. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16(2):331–48, June 2002.

- [RBR<sup>+</sup>14] S Ruiz, K Buyukturkoglu, M Rana, N Birbaumer, and R Sitaram. Real-time fMRI brain computer interfaces: Self-regulation of single brain regions to networks. *Biological psychology*, 95:4–20, January 2014.
- [RD08] SL Rauch and WC Drevets. *Stress-Induced and Fear Circuitry Disorders*, chapter 'Neuroimaging and the neuroanatomy of stress-induced and fear circuitry disorders: the agenda for future research', pages 235–278. American Psychiatric association, 2008.
- [RKG<sup>+</sup>17] M Ramot, S Kimmich, J Gonzalez-Castillo, V Roopchansingh, H Popal, E White, SJ Gotts, and A Martin. Direct modulation of aberrant brain network connectivity through real-time neurofeedback. *Elife*, 6:e28974, September 2017.
- [RLA<sup>+</sup>97] EM Reiman, RD Lane, GL Ahern, GE Schwartz, RJ Davidson, KJ Friston, LS Yun, and K Chen. Neuroanatomical correlates of externally and internally generated human emotion. *American Journal of Psychiatry*, 154(7):918–25, July 1997.
- [RLS<sup>+</sup>13] S Ruiz, S Lee, SR Soekadar, A Caria, R Veit, T Kircher, N Birbaumer, and R Sitaram. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. *Human Brain Mapping*, 34(1):200–12, January 2013.
- [RM07] KJ Ressler and HS Mayberg. Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nature Neuroscience*, 10(9):1116–24, September 2007.
- [RMK<sup>+</sup>17] F Robineau, DE Meskaldji, Y Koush, SW Rieger, C Mermoud, S Morgenthaler, D Van de Ville, P Vuilleumier, and F Scharnowski. Maintenance of voluntary self-regulation learned through real-time fMRI neurofeedback. *Frontiers in Human Neuroscience*, 11:131, March 2017.
- [RMR<sup>+</sup>10] T Ros, MAM Munneke, D Ruge, JH Gruzelier, and JC Rothwell JC. Endogenous control of waking brain rhythms induces neuroplasticity in humans. *The European Journal of Neuroscience*, 31(4):770–8, February 2010.
- [RTW<sup>+</sup>06] AJ Rush, MH Trivedi, SR Wisniewski, AA Nierenberg, JW Stewart, D Warden, G Niederehe, ME Thase, PW Lavor, BD Lebowitz, PJ McGrath, JF Rosenbaum, HA Sackeim, DJ Kupfer, J Luther, and M Fava. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *American Journal of Psychiatry*, 163(1):1905–17, November 2006.
- [SA03] MS Schwartz and FE Andrasik. *Biofeedback: A Practitioner's Guide*, page 930. The Guilford Press, New York, 2003.

- 
- [SAL<sup>+</sup>11] L Sherlin, M Arns, J Lubar, H Heinrich, C Kerson, U Strehl, and B Sterman. Neurofeedback and basic learning theory: Implications for research and practice. *Journal of Neurotherapy*, 15:292–304, October 2011.
- [Sap02] CB Saper. The central autonomic nervous system: conscious visceral perception and autonomic pattern generation. *Annual Review of Neuroscience*, 25:433–469, March 2002.
- [SCA08] K Sergerie, C Chochol, and JL Armony. The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 32(4):811–30, January 2008.
- [SCM06] SR Searle, G Casella, and CE McCulloch. *Variance Components*. Wiley-Interscience, Hoboken, 2006.
- [SCV<sup>+</sup>07] R Sitaram, A Caria, R Veit, T Gaber, G Rota, A Kuebler, and N Birbaumer. fMRI brain-computer interface: a tool for neuroscientific research and treatment. *Computational intelligence and neuroscience*, 2007:25487, 2007.
- [SD05] T Suslow and U Dannlowski. *Mood state and health*, chapter 'Detection of Facial Emotion in Depression', pages 1–32. Nova Biomedical Books, 2005.
- [SDLM<sup>+</sup>04] T Suslow, U Dannlowski, J Lalee-Mentzel, US Donges, V Arolt, and A Kersting. Spatial processing of facial emotion in patients with unipolar depression: a longitudinal study. *Journal of Affective Disorder*, 83(1):59–63, November 2004.
- [SE06] MB Serman and T Egner. Foundation and practice of neurofeedback for the treatment of epilepsy. *Applied Psychophysiological Biofeedback*, 31(1):21–35, March 2006.
- [SGG<sup>+</sup>14] LE Stoeckel, KA Garrison, SS Ghosh, P Wighton, CA Hanlon, JM Gilman, S Green, NB Turk-Browne, MT deBettencourt, D Scheinost, C Craddock, T Thompson, V Calderon, CC Bauer, M George, HC Breiter, S Whitfield-Gabrieli, JD Gabrieli, SM LaConte, L Hirshberg, JA Brewer, M Hampson, A Van der Kouwe, S Mackey, and AE Evins. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *NeuroImage.Clinical*, 5:245–55, July 2014.
- [SGGM94] F Schneider, RC Gur, RE Gur, and LR Muenz. Standardized mood induction with happy and sad facial expressions. *Psychiatry Research*, 51(1):19–31, January 1994.
- [SH98] O Speck and J Hennig. Functional imaging by Io- and T2\*-parameter mapping using multi-image EPI. *Magnetic resonance in Medicine*, 40(2):243–248, August 1998.

- [SHJ<sup>+</sup>11] L Subramanian, JV Hindle, S Johnston, MV Roberts, M Husain, R Goebel, and D Linden. Real-time functional magnetic resonance imaging neurofeedback for treatment of Parkinson's disease. *Journal of Neuroscience*, 31(45):16309–17, November 2011.
- [SHK<sup>+</sup>00] F Schneider, U Habel, C Kessler, JB Salloum, and S Posse. Gender differences in regional cerebral activity during sadness. *Human Brain Mapping*, 9(4):226–38, April 2000.
- [SHS<sup>+</sup>13] J Sulzer, S Haller, F Scharnowski, N Weisskopf, N Birbaumer, ML Blefari, AB Bruehl, LG Cohen, RC DeCharms, R Gassert, R Goebel, U Herwig, S LaConte, D Linden, A Luft, E Seifritz, and R Sitaram. Real-time fMRI neurofeedback: progress and challenges. *Neuroimage*, 76:386–99, August 2013.
- [SJH<sup>+</sup>02] GS Sanders, GG Gallup Jr., H Heinsen, PR Hof, and C Schmitz. Cognitive deficits, schizophrenia, and the anterior cingulate cortex. *Trends in Cognitive Science*, 6(5):190–192, May 2002.
- [SKK<sup>+</sup>10] T Suslow, C Konrad, H Kugel, D Rumstadt, P Zwitserlood, A Schöning, P Ohrmann, J Bauer, M Pyka, A Kersting, V Arolt, W Heindel, and DU Dannlowski. Automatic mood-congruent amygdala responses to masked facial expressions in major depression. *Biological Psychiatry*, 67(2):155–160, January 2010.
- [SKY<sup>+</sup>04] W Sato, T Kochiyama, S Yoshikawa, E Naito, and M Matsumura. Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Brain research. Cognitive brain research*, 20(1):81–91, June 2004.
- [SLG<sup>+</sup>06] U Strehl, U Leins, G Goth, C Klinger, T Hinterberger, and N Birbaumer. Self-regulation of slow cortical potentials: a new treatment for children with attention-deficit/hyperactivity disorder. *Pediatrics*, 118(5):e1530–40, November 2006.
- [SLI<sup>+</sup>14] MO Sokunbi, DEJ Linden, I Habel I, S Johnston, and N Ihssen. Real-time fMRI brain-computer interface: development of a "motivational feedback" subsystem for the regulation of visual cue reactivity. *Frontiers in Behavioral Neuroscience*, 8:392, November 2014.
- [SNP<sup>+</sup>20] F Scharnowski, AA Nicholson, S Pichon, MJ Rosa, G Rey, SB Eickhoff, D Van De Ville, P Vuilleumier, and Y Koush Y. The role of the subgenual anterior cingulate cortex in dorsomedial prefrontal-amygdala neural circuitry during positive-social emotion regulation. *Human Brain Mapping*, 41(11):3100–3118, August 2020.

- [SPC17] Y Shoji, CR Patti, and D Cvetkovic. Electroencephalographic neurofeedback to up-regulate frontal theta rhythms: Preliminary results. *Annual International Conference of the IEEE engineering in medicine and biology society*, pages 1425–1428, July 2017.
- [SREP98] R Sprengelmeyer, M Rausch, UT Eysel, and H Przuntek. Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society B: Biological Sciences*, 265(1409):1927–31, October 1998.
- [SSD11] A Stuhmann, T Suslow, and U Danlowski. Facial emotion processing in major depression: a systematic review of neuroimaging findings. *Biology of Mood and Anxiety Disorders*, 1(1):10, November 2011.
- [SSR<sup>+</sup>16] P Sepulveda, R Sitaram, M Rana, C Montalba, C Tejos, and S Ruiz. How feedback, motor imagery, and reward influence brain self-regulation using real-time fMRI. *Human Brain Mapping*, 37(9):3153–71, September 2016.
- [SSS<sup>+</sup>05] A Schienle, A Schäfer, R Stark, B Walter, and D Vaitl. Gender differences in the processing of disgust- and fear-inducing pictures: an fMRI study. *Neuroreport*, 16(3):277–80, February 2005.
- [STH<sup>+</sup>12] Y Sagi, I Tavor, S Hofstetter, S Tzur-Moryosef, T Blumenfeld-Katzir, and Y Assaf. Learning in the fast lane: New insights into neuroplasticity. *Neuron*, 73(6):1195–203, March 2012.
- [Str14] U Strehl. What learning theories can teach us in designing neurofeedback treatments. *Frontiers in Human Neuroscience*, 8:894, November 2014.
- [SWR69] MB Serman, W Wyrwicka, and S Roth. Electrophysiological correlates and neural substrates of alimentary behavior in the cat. *Annals of the New York Academy of Science*, 157(2):723–39, May 1969.
- [TA12] J Townsend and LL Altshuler. Emotion processing and regulation in bipolar disorder: a review. *Bipolar Disorders*, 14(4):326–39, June 2012.
- [TC02] C Tsigos and GP Chrousos. Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. *Journal of psychosomatic research*, 53(4):865–871, October 2002.
- [TFM09] SA Trautmann, T Fehr, and M Herrmann M. Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain research*, 1284:100–115, August 2009.

- [THC<sup>+</sup>99] JD Teasdale, RJ Howard, SG Cox, Y Ha, MJ Brammer, and SC Williams SA Checkley. Functional MRI study of the cognitive generation of affect. *American Journal of Psychiatry*, 156(2):209–215, February 1999.
- [TJ00] JF Thayer and BH Johnson. Sex differences in judgement of facial affect: a multivariate analysis of recognition errors. *Scandinavian Journal of Psychology*, 41(3):243–6, September 2000.
- [TSH<sup>+</sup>04] T Takahashi, M Suzuki, H Hagino, SY Zhou, Y Kawasaki, S Nohara, K Nakamura, I Yamashita, H Seto, and M Kurachi. Bilateral volume reduction of the insular cortex in patients with schizophrenia: a volumetric MRI study. *Psychiatry Research*, 131(3):185–94, September 2004.
- [TTH<sup>+</sup>09] G Tan, J Thornby, D Corydon Hammond, U Strehl, B Canady, K Arne-  
mann, and DA Kaiser. Meta-analysis of EEG biofeedback in treating  
epilepsy. *Clinical EEG and Neuroscience*, 40(3):173–9, July 2009.
- [Ver05] DJ Vernon. Can neurofeedback training enhance performance? An eval-  
uation of the evidence with implications for future research. *Applied  
Psychophysiological Biofeedback*, 30(4):347–64, December 2005.
- [VV05] VGS Vasdekis and IG Vlachonikolis. On the difference between ML  
and REML estimators in the modelling of multivariate longitudinal data.  
*Journal of Statistical Planning and Inference*, 134(1):294–205, September  
2005.
- [WEE20] LA Weber, T Ethofer, and AC Ehlis. Predictors of neurofeedback training  
outcome: A systematic review. *Neuroimage.Clinical*, 27:102301, May 2020.
- [Wei12] N Weisskopf. Real-time fMRI and its application to neurofeedback. *Neu-  
roimage*, 62(2):682–92, August 2012.
- [WHM<sup>+</sup>14] O Wilhelm, A Hildebrandt, K Manske, A Schacht, and W Sommer. Test  
battery for measuring the perception and recognition of facial expressions  
of emotion. *Frontiers of Psychology*, pages 5–404, May 2014.
- [WKH07] A Wohlschläger, T Kellermann, and U Habel. *Funktionelle MRT in Psy-  
chiatric und Neurologie*, chapter 'Datenanalyse: Vorverarbeitung, Statistik  
und Auswertung', pages 134–147. Springer Medizin Verlag, Heidelberg,  
2007.
- [WKN<sup>+</sup>13] M Witte, SE Kober, M Ninaus, C Neuper, and G Wood. Control be-  
liefs can predict the ability to up-regulate sensorimotor rhythm during  
neurofeedback training. *Frontiers in Human Neuroscience*, 7:478, August  
2013.

- 
- [WMS<sup>+</sup>02] CI Wrioth, B Martis, LM Shin, H Fischer, and SL Rauch. Enhanced amygdala responses to emotional versus neutral schematic facial expressions. *Neuroreport*, 13(6):785–90, May 2002.
- [WRE<sup>+</sup>98] PJ Whalen, SL Rauch, NL Etcoff, SC McInerney, MB Lee, and MA Jenike. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, 18(1):411–418, January 1998.
- [YLO<sup>+</sup>07] SS Yoo, JH Lee, H O’Leary, V Lee, SE Choo, and FA Jolesz. Functional magnetic resonance imaging-mediated learning of increased activity in auditory areas. *Neuroreport*, 18(18):1915–20, December 2007.
- [YLO<sup>+</sup>08] SS Yoo, JH Lee, H O’Leary, LP Panych, and FA Jolesz. Neurofeedback fMRI-mediated learning and consolidation of regional brain activation during motor imagery. *International Journal of Imaging Systems and Technology*, 18(1):69–78, June 2008.
- [YR09] Y Yang and A Raine. Neuroimaging: Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: A meta-analysis. *Psychiatric Research*, 174:81–88, November 2009.
- [YS06] A Yoshikawa and W Sato. Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotion. *Japanese Psychological Research*, 48(3):213–222, September 2006.
- [YZP<sup>+</sup>14] KD Young, V Zotev, R Phillips, M Misaki, H Yuan, WC Drevets, and J Bodurka. Real-time fMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PLoS One*, 9(2):e88785, February 2014.
- [ZWB<sup>+</sup>19] R Zahn, JH Weingartner, R Basilio, P Bado, P Mattos, JR Sato, R de Oliveira-Souza, LF Fontenelle, AH Young, and J Moll. Blame-rebalance fMRI neurofeedback in major depressive disorder: A randomised proof-of-concept trial. *NeuroImage: Clinical*, 24:101992, August 2019.